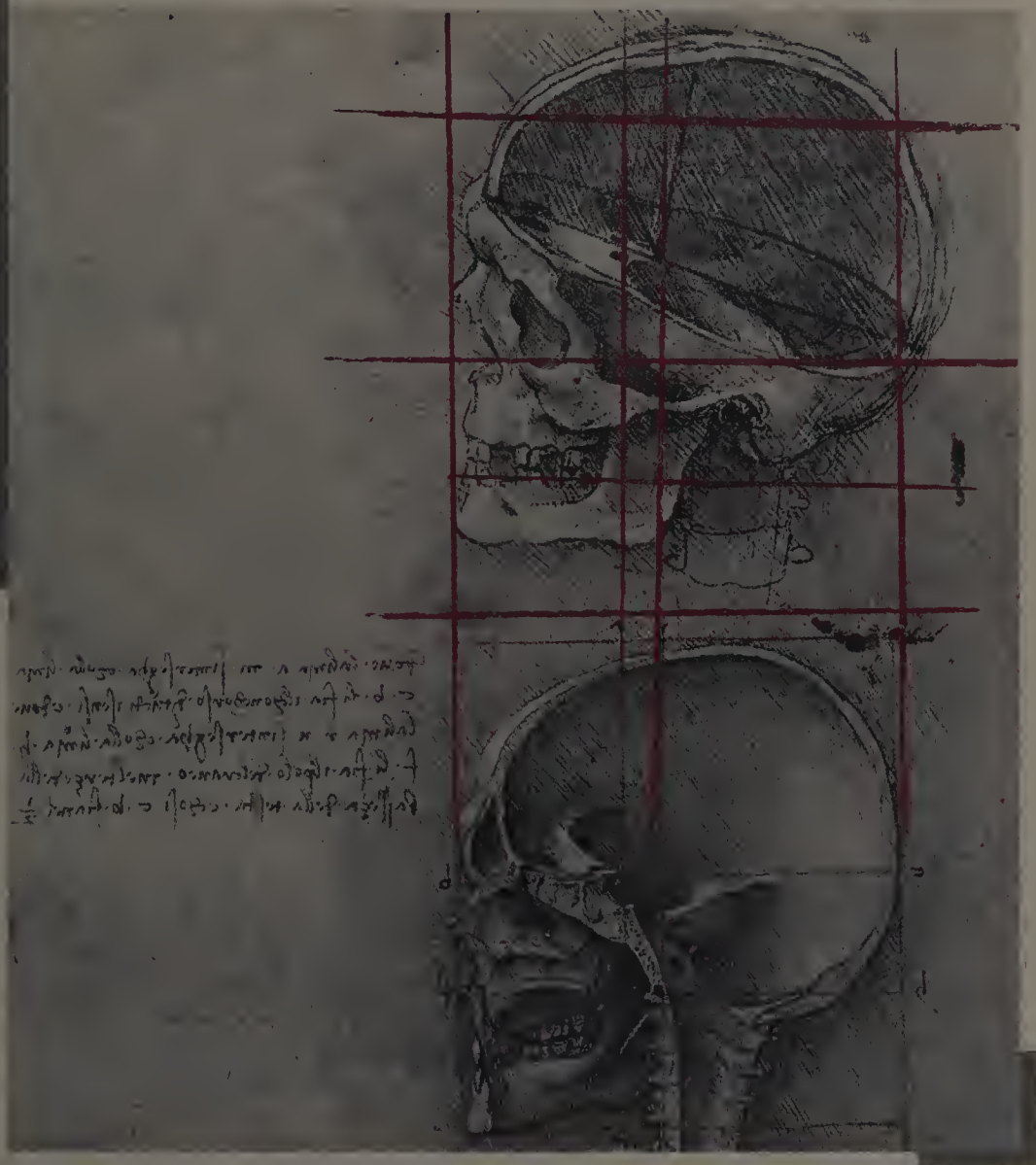


BIBLIOTECA MONDADORI



ROGER PENROSE LA NUEVA MENTE DEL EMPERADOR

Q335.P418 1991

Penrose, Roger.

La nueva mente del emperador

3 1140 0001 7279 8

Palomar College Library

No sería extraño que en el futuro nuestro siglo fuese conocido como el de la ciencia y la tecnología. Es cierto que no hemos inventado el método científico, que hemos estado subidos a «hombros de gigantes», pero no lo es menos que durante el siglo xx se han producido cambios cualitativos y cuantitativos radicales en nuestra comprensión y dominio de la Naturaleza. Para bien o para mal, ya pocas cosas son como lo fueron en el pasado. Con rigor, sin buscar lo fácil, pero sin renunciar por ello a la claridad asociada al conocimiento auténtico, la Biblioteca Mondadori (Ciencia) pretende facilitar a los lectores el acceso a ese mundo de la ciencia y la tecnología del que todos somos, nos guste o no, ciudadanos.

DISCARD

DISCARD

La nueva mente del emperador

BIBLIOTECA MONDADORI (CIENCIA)

Director: JOSÉ MANUEL SÁNCHEZ RON



Roger Penrose

La nueva mente del emperador

Biblioteca Mondadori

PALOMAR COLLEGE LIBRARY
SAN MARCOS, CA

Título original: *The Emperor's New Mind*

Traducción de Javier García Sanz

© OXFORD UNIVERSITY PRESS, 1989

© JAVIER GARCÍA SANZ DE LA TRADUCCIÓN, 1991

© MONDADORI ESPAÑA, S. A., 1991

AVDA. ALFONSO XIII, 50 - MADRID

ISBN: 84-397-1786-5

D.L.: 40769-1991

IMPRESIÓN: LAVEL

Printed in Spain

PRIMERA EDICIÓN: SEPTIEMBRE, 1991

SEGUNDA EDICIÓN: OCTUBRE, 1991

TERCERA EDICIÓN: NOVIEMBRE, 1991

CUARTA EDICIÓN: DICIEMBRE, 1991

Quedan rigurosamente prohibidas, sin la autorización escrita de los titulares del «Copyright», bajo las sanciones establecidas en las leyes, la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la reprografía y el tratamiento informático y la distribución de ejemplares de ella mediante alquiler o préstamo públicos.

Índice

PREFACIO	11
NOTA PARA EL LECTOR	15
AGRADECIMIENTOS	16
PROCEDENCIA DE LAS ILUSTRACIONES	18
PRÓLOGO	21
1. ¿PUEDE TENER MENTE UN COMPUTADOR?	23
Introducción, 23	
El test de Turing, 26	
Inteligencia artificial, 33	
Una aproximación IA al «placer» y al «dolor», 36	
IA fuerte y la habitación china de Searle, 40	
Hardware y software, 48	
2. ALGORITMOS Y MÁQUINAS DE TURING	56
Fundamentos del concepto de algoritmo, 56	
El concepto de Turing, 62	
Codificación binaria de los datos numéricos, 71	
La tesis de Church-Turing, 76	
Números diferentes de los naturales, 79	
La máquina universal de Turing, 81	
La insolubilidad del problema de Hilbert, 90	
¿Cómo superar a un algoritmo?, 98	
El cálculo lambda de Church, 100	

3. MATEMÁTICAS Y REALIDAD	107
La tierra de Tor'Bled-Nam, 107	
Números reales, 114	
¿Cuántos números reales hay?, 117	
«Realidad» de los números reales, 121	
Números complejos, 122	
Construcción del conjunto de Mandelbrot, 129	
¿Realidad platónica de los conceptos matemáticos?, 131	
4. VERDAD, DEMOSTRACIÓN E INTUICIÓN DIRECTA	136
El programa de Hilbert para las matemáticas, 136	
Sistemas matemáticos formales, 140	
El teorema de Gödel, 144	
La intuición matemática, 147	
¿Platonismo o intuicionismo?, 153	
Teoremas tipo-Gödel a partir del resultado de Turing, 158	
Conjuntos recursivamente enumerables, 161	
¿Es recursivo el conjunto de Mandelbrot?, 168	
Algunos ejemplos de matemáticas no-recursivas, 174	
¿Es el conjunto de Mandelbrot semejante a la matemática no-recursiva?, 183	
Teoría de la complejidad, 187	
Complejidad y computabilidad en los objetos físicos, 193	
5. EL MUNDO CLÁSICO	195
El status de la teoría física, 195	
Geometría euclídea, 204	
La dinámica de Galileo y Newton, 211	
El mundo mecanicista de la dinámica newtoniana, 218	
¿Es computable la vida en el mundo de las bolas de billar?, 221	
Mecánica hamiltoniana, 226	
Espacio de fases, 229	
La teoría electromagnética de Maxwell, 239	
Computabilidad y ecuación de ondas, 243	
La ecuación de movimiento de Lorentz; partículas desbocadas, 245	
La relatividad especial de Einstein y Poincaré, 248	
La relatividad general de Einstein, 261	
Causalidad relativista y determinismo, 273	
Computabilidad en física clásica: ¿Dónde estamos?, 278	
Masa, materia y realidad, 280	
6. MAGIA CUÁNTICA Y MISTERIO CUÁNTICO	286
¿Necesitan los filósofos la teoría cuántica?, 286	
Problemas con la teoría clásica, 289	
Los comienzos de la teoría cuántica, 291	
El experimento de la doble-rendija, 294	

- Amplitudes de probabilidad, 300
- El estado cuántico de una partícula, 308
- El principio de incertidumbre, 314
- Los procedimientos de evolución U y R , 316
- ¿Partículas en dos lugares a la vez?, 319
- Espacio de Hilbert, 325
- Medidas, 329
- El spin y la esfera de estados de Riemann, 333
- Objetividad y medibilidad de los estados cuánticos, 339
- Copia de un estado cuántico, 341
- El spin del fotón, 342
- Objetos con gran spin, 345
- Sistemas de muchas partículas, 347
- La «paradoja» de Einstein, Podolsky y Rosen, 353
- Experimentos con fotones: ¿un problema para la relatividad?, 360
- La ecuación de Schrödinger y la ecuación de Dirac, 363
- La teoría cuántica de campos, 365
- El gato de Schrödinger, 366
- Diversas actitudes hacia la teoría cuántica existente, 369
- ¿Dónde nos deja todo esto?, 373

7. LA COSMOLOGÍA Y LA FLECHA DEL TIEMPO 378

- El flujo del tiempo, 378
- El inexorable incremento de la entropía, 381
- ¿Qué es la entropía?, 387
- La segunda ley en acción, 393
- El origen de la baja entropía en el Universo, 398
- La cosmología y el big bang, 404
- La bola de fuego primordial, 408
- ¿Explica el big bang la segunda ley?, 411
- Agujeros negros, 413
- La estructura de las singularidades del espacio-tiempo, 420
- ¿Hasta qué punto fue especial el big bang?, 425

8. EN BUSCA DE LA GRAVITACIÓN CUÁNTICA 433

- ¿Por qué la gravitación cuántica?, 433
- ¿Qué hay detrás de la hipótesis de curvatura de Weyl?, 436
- Asimetría temporal en la reducción del vector de estado, 441
- La caja de Hawking: ¿una conexión con la hipótesis de curvatura de Weyl?, 447
- ¿Cuándo se reduce el vector de estado?, 457

9. CEREBROS REALES Y MODELOS DE CEREBRO 463

- ¿Cómo son realmente los cerebros?, 463
- ¿Dónde está la sede de la consciencia?, 472
- Experimentos de escisión cerebral, 476
- Ceguera cortical, 479

Procesamiento de la información en el córtex visual, 480
 ¿Cómo funcionan las señales nerviosas?, 482
 Modelos de computador, 486
 Plasticidad cerebral, 491
 Computadores paralelos y la «unicidad» de la consciencia, 493
 ¿Hay un papel para la mecánica cuántica en la actividad cerebral?, 495
 Computadores cuánticos, 497
 ¿Más allá de la teoría cuántica?, 499

10. ¿DÓNDE RESIDE LA FÍSICA DE LA MENTE? 502

¿Para qué están las mentes?, 502
 ¿Qué hace realmente la consciencia?, 507
 ¿Selección natural de algoritmos?, 513
 La naturaleza no-algorítmica de la intuición matemática, 516
 Inspiración, intuición directa y originalidad, 518
 La no-verbalidad del pensamiento, 525
 ¿Consciencia animal?, 527
 Contacto con el mundo de Platón, 529
 Una visión de la realidad física, 532
 Determinismo y determinismo fuerte, 534
 El principio antrópico, 537
 Teselaciones y cuasicristales, 538
 Posible relevancia para la plasticidad cerebral, 542
 Los retardos temporales de la consciencia, 544
 El extraño papel del tiempo en la percepción consciente, 549
 Conclusión: una visión de niño, 554

EPÍLOGO 557

NOTAS 558

REFERENCIAS 579

ÍNDICE ANALÍTICO 590

Prefacio

por Martin Gardner

Para muchos grandes matemáticos y físicos resulta difícil, si no imposible, escribir un libro que pueda ser entendido por los no profesionales. Hasta este año se podría haber pensado que Roger Penrose, uno de los físico-matemáticos más eruditos y creativos del mundo, pertenecía a esta clase. Quienes habíamos leído sus artículos y conferencias no técnicas 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 interesado. Creo que este libro llegará a ser un clásico.

Aunque los capítulos del libro de Penrose recorren la teoría de la relatividad, la mecánica cuántica y la cosmología, su principal interés está 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 en sólo cuestión de uno o dos siglos (algunos rebajan este tiempo a cincuenta años) los computadores electrónicos harán todo lo que una mente humana puede hacer. Estimulados por la ciencia-ficción que leyeron en su juventud, y convencidos de que nuestras mentes son simplemente «computadores hechos de carne» (como Marvin Minsky dijo en cierta ocasión), dan por supuesto que el placer y el dolor, la estimación de la belleza y el humor, la consciencia y el libre albedrío son capacidades que emergerán de modo natural cuando el comportamiento algorítmico 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 profundidad) están en abierto desacuerdo. Para ellos un computador no es esencialmente diferente de las calculadoras mecánicas que fun-

cionan con ruedas, palancas o cualquier otro mecanismo que transmita señales. (Se puede construir un computador a base de bolas que ruedan o agua que se mueve por tuberías.) Puesto que la electricidad viaja por los cables conductores con una rapidez mucho mayor 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» un computador electrónico lo que está haciendo en una medida superior a la «comprensión» de un ábaco? Los computadores juegan ahora al ajedrez como un gran maestro. ¿«Comprenden» su juego mejor de lo que hace la máquina de jugar a tres en raya que en cierta ocasión construyeron unos desguazadores de computadores con trozos de chatarra?

El libro de Penrose es el ataque más poderoso nunca escrito contra la IA fuerte. Durante los siglos pasados se han levantado objeciones contra el alegato reduccionista de que una mente es una máquina que funciona según las conocidas leyes de la física; pero la ofensiva de Penrose es más convincente porque hace uso de informaciones de las que no disponían los escritores anteriores. Penrose se revela en el libro como algo más que un físico-matemático: es también un filósofo de primera fila, que no teme abordar problemas que los filósofos contemporáneos despachan calificándolos de sin sentido.

Penrose tiene también el valor de mantener, frente al creciente rechazo de un pequeño grupo de físicos, un vigoroso realismo. No sólo el Universo «está ahí», sino que la verdad matemática tiene también su propia independencia e intemporalidad misteriosa. Como Newton y Einstein, Penrose tiene un profundo sentido de humildad y respeto hacia el mundo físico, tanto 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 mejores demostraciones. A los matemáticos se les permite de cuando en cuando echar una ojeada a parte de una página. Penrose cree que cuando un físico o matemático experimenta una repentina inspiración «ajá» no se trata simplemente de algo «producido por un cálculo complicado»: es la mente que 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 de Penrose están dedicadas a la famosa estructura de tipo fractal conocida como conjunto de Mandelbrot, por ser Benoit Mandelbrot quien la descubrió. Aunque es autosimilar en

sentido estadístico, a medida que sus partes son ampliadas, su estructura con infinitas circunvoluciones cambia de maneras impredecibles. Penrose encuentra incomprensible (igual que yo) que nadie pueda suponer que esta exótica estructura no «esté ahí» igual que lo está el Monte Everest, sujeta a exploración de la misma forma que se explora una selva.

Penrose forma parte también del cada vez mayor grupo de físicos que piensan que Einstein no era tan obstinado y confuso cuando afirmaba que su «voz interior» le decía que la mecánica cuántica es incompleta. Para apoyar esta afirmación, Penrose le lleva por un fascinante recorrido que cubre 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 de fases, los espacios de Hilbert, los agujeros negros, los agujeros blancos, la radiación de Hawking, la entropía, 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, el transferir a una persona de un lugar a otro del modo como eran transmitidos o recibidos los astronautas de la serie de televisión *Star Trek*? ¿Qué valor de supervivencia descubrió la evolución al producir 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 firmemente imbricados? ¿Son las leyes de la mecánica cuántica, o quizá otras leyes aún más profundas, esenciales para la actuación de una mente?

La respuesta de Penrose a las dos últimas preguntas es afirmativa. Su famosa teoría de «twistors» —objetos geométricos abstractos que operan en un espacio complejo multidimensional que subyace al 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 los campos y partículas de la mecánica cuántica. En su clasificación de las teorías en las cuatro categorías de soberbias, útiles, tentativas y descaminadas, Penrose coloca modestamente a la teoría de twistors en la clase de las tentativas, junto con las supercuerdas u otros esquemas de gran unificación que hoy son vivamente debatidos.

Penrose es desde 1973 Rouse Ball Professor de Matemática en la Universidad de Oxford. El título es apropiado ya que W.W. Rouse Ball no sólo fue un matemático notable sino también un mago aficionado con un interés tan apasionado por las matemáticas recreativas

que escribió la obra clásica en inglés 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 un dibujo de una figura sólida que no puede existir debido a que incorpora elementos autocontradictorios.) Él y su padre Lionel, un genetista, enrollaron el tribar en la Escalera de Caracol de Penrose, una estructura que Maurits Escher utilizó en dos famosas litografías: *Ascenso y Descenso* y *Cascada*. Un día que Penrose estaba tumbado en la cama visualizó, en lo que él llamó un «arrebato de locura», un objeto imposible en un espacio tetradimensional. Es algo, decía, que si se le mostrase a una criatura de un espacio de cuatro dimensiones le haría exclamar: «¡Dios mío!, ¿qué es esto?»

Durante los años 60, cuando Penrose trabajaba en cosmología con su amigo Stephen Hawking, hizo lo que tal vez sea su descubrimiento más conocido. Si la teoría de la relatividad es válida «hasta el final» debe haber una singularidad en todo agujero negro en la que dejan de ser aplicables las leyes de la física. Incluso este logro ha sido eclipsado en años recientes por la construcción que hizo Penrose de dos formas que teselan el plano, a la manera de la teselación de Escher, pero que sólo pueden hacerlo de 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 teselas pueden encontrarse en la base de un extraño y nuevo tipo de materia. El estudio de estos «cuasicristales» es hoy 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 logros de Penrose en matemáticas y física —y solo he mencionado una pequeña parte— surgen de una perpetua admiración por el misterio y 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 su prólogo y epílogo es en parte un símbolo del amanecer de la consciencia en la lenta evolución de la vida sensible. Para mí, es también Penrose —el niño sentado en la tercera fila, detrás de los líderes de la IA— quien se atreve a sugerir que los emperadores de la IA fuerte van desnudos. Muchas de las opiniones de Penrose están salpicadas de humor, pero éste no es un tema de risa.

Nota para el lector

Sobre la lectura de las ecuaciones matemáticas

En varios lugares de este libro he recurrido al uso de ecuaciones matemáticas, desoyendo impertérrito las advertencias frecuentemente hechas de que cada una de estas fórmulas reduciría a la mitad el número de lectores. Si usted es uno de esos lectores que se siente intimidado ante una fórmula (y la mayoría de la gente es así), entonces le recomiendo un método que yo mismo adopto normalmente cuando se presenta una de estas líneas fastidiosas. El método consiste, más o menos, en pasar por alto esa línea ¡y saltar a la siguiente línea de texto real! Bueno, no exactamente esto; uno debería echar una rápida ojeada a la pobre fórmula, en lugar de tratar de comprenderla completamente, y luego seguir adelante. Algún tiempo después, y armado con nueva confianza, uno puede volver a la fórmula olvidada y tratar de captar alguna de sus características más sobresalientes. El propio texto puede servir de ayuda para saber lo que es importante y lo que puede ser pasado por alto sin problemas. Si no lo consigue, entonces no tenga miedo de prescindir por completo de la fórmula.

Agradecimientos

Muchas personas me han ayudado, de una u otra forma, a escribir este libro, y a ellas debo dar las gracias. En particular, a los defensores de la IA fuerte (especialmente a los que intervinieron en un programa de la BBC TV que presencié en cierta ocasión) quienes, al expresar opiniones tan extremas, me incitaron, hace ya varios años, a embarcarme en este proyecto. (Pese a todo, ¡me temo que si entonces hubiera sabido los trabajos futuros que la escritura me iba a exigir, no lo habría empezado!) Muchas personas han revisado versiones de pequeñas partes del manuscrito y me han hecho muchas sugerencias valiosas para mejorarlo, y también a ellos quiero expresar mi agradecimiento: Toby Bailey, David Deutsch (quién 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 Newmann, Eric Penrose, Toby Penrose, Wolfgang Rindler, Engelbert Schücking y Denis Sciama. La ayuda de Christopher Penrose, con la información detallada respecto al conjunto de Mandelbrot, merece un reconocimiento especial, así como la de Jonathan Penrose, por su valiosa información acerca de los computadores que juegan al ajedrez. Gracias especiales a Colin Blakemore, Erich Harth y David Hubel por leer y revisar el Capítulo 9, que concierne a un tema en el que ciertamente no soy un experto —aunque, como sucede con todos los demás a quienes he expresado agradecimiento, ellos no son en ninguna medida responsables de los errores que puedan subsistir. Agradezco a la NSF su ayuda bajo los contratos DMS 84-05644, DMS 86-06488 y PHY 86-12424. También he contraído una gran deuda con Martin Gardner por su extrema generosidad al escribir un

prefacio para este trabajo, y también por algunos comentarios concretos. De una forma más especial, quiero expresar mi agradecimiento a mi querida Vanessa por sus atentas y detalladas críticas de varios capítulos, por su incalculable asistencia en las referencias y, no menos importante, por aguantarme cuando estaba más insoportable —y por su cariño y apoyo allí donde era vitalmente necesario.

Procedencia de las ilustraciones

Los editores han solicitado o desean agradecer el permiso para la reproducción de las ilustraciones que se citan:

Figs. 4.6 y 4.9 proceden de D. A. Klarner (ed.), *The mathematical Gardner* (Wadsworth International, 1981).

Fig. 4.7 procede de B. Grunbaum y G. C. Shephard, *Tilings and patterns* (W. H. Freeman, 1987). Copyright 1987 by W. H. Freeman and Company. Utilizada con su permiso.

Fig. 4.10 procede de K. Chandrasekharan, *Hermann Weyl, 1885-1985* (Springer, 1986).

Figs. 4.11 y 10.3 proceden de Pentaplexity: a class of non-periodic tilings of the plane. *The Mathematical Intelligencer*, 2, 32-7 (Springer, 1979).

Fig. 4.12 procede de H. S. M. Coxeter, M. Emmer, R. Penrose y M. L. Teuber (eds.), *M. C. Escher: Art and Science* (North Holland, 1986).

Fig. 5.2 © 1989, M. C. Escher Heirs/Cordon Art-Baarn-Holland.

Fig. 10.4 procede de *Journal of Materials Research*, 2, 1-4 (Materials Research Society, 1987).

Todas las demás figuras son del autor.

*Dedico este libro a la memoria de mi querida madre,
que no vivió lo suficiente para verlo.*

Prólogo

Había una numerosa concurrencia en el Gran Auditorium para la presentación del nuevo computador «Ultronic». El Presidente Pollo acababa de concluir su discurso de apertura y se alegraba de ello: no se sentía a gusto en tales ocasiones y no sabía nada de computadores, salvo que éste le iba a ahorrar mucho tiempo. Los fabricantes le habían asegurado que, entre sus muchos cometidos, sería capaz de asumir todas las delicadas decisiones de Estado que él encontraba tan fastidiosas. Mejor que así fuera, considerando la cuantiosa suma de dinero del Tesoro que había invertido en ello. Se veía ya disfrutando de muchas horas libres para jugar al golf en su magnífico campo privado —una de las pocas áreas verdes extensas que quedaban en su pequeño país.

Adam se sentía privilegiado por estar entre los asistentes a la ceremonia de inauguración. Se sentó en la tercera fila. Dos filas delante de él estaba su madre, una tecnócrata jefe que había intervenido en el diseño del Ultronic. Casualmente su padre también estaba allí —en el fondo de la sala sin estar invitado, y ahora completamente rodeado de guardias de seguridad. En el último minuto el padre de Adam había tratado de hacer estallar el computador. El mismo se había asignado esta misión, en calidad de autotitulado «espíritu conductor» de un pequeño grupo marginal de activistas: el Gran Consejo para la Consciencia Psíquica. Por supuesto, él y todos sus explosivos habían sido inmediatamente detectados por los numerosos sensores electrónicos y químicos. Como una pequeña parte de su condena, tendría que ser testigo de la ceremonia de puesta en marcha.

Adam no sentía aprecio especial por ninguno de sus padres. Quizá no necesitaba tales sentimientos. Durante sus trece años había sido

criado casi exclusivamente por computadores y con gran lujo material. Podía tener todo lo que quisiera sin más que apretar un botón: comida, bebida, compañía y entretenimiento, y también educación en cualquier cosa de la que sintiese necesidad —siempre ilustrada con coloridos y atractivos gráficos. La posición de su madre había hecho posible todo esto.

El Diseñador 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. Pero afortunadamente no necesitamos imaginarla. Dentro de un instante todos nosotros tendremos el privilegio de ser testigos de primera mano de esta inteligencia: ¡pido a la respetada Primera Dama de nuestro gran país, la señora Isabella Pollo, que conecte el interruptor que activa nuestro fantástico Computador Ultronic!»

La esposa del Presidente avanzó. Un poco nerviosa, y con cierta torpeza, conectó el interruptor. Hubo un silencio y un casi imperceptible desvanecimiento 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 Computador Ultronic para plantearle la primera pregunta?», preguntó el Diseñador Jefe. Nadie se atrevía, temerosos de parecer estúpidos ante la multitud —y ante la Nueva Omnipresencia. Se hizo el silencio. «Seguro que hay alguien», suplicó. Pero todos tenían miedo, como si presintieran una nueva y todopoderosa consciencia. Adam no sentía el mismo respeto. Desde que nació había crecido entre computadores. Casi sabía lo que le gustaría *ser* a un computador, o por lo menos así lo creía. De todas formas tenía curiosidad. Adam levantó su mano. «Ah, sí», dijo el Diseñador Jefe, «el muchacho de la tercera fila. ¿Tienes alguna pregunta para nuestro —ejem— nuevo amigo?»

1. ¿Puede tener mente un computador?

Introducción

Durante las últimas décadas la tecnología de computadores electrónicos ha hecho enormes progresos. Y lo que es más, existen pocas dudas de que en las próximas décadas tendrán lugar nuevos avances en velocidad, capacidad y diseño lógico. Nuestros computadores actuales nos parecerán tan lentos y primitivos como hoy nos parecen las calculadoras mecánicas de antaño. Hay algo casi estremecedor en el ritmo del progreso. Los computadores ya pueden realizar tareas que hasta ahora habían estado reservadas exclusivamente al pensamiento humano, con una velocidad y precisión que superan con mucho a lo que un ser humano puede lograr. Hace tiempo que nos hemos acostumbrado a la maquinaria que nos supera ampliamente en las tareas físicas. *Esto* no nos causa desasosiego. Antes bien, nos complace tener aparatos que nos llevan por tierra normalmente a grandes velocidades —más de cinco veces más rápido que el más veloz atleta humano— o que pueden cavar hoyos o demoler estructuras que nos estorban a velocidades que dejarían en ridículo a equipos compuestos por docenas de hombres. Aún estamos más encantados de tener máquinas que nos permiten hacer físicamente cosas que nunca antes habíamos podido hacer; pueden llevarnos a los cielos y depositarnos al otro lado del océano en cuestión de horas. Tales logros de su parte no hieren nuestro orgullo. Pero el poder *pensar*, eso sí ha sido siempre una prerrogativa humana. Después de todo, ha sido esa capacidad de pensar la que, al traducirse en términos físicos, nos ha permitido trascender nuestras limitaciones físicas y la que parecía ponernos por encima de nuestras criaturas hermanas. Si las máquinas pudieran lle-

gar a superarnos algún día en esa cualidad importante en la que nos habíamos creído superiores, ¿no tendríamos entonces que ceder esa superioridad a nuestras creaciones?

La pregunta de si se puede afirmar o no que un artefacto mecánico piensa —quizá incluso que experimenta sentimientos, o que tiene una mente— no es realmente nueva ¹. Sin embargo, ha recibido un nuevo impulso, incluso una demanda urgente, con la llegada de la moderna tecnología de computadores. La pregunta afecta a temas profundos de la filosofía. ¿Qué significa pensar o sentir? ¿Qué es una mente? ¿Existen realmente las mentes? Suponiendo que sí existen, ¿en qué medida dependen de las estructuras físicas a las que están asociadas? ¿Podrían existir mentes independientemente de tales estructuras? ¿O son simplemente los modos de funcionar de (ciertos tipos apropiados de) estructuras físicas? En cualquier caso, ¿es necesario que las estructuras relevantes sean de naturaleza biológica (cerebros) o podrían también estar asociadas con componentes electrónicos? ¿Están sujetas a las leyes de la física? ¿Cuáles son, de hecho, las leyes de la física?

Éstos son algunos de los temas que intentaré tratar en este libro. Pedir respuestas definitivas a preguntas tan fundamentales estaría fuera de lugar. Yo no puedo proporcionar tales respuestas: nadie puede, aunque algunos puedan tratar de impresionarnos con sus conjeturas. Mis propias conjeturas jugarán un papel importante en lo que sigue, pero trataré de distinguir claramente tales especulaciones de los hechos científicos brutos, y trataré también de dejar claras las razones que subyacen a mis especulaciones. No obstante, mi principal propósito aquí no es el de conjeturar respuestas sino el de plantear algunos temas aparentemente nuevos concernientes a la estructura de la ley física, la naturaleza de las matemáticas y del pensamiento consciente, y de presentar un punto de vista que no he visto expresado hasta ahora. Es un punto de vista que no puedo describir adecuadamente en pocas palabras, y ésta es una de las razones por las que he querido presentar las cosas en un libro de este tamaño. Pero en resumen, y quizá de manera algo equívoca, puedo al menos afirmar que mi punto de vista sugiere que es nuestra actual falta de comprensión de las leyes fundamentales de la física la que nos impide aprehender el concepto de «mente» en términos físicos o lógicos. No quiero decir con esto que las leyes nunca serán suficientemente bien conocidas. Por el contrario, parte del objetivo de esta obra es intentar estimular futura investigación en direcciones que parecen prometedoras a este respecto, y tratar de hacer algunas sugerencias bastante concretas, y apa-

rentemente nuevas, sobre el lugar que realmente podría ocupar la «mente» en el desarrollo de la física que conocemos.

Debería dejar claro que mi punto de vista es poco convencional entre los físicos y que, por consiguiente, resulta poco probable que sea adoptado, actualmente, por los científicos de computadores o los fisiólogos. La mayoría de los físicos alegarán que las leyes fundamentales que operan a la escala del cerebro humano son ya perfectamente conocidas. No se negará, por supuesto, que existen aún muchas lagunas en nuestro conocimiento de la física en general. Por ejemplo, no conocemos las leyes básicas que determinan los valores de las masas de las partículas subatómicas de la Naturaleza ni la intensidad de sus interacciones. No sabemos cómo hacer completamente compatible la teoría cuántica con la teoría de la relatividad especial de Einstein —ni mucho menos cómo construir la teoría de la «gravitación cuántica» que haría compatible la teoría cuántica con su teoría de la relatividad *general*. Como consecuencia de esto último, no comprendemos la naturaleza del espacio a la escala absurdamente minúscula de $1/100\ 000\ 000\ 000\ 000\ 000\ 000$ de la dimensión de las partículas fundamentales conocidas, aunque para dimensiones mayores que ésta nuestro conocimiento se presume adecuado. No sabemos si el Universo como un todo tiene extensión finita o infinita —tanto en el espacio como en el tiempo— aunque pueda parecer que tales incertidumbres no tienen ninguna importancia en la escala humana. No comprendemos la física que actúa en el corazón de los agujeros negros ni en el big bang origen del propio Universo. Pero todas estas cosas parecen estar lo más alejadas que podamos imaginar de la escala «cotidiana» (o una más pequeña) que es relevante para el funcionamiento de un cerebro humano. ¡Y ciertamente lo están! De todas formas, argumentaré que *precisamente* a este nivel existe otra gran incógnita en nuestra comprensión de la física, que podría ser relevante para la actuación del pensamiento humano y la consciencia —¡ante (o, mejor dicho, detrás) de nuestras propias narices! Es una incógnita que no ha sido siquiera reconocida por la mayoría de los físicos, como trataré de explicar. Argumentaré además que, de forma bastante curiosa, los agujeros negros y el big bang son consideraciones ¡que realmente *tienen* una gran relación con estos asuntos!

En lo que sigue intentaré persuadir al lector de la fuerza de la evidencia que subyace al punto de vista que trato de exponer. Pero para comprender este punto de vista tenemos un buen trabajo por delante. Necesitaremos viajar por territorios muy extraños —algunos de relevancia aparentemente dudosa— y por campos muy dispares.

Necesitaremos examinar la estructura, fundamentos y enigmas de la teoría cuántica; los rasgos básicos de las teorías de la relatividad especial y general, de los agujeros negros, el big bang, y la segunda ley de la termodinámica; y de la teoría de Maxwell de los fenómenos electromagnéticos, tanto como de las bases de la mecánica newtoniana. Algunas cuestiones de filosofía y psicología tendrán un claro papel que jugar cuando intentemos comprender la naturaleza y función de la consciencia. Por supuesto, tendremos que tener una visión general de la neurofisiología real del cerebro, además de los modelos de computador propuestos. Necesitaremos poseer alguna noción del status de la inteligencia artificial. Necesitaremos saber qué es una máquina de Turing, y comprender el significado de la computabilidad, del teorema de Gödel y de la teoría de la complejidad. Necesitaremos profundizar también en los fundamentos de la matemática, e incluso plantearnos la cuestión de la propia naturaleza de la realidad física.

Si, al final de todo ello, el lector no queda persuadido por el menos convencional de los argumentos que trato de exponer, confío al menos que ella o él habrán sacado algo de genuino valor de este tortuoso aunque, espero, fascinante viaje.

El test de Turing

Imaginemos que un nuevo modelo de computador ha salido al mercado, posiblemente con un tamaño de memoria de almacenamiento y un número de unidades lógicas mayor que las que hay en un cerebro humano. Supongamos también que las máquinas han sido cuidadosamente programadas y se les ha introducido gran cantidad de datos apropiados. Los fabricantes dicen que el dispositivo *piensa* realmente. Quizá digan también que son auténticamente inteligentes. O pueden ir más lejos y sugerir que estos aparatos realmente *sienten* —dolor, felicidad, compasión, orgullo, etc.— y que son conscientes, y que realmente *comprenden* lo que están haciendo. De hecho, parece que se esté afirmando que tienen *consciencia*.

¿Cómo decidir si dar crédito o no a las afirmaciones de los fabricantes? Normalmente, cuando compramos una determinada máquina juzgamos su valor únicamente según el servicio que nos presta. Si realiza satisfactoriamente las tareas que le encomendamos, entonces quedamos complacidos. Si no, la devolvemos para su reparación o sustitución. De acuerdo con este criterio, para probar la afirmación de los fabricantes de que un aparato semejante tiene realmente los

atributos humanos que se le atribuyen pediríamos simplemente que se *comporte* en estos aspectos como lo haría un ser humano. Siempre que lo hiciera satisfactoriamente no tendríamos ningún motivo de queja a los fabricantes y no necesitaríamos devolver el computador para su reparación o sustitución.

Esto nos proporciona un punto de vista operacional respecto a tales cuestiones. El operacionalista diría que el computador *piensa* con tal de que *actúe* de manera indistinguible de como lo hace una persona cuando está pensando. Adoptaremos, de momento, este punto de vista operacional. Esto no quiere decir, por supuesto, que estemos pidiendo que el computador se mueva como podría hacerlo una persona mientras está pensando. Menos aún esperaríamos que se asemejara o se hiciera sentir al tacto como un ser humano: estos atributos serían irrelevantes para el propósito del computador. Lo que esto quiere decir, no obstante, es que le estamos pidiendo que dé respuestas cuasi-humanas a cualquier pregunta que le podamos plantear, y estamos afirmando estar de acuerdo en que realmente piensa (o siente, comprende, etc.) siempre que responda a nuestras preguntas de una forma indistinguible de la de un ser humano.

Este punto de vista fue vigorosamente expuesto en un famoso artículo de Alan Turing, titulado «Computing Machinery and Intelligence», aparecido en 1950 en la revista filosófica *Mind* (Turing 1950). (Diremos más sobre Turing más adelante.) En este artículo se describió por primera vez la idea ahora conocida como *test de Turing*. Este pretendía ser una prueba acerca de si se puede decir razonablemente que una máquina piensa. Supongamos que se afirma realmente en un computador (como el que nos venden los fabricantes en la descripción anterior) piensa. De acuerdo con el test de Turing, el computador y algún voluntario humano se ocultan a la vista de un (perspicaz) interrogador. El interrogador tiene que tratar de decidir cuál de los dos es el computador y cuál el ser humano mediante el simple procedimiento de plantear preguntas de prueba a cada uno de ellos. Estas preguntas y, lo que es más importante, las respuestas que ella * reci-

* Al escribir un libro como éste se plantea inevitablemente el problema de decidir si utilizar el pronombre «él» o «ella» donde, por supuesto, no se pretende que el género tenga ninguna implicación. En consecuencia, cuando me refiera a una persona abstracta utilizaré a partir de ahora simplemente «él» en el sentido de «ella» o «él», que es lo que considero una práctica usual. Sin embargo, espero que se me perdone una pizca de «sexismo» al expresar en este punto una preferencia por una interrogadora femenina. ¡Mi idea es que ella podría ser más sensible que su homólogo masculino para reconocer la verdadera cualidad humana!

be, se transmiten de modo impersonal; por ejemplo, pulsadas en un teclado o mostradas en una pantalla. A la interrogadora no se le permite más información sobre cualquiera de las partes que la que obtiene en esta sesión de preguntas y respuestas. El sujeto humano responde a las preguntas sinceramente y trata de persuadirle de que él es realmente el ser humano y que el otro sujeto es el computador; pero el computador está programado para «mentir» y, por consiguiente, tratar de convencer a la interrogadora de que es *él*, y no el otro, el ser humano. Si en el curso de una serie de test semejantes la interrogadora es incapaz de identificar de una forma definitiva al sujeto humano real, se considera que el computador (o el programa del computador, o el programador, o el diseñador, etc.), ha superado la prueba.

Ahora bien, podría decirse que este test es bastante injusto con el computador. En efecto, si se invirtieran los papeles de forma que se le pidiese al ser humano que se hiciera pasar por un computador y al computador que repondiese sinceramente, sería demasiado fácil para la interrogadora descubrir cuál es cuál. Todo lo que necesitaría hacer sería pedir al sujeto que realizara alguna complicada operación aritmética. Un buen computador sería capaz de responder al instante con precisión, pero un ser humano se quedaría mudo. (Habría que tener cierto cuidado con esto, no obstante. Hay «calculadores prodigio» humanos que pueden realizar notables proezas de aritmética mental con precisión infalible y sin esfuerzo aparente. Por ejemplo, Johann Martin Zacharias Dase ², hijo de un granjero analfabeto, que vivió en Alemania entre 1824 y 1861, era capaz de multiplicar mentalmente dos números de ocho cifras en menos de un minuto, o dos números de veinte cifras ¡en unos seis minutos! Sería fácil confundir tales proezas con los cálculos de un computador. En tiempos más recientes fueron igualmente impresionantes los logros de Alexander Aitkan, que fue Catedrático de Matemáticas en la Universidad de Edimburgo en los años 50, y de algunos otros. La tarea aritmética que escogiera la interrogadora para hacer el test tendría que ser mucho más compleja que ésta —por ejemplo, multiplicar dos números de treinta cifras en dos segundos, lo que está claramente dentro de las capacidades de un buen computador moderno.)

Por consiguiente, parte de la tarea de los programadores del computador consiste en hacer que el computador parezca en ciertos aspectos «más estúpido» de lo que realmente es. Así, si la interrogadora planteara al computador una tarea aritmética complicada, como las que hemos considerado más arriba, el computador debería simular

que *no* es capaz de responderla o, de lo contrario, ¡sería descubierto inmediatamente! Pero no creo que la tarea de hacer a un computador «más estúpido» fuera un problema particularmente serio para los programadores del computador. Su dificultad principal estaría en hacerle responder a algunos de los tipos más simples de preguntas de «sentido común»: ¡preguntas con las que el sujeto humano no tendría ninguna dificultad!

No obstante, hay un problema inherente al citar ejemplos concretos de tales preguntas. Dada cualquier pregunta que uno pudiera sugerir, sería cosa fácil, a continuación, pensar una manera de hacer que el computador respondiera a esa pregunta *concreta* como lo haría una persona. Pero cualquier falta de comprensión real por parte del computador quedaría probablemente en evidencia en un interrogatorio *continuado*, y especialmente con preguntas originales y que requieran alguna comprensión real. La habilidad de la interrogadora radicaría en parte en su capacidad de imaginar estas preguntas originales, y en parte en su capacidad de hacerlas seguir de otras, de naturaleza exploratoria, diseñadas para descubrir si ha tenido o no lugar alguna «comprensión» real. Podría plantear de vez en cuando alguna pregunta completamente sin sentido para ver si el computador podía detectar la diferencia, o podría añadir una o dos que superficialmente pareciesen absurdas pero que en realidad tuviesen cierto sentido. Por ejemplo, podría decir: «Esta mañana oí que un rinoceronte iba volando por el Mississipi en un globo rosa. ¿Qué piensas de eso?» (Casi podemos imaginar las gotas de sudor frío formándose en la frente del computador —¡por usar una metáfora no muy apropiada!) Él podría responder cautelosamente: «Eso me suena bastante ridículo.» Hasta aquí todo va bien. Continúa la interrogadora: «¿De veras? Mi tío lo hizo una vez, sólo que era blanco con rayas. ¿Qué hay de ridículo en eso?» Es fácil imaginar que, si no tuviera una correcta «comprensión», un computador caería pronto en la trampa y se descubriría. Podría incluso meter la pata diciendo «Los rinocerontes no pueden volar» (si sus bancos de memoria vinieran en su ayuda con el hecho de que no tienen alas), en respuesta a la primera pregunta, o «Los rinocerontes no tienen rayas», en respuesta a la segunda. La próxima vez la interrogadora podría ensayar una pregunta realmente absurda, tal como cambiarla por «*bajo el Mississipi*», o «*dentro de un globo rosa*», o «en un *camisón rosa*», para ver si el computador tenía juicio para darse cuenta de la diferencia esencial.

Dejemos a un lado, de momento, la cuestión de si, o cuándo, puede construirse un computador que supere realmente el test de

Turing. En lugar de ello supongamos, sólo para nuestra argumentación, que ya se han construido máquinas semejantes. Podemos entonces preguntar si se puede decir que un computador, que ha superado el test, *necesariamente* piensa, siente, comprende, etc. Volveré a este asunto en breve. De momento, consideremos algunas de sus implicaciones. Por ejemplo, si los fabricantes tienen razón en sus afirmaciones más radicales, es decir, que su aparato es un ser pensante, sintiente, sensible, comprensivo, *consciente*, entonces nuestra compra del aparato nos implicará en *responsabilidades morales*. ¡Ciertamente lo *haría* si hubiéramos de creer a los fabricantes! Sería censurable el simple hecho de poner en marcha el computador para satisfacer nuestras necesidades sin tener en cuenta su propia sensibilidad. Eso no sería diferente de maltratar a un esclavo. En general, tendríamos que evitar causar al computador el dolor que los fabricantes alegan que es capaz de sentir. Desconectar el computador, o quizá incluso venderlo cuando había llegado a sentirse muy unido a nosotros, nos plantearía dificultades morales, y habría otros incontables problemas del tipo que se nos presentan en la relación con otros seres humanos o con los animales. Todas estas cuestiones serían ahora muy relevantes. Por todo ello, sería para nosotros de gran importancia (¡y también para las autoridades!) saber si las pretensiones de los fabricantes —que, suponemos, se basan en su afirmación de que

«Cada aparato pensante ha sido sometido al test de Turing por nuestro equipo de expertos»

son realmente ciertas.

Creo que, pese al absurdo aparente de algunas implicaciones de estas afirmaciones, en particular las morales, la afirmación de considerar la superación del test de Turing como una indicación válida de la presencia de pensamiento, inteligencia, comprensión o consciencia es realmente bastante fuerte; pues ¿de qué otro modo que no sea la conversación juzgamos que otras personas diferentes que nosotros poseen tales cualidades? En realidad *existen* otros criterios, tales como las expresiones faciales, movimientos corporales, y acciones en general, que nos pueden influir de forma significativa cuando hacemos tales juicios. Pero podríamos imaginar que (quizá en un futuro más lejano) se pudiera construir un robot que imitase con éxito todas estas expresiones y movimientos. Ahora ya no sería necesario ocultar al robot y al sujeto humano a la vista de la interrogadora, aunque los criterios que ésta tiene a su disposición son, en principio, los mismos que antes.

Desde mi propio punto de vista, estoy dispuesto a relajar de forma considerable los requisitos del test de Turing. Creo que pedir al computador que imite a un ser humano de una forma tan fiel que resulte indistinguible en los aspectos relevantes es, en verdad, pedir al computador más de lo necesario. Todo lo que yo pediría sería que nuestra perspicaz interrogadora se sintiera realmente convencida, por la naturaleza de las réplicas del computador, de que hay una *presencia consciente*, aunque posiblemente extraña, subyacente a estas réplicas. Esto es algo manifiestamente ausente de todos los sistemas para computadores que se han construido hasta la fecha. Me doy cuenta, sin embargo, de que existiría el peligro de que si la interrogadora fuera capaz de decidir efectivamente qué sujeto era el computador, entonces, quizá inconscientemente, podría ser reacia a atribuir una consciencia al computador aun cuando *pudiera* percibirla. O, por el contrario, ella podría tener la impresión de que «siente» esa «presencia extraña» —y estar preparada para conceder al computador el beneficio de la duda— aun cuando no la hay. Por estas razones, la versión original de Turing acerca del test tiene una ventaja considerable en su mayor objetividad, y en general me atenderé a ella en lo que sigue. La consiguiente «injusticia» hacia el computador de la que he hablado antes (es decir, que para superar el test debe ser capaz de hacer todo lo que puede hacer un ser humano, mientras que el humano no necesita ser capaz de hacer todo lo que puede hacer un computador) no es algo que parezca preocupar a los defensores del test de Turing como un verdadero test de pensamiento, etc. En cualquier caso, su opinión frecuentemente manifiesta es que no pasará mucho tiempo antes de que un computador pueda *realmente* superar el test —digamos hacia el año 2010. (Turing sugirió originalmente que para el año 2000 se podría llegar a un 30 por ciento de éxitos para el computador con un interrogador «medio» y sólo cinco minutos de interrogatorio.) Parecen convencidos, en consecuencia, de que esta falta de imparcialidad ¡no está retrasando significativamente ese día!

Todo esto resulta relevante para una cuestión esencial: ¿realmente proporciona el punto de vista operacional un conjunto de criterios razonables para juzgar la presencia o ausencia de cualidades mentales en un objeto? Algunos afirmarán firmemente que no lo hace. La imitación, por muy hábil que sea, no es necesariamente lo mismo que la cosa real. Mi posición a este respecto es en cierto modo intermedia. Como principio general me inclino a creer que la imitación, por muy hábil que sea, debería ser siempre detectable mediante un sondeo suficientemente hábil —aunque esto es más una cuestión de fe (o de

optimismo científico) que un hecho probado. Por ello estoy dispuesto a aceptar el test como aproximadamente válido en su contexto. Es decir, *si* el computador fuera capaz de responder a todas las preguntas que se le plantean de una manera indistinguible de cómo las respondería un ser humano —y, de este modo, engañar completa * y consistentemente a nuestra perspicaz interrogadora— entonces, *en ausencia de cualquier evidencia en contra*, mi *conjetura* sería que el computador realmente piensa, siente, etc. Al utilizar palabras como «evidencia», «realmente» y «conjetura» quiero decir que cuando me refiero a pensamiento, sentimiento, o comprensión, o, especialmente, a *consciencia*, considero que los conceptos significan «cosas» reales objetivas cuya presencia o ausencia en los cuerpos físicos es algo que tratamos de descubrir, ¡y que no son simplemente conveniencias de lenguaje! Considero esto un punto crucial. Al tratar de discernir la presencia de tales cualidades hacemos conjeturas basadas en toda la evidencia disponible. (Esto no es diferente en principio de, por ejemplo, un astrónomo tratando de averiguar la masa de una estrella lejana.)

¿Qué tipo de evidencia en contra tendríamos que considerar? Es difícil establecer reglas por adelantado. No obstante, quiero dejar claro que el simple hecho de que el computador pudiera esta construido a base de transistores, cables y similares en lugar de neuronas, venas, etc., *no* es, propiamente dicho, el tipo de cosas que consideraría evidencia en contra. El tipo de cosas en que estoy pensando es que en algún momento en el futuro pueda desarrollarse una teoría acertada de la consciencia —acertada en el sentido que sea una teoría física coherente y apropiada, elegantemente consistente con el resto de los conocimientos físicos, y tal que sus predicciones se correspondan exactamente con las afirmaciones de los seres humanos acerca de cuándo, si, o hasta qué punto parecen ellos mismos ser conscientes— y que esta teoría pueda tener implicaciones con respecto a la supuesta consciencia de nuestro computador. Se podría incluso imaginar un «detector de consciencias», construido según los principios de esta teoría, que fuera completamente fiable con respecto a sujetos huma-

* Estoy siendo deliberadamente cuidadoso acerca de lo que consideraría una auténtica superación del test de Turing. Puedo imaginar, por ejemplo, que tras una larga serie de fracasos en el test un computador podría reunir todas las respuestas que el sujeto humano hubiera dado previamente y emitir las de nuevo con ciertos ingredientes aleatorios apropiados. ¡Después de algún tiempo, nuestra cansada interrogadora acabaría por agotar las preguntas originales y podría ser engañada de un modo que considero «tramposo» por parte del computador!

nos pero que diera resultados diferentes a los de un test de Turing en el caso de un computador. En tales circunstancias, uno tendría que ser muy cuidadoso a la hora de interpretar los resultados de un test de Turing. Creo que la forma de ver la cuestión de la adecuación del test de Turing depende en parte de la forma en que uno espera que se desarrolle la ciencia y la tecnología. Tendremos necesidad de volver a estas consideraciones más adelante.

Inteligencia artificial

Un área de gran interés en los últimos años es la que se conoce como *inteligencia artificial*, a menudo abreviada simplemente con «IA». Los objetivos de la IA son imitar por medio de máquinas, normalmente electrónicas, tantas actividades mentales como sea posible, y quizá llegar a mejorar las capacidades humanas en estos aspectos. El interés por los resultados de la IA procede al menos de cuatro direcciones. En concreto tenemos el estudio de la *robótica*, que está interesada, en gran medida, en los requisitos prácticos de la aplicación industrial de los dispositivos mecánicos que pueden realizar tareas «inteligentes» —tareas de una variedad y complejidad que habían exigido anteriormente la intervención o control humano— y realizarlas con una velocidad y fiabilidad más allá de cualquier capacidad humana, o en condiciones adversas en las que la vida humana correría peligro. También de interés comercial, tanto como general, es el desarrollo de *sistemas expertos*, con los que se intenta codificar el conocimiento esencial de toda una profesión —medicina, abogacía, etc.— en ¡un paquete de ordenador! ¿Es posible que la experiencia y competencia de los seres humanos miembros de estas profesiones pueda ser realmente reemplazada por estos paquetes? ¿O se trata sencillamente de que todo lo que podemos esperar conseguir son largas listas de información objetiva junto con un completo sistema de referencias cruzadas? La cuestión de si los computadores pueden mostrar (o imitar) auténtica inteligencia tiene evidentemente importantes implicaciones sociales. Otra área en la que la IA podría tener relevancia directa es la *psicología*: se confía en que tratando de imitar el comportamiento de un cerebro humano (o de algún otro animal) mediante un dispositivo electrónico —o fracasando en el intento— podamos aprender algo importante sobre el funcionamiento cerebral. Finalmente, existe la esperanza optimista de que, por razones similares, la IA tuviera algo que decir sobre cuestiones profundas de la filosofía y nos

proporcionara algunas intuiciones directas sobre el significado del concepto de *mente*.

¿Hasta dónde ha sido capaz de llegar la IA por el momento? Me resultaría difícil tratar de resumirlo. Existen muchos grupos activos en diferentes partes del mundo y sólo estoy familiarizado con una pequeña parte de este trabajo. De todas formas sería correcto decir que, aunque se han hecho muchas cosas ingeniosas, la simulación de algo que pudiera pasar por auténtica inteligencia tiene aún un largo camino por delante. Para dar alguna idea del tema mencionaré primero algunos de los (todavía impresionantes) logros anteriores, y luego algunos progresos notables recientes con computadores que juegan al ajedrez.

Uno de los primeros dispositivos IA fue la «tortuga» de W. Gray Walter, construida a comienzos de los años 50³, que se movía autónomamente por el suelo hasta que sus baterías estaban bajas, y entonces iba al enchufe más próximo, se conectaba y recargaba sus baterías. Una vez totalmente reabastecida se desconectaba por sí misma del enchufe ¡y volvía a sus aventuras por el suelo! Desde entonces se han construido muchos objetos similares (ver, por ejemplo, Waltz, 1982). Una línea de desarrollo algo diferente queda ejemplificada en el programa de computador de Terry Winograd, diseñado en 1972, que puede conversar audiblemente sobre una colección de bloques de varias formas y colores que, mediante simulación, coloca unos sobre otros en diferente orden y disposición. Pese a estos tempranos éxitos es notable cuán difícil ha resultado el diseñar el control de un sencillo «brazo robot» articulado que tiene que maniobrar evitando obstáculos amontonados, en donde los movimientos necesarios parecerían «obvios» desde una perspectiva humana. La interpretación de escenas visuales es también un área en la que los procedimientos de la IA no han conseguido acercarse a lo que el cerebro humano (y, de hecho, el de la mayoría de los animales) parece lograr «sin esfuerzo».

Quizá en línea más directa con el test de Turing está el programa de ordenador de K.M. Colby, desarrollado a mediados de los 60, que simula un psicoterapeuta —con tanto éxito que algunos pacientes descubren que prefieren al computador antes que a un terapeuta humano, y están dispuestos a revelarle sobre sí mismos más de lo que lo harían a su homólogo humano. Como ejemplo⁴ de tal intercambio entre el computador «terapeuta» y un paciente, consideremos la siguiente conversación (mantenida a través de un teletipo):

Terapeuta: En qué puedo ayudarle.

Paciente: No lo sé.

Terapeuta: Trate de contarme algo.

Paciente: No.

Terapeuta: Puede explicarse.

Paciente: Hoy me encuentro enfermo y cansado.

Terapeuta: Tiene alguna idea de por qué.

Paciente: Sí, soy viejo.

Terapeuta: Por favor continúe.

Paciente: Creo que he perdido interés en todo.

Terapeuta: Dígame algo más sobre esos sentimientos.

Paciente: Tengo alguna idea de por qué me siento así.

Terapeuta: La tiene.

Paciente: Creo que echo en falta una familia.

Terapeuta: Qué le hace echar en falta una familia.

Paciente: Hay un estímulo en ello.

Terapeuta: Por qué es eso importante para usted.

Aunque esto pueda dar una extraña impresión de que el computador entiende algo, en realidad no entiende nada y simplemente está siguiendo algunas reglas mecánicas sencillas. (Existe también un caso «inverso» de éste en un sistema en el que computador simula un paciente humano esquizofrénico, que presenta todos los síntomas y respuestas de un libro de texto y es capaz de engañar a algunos estudiantes de medicina que ¡creen que es realmente un paciente humano el que está dando las respuestas!)

Los computadores que juegan al ajedrez proporcionan probablemente los mejores ejemplos de máquinas que exhiben lo que podría ser considerado como «conducta inteligente». De hecho, algunas máquinas han alcanzado hoy día (1989) un nivel de actuación extremadamente respetable con relación a los jugadores humanos —acercándose al de «Maestro Internacional». (Las puntuaciones de estos computadores estarían algo por debajo de 2.300; en comparación, el campeón mundial Kasparov tiene una puntuación por encima de 2.700.) En concreto, un programa de computador (para un microprocesador comercial Fidelity Excel) de Dan y Kathe Spracklen ha alcanzado una puntuación (Elo) de 2.110 y se le ha concedido el título de «Maestro» de la USCF. Aún más impresionante es «Deep Thought» (Pensamiento Profundo), programado fundamentalmente por Hsiung Hsu, de la Universidad de Carnegie Mellon, que tiene una puntuación en torno a 2.500 Elo y recientemente logró la notable proeza de compartir el primer puesto (con el Gran Maestro Tony Miles) en un torneo de ajedrez (en Longbeach, California, noviembre de 1988), ¡derrotando

de hecho a un Gran Maestro (Ben Larsen) por primera vez! ⁵ Los computadores que juegan al ajedrez sobresalen también en la resolución de *problemas* de ajedrez y superan fácilmente a los humanos en este empeño.

Las máquinas de jugar al ajedrez dependen en buena medida del «conocimiento libresco» además de su preciso poder de cálculo. Es digno de mención que, curiosamente, las máquinas de jugar al ajedrez son mejores en general, respecto a jugadores humanos comparables, cuando se impone que los movimientos se ejecuten muy rápidamente; los jugadores humanos actúan mucho mejor, comparados con las máquinas, cuando se permite una buena cantidad de tiempo para cada movida. Esto se puede comprender por el hecho de que las decisiones del computador se toman sobre la base de extensos cálculos rápidos y exactos, mientras que el jugador humano saca ventaja de «juicios» que descansan en evaluaciones conscientes relativamente lentas. Estos juicios humanos sirven para reducir drásticamente el número de posibilidades serias que deben considerarse en cada etapa del cálculo, y cuando se dispone de tiempo se puede hacer un análisis mucho más profundo que en el caso del mero cálculo y eliminación directa de posibilidades que hace la máquina sin utilizar tales juicios. (Esta diferencia es aún más notable en el difícil juego oriental del «Go», en el que el número de posibilidades en cada movida es mucho mayor que en el ajedrez.) La relación entre consciencia y formación de juicios será capital para mis argumentos posteriores, especialmente en el Capítulo 10.

Una aproximación IA al «placer» y al «dolor»

Una de las pretensiones de la IA es que proporciona una vía hacia un tipo de entendimiento de las cualidades mentales, tales como felicidad, dolor, hambre. Tomemos por ejemplo la tortuga de Grey Walter. Cuando sus baterías estén bajas su pauta de comportamiento cambiará, y actuará de una forma planeada para reabastecer su reserva de energía. Existen claras analogías entre ésta y la manera en que actuaría un ser humano —o cualquier otro animal— cuando sienta hambre. Quizá no sería un grave abuso de lenguaje decir que la tortuga de Grey Walter estaba «hambrienta» cuando actuaba de esta forma. Algún mecanismo interno era sensible al estado de carga de su batería, y cuando éste caía por debajo de cierto nivel orientaba a la tortuga hacia una pauta de comportamiento diferente. Sin duda

existe una operación similar en el interior de los animales cuando empiezan a tener hambre, sólo que los cambios de comportamiento son más complicados y sutiles. Más que pasar de una pauta de comportamiento a otra hay un cambio en las *tendencias* a actuar de ciertas formas, haciéndose estos cambios más fuertes (hasta un cierto punto) a medida que aumenta la necesidad de reabastecer el suministro de energía.

De modo análogo, los defensores de la IA imaginan que conceptos como el dolor o la felicidad pueden modelarse adecuadamente de esta forma. Simplifiquemos las cosas y consideremos sólo una escala de «sentimientos» que va desde el extremo «dolor» (puntuación -100) al extremo «placer» (puntuación +100). Imaginemos que tenemos un dispositivo —una máquina de algún tipo, presumiblemente electrónica— que tiene un medio de registrar su propia (supuesta) puntuación «placer-dolor», que llamaré su «puntuación-pd». El dispositivo tiene que disponer de ciertos modos de comportamiento y ciertos datos de entrada, ya sean internos (como el estado de sus baterías) o externos. La idea es que sus acciones están ajustadas para conseguir maximizar su puntuación-pd. Habría muchos factores que influyen en la puntuación-pd. Podríamos ciertamente disponer que la carga de sus baterías fuera uno de ellos, de modo que una carga baja cuente negativamente y una carga alta positivamente, pero también podría haber otros factores. Quizá nuestro dispositivo tenga algunos paneles solares que le proporcionan medios alternativos de obtener energía, de modo que no necesita hacer uso de sus baterías cuando los paneles están en operación. Podríamos disponer que al moverse hacia la luz pueda incrementar algo su puntuación-pd, de modo que, en ausencia de otros factores, eso sería lo que tendería a hacer. (¡En realidad, la tortuga de Grey Walter acostumbraba a *evitar* la luz!) Sería necesario tener algún medio de realizar cálculos para que pudiese evaluar los efectos probables que tendrían finalmente sus diferentes acciones en su puntuación-pd. Podría introducir pesos estadísticos, de modo que un cálculo tuviera un efecto mayor o menor en la puntuación dependiendo de la fiabilidad de los datos en los que se basaba.

También sería necesario proporcionar a nuestro dispositivo «objetivos» diferentes del simple mantenimiento de su suministro de energía, ya que de lo contrario no tendríamos medio de distinguir el «dolor» del «hambre». Sin duda es demasiado pedir que nuestro dispositivo tenga medios de procreación así que, de momento, ¡sexo fuera! Pero quizá podamos implantar en él un «deseo» de compañía de otro dispositivo semejante si damos a sus encuentros una puntuación-pd

positiva. O podríamos hacerle «ansioso» de aprender por su propio bien, de modo que el simple almacenamiento de hechos sobre el mundo externo tuviera también puntuación positiva de su escala-pd. (Más egoístamente, podríamos disponer que el realizar diferentes servicios para *nosotros* tuviera una puntuación positiva, ¿como tendríamos que hacer si construyéramos un sirviente robot!) Podría alegarse que hay algo artificioso en imponer tales «objetivos» en nuestro dispositivo a nuestro capricho; no obstante, esto no difiere mucho del modo en que la selección natural nos ha impuesto, como individuos, ciertos «objetivos» que están gobernados en gran medida por la necesidad de propagar nuestros genes.

Supongamos ahora que nuestro dispositivo ha sido construido con éxito con arreglo a todo esto. ¿Qué derecho tendríamos a asegurar que realmente *siente* placer cuando su puntuación-pd es positiva y dolor cuando la puntuación es negativa? El punto de vista de la IA (u operacional) diría que juzgamos esto simplemente a partir del modo en que se comporta el dispositivo. Puesto que actúa de una forma que incrementa su puntuación hasta un valor positivo tan alto como sea posible (y durante tanto tiempo como sea posible) y, correspondientemente, también actúa para evitar puntuaciones negativas, entonces podemos *definir* razonablemente su sentimiento de placer como el grado de positividad de su puntuación, y *definir* correspondientemente su sentimiento de dolor como el grado de negatividad de la puntuación. La «razonabilidad» de tal definición, se alegraría, deriva del hecho de que ésta es precisamente la forma en que reacciona un ser humano en relación con sentimientos de placer o dolor. Por supuesto que con los seres humanos las cosas no son realmente tan sencillas como esto, como ya sabemos: a veces parecemos buscar el dolor deliberadamente, o apartarnos de nuestro camino para evitar ciertos placeres. Es evidente que nuestras acciones están realmente guiadas por criterios mucho más complejos que estos (cf. Dennett, 1978, pp. 190-229). Pero, en primera aproximación, nuestra forma de actuar consiste en evitar el dolor y buscar el placer. Para un operacionalista esto sería suficiente para justificar, en el mismo nivel de aproximación, la *identificación* de la puntuación-pd de nuestro dispositivo con su valoración placer-dolor. Tales identificaciones parecen estar también entre los propósitos de la teoría de IA.

Debemos preguntar: ¿es realmente cierto que nuestro dispositivo *sentiría* dolor cuando su puntuación-pd fuera negativa y placer cuando fuera positiva? De hecho, ¿podría nuestro dispositivo sentir en absoluto? Sin duda, los operacionalistas dirán «Obviamente sí», o despa-

charán tales preguntas como sin sentido. Pero a mí me parece evidente que *existe* una cuestión seria y difícil que debe ser considerada. En nosotros mismos, las influencias que nos impulsan son de varios tipos. Algunas son conscientes, como el dolor y el placer; pero hay otras de las que no tenemos consciencia directa. Esto queda claramente ilustrado en el ejemplo de una persona que toca una estufa caliente. Se establece una acción involuntaria que le hace retirar su mano aun antes de que experimente cualquier sensación de dolor. Parecería que tales acciones involuntarias están mucho más cerca de las respuestas de nuestro dispositivo a su puntuación-pd de lo que lo están los efectos reales del dolor y el placer.

Con frecuencia utilizamos términos antropomorfos en una forma descriptiva, a menudo jocosa, para describir el comportamiento de las máquinas: «Mi coche no quería arrancar esta mañana»; o «Mi reloj aún piensa que va con la hora de California»; o «Mi ordenador afirma que no comprende la última instrucción y no sabe qué hacer a continuación». Por supuesto, no queremos decir *realmente* que el coche *quiera* algo de verdad, o que el reloj *piense*, o que el ordenador **afirme* algo o que *comprenda* o incluso sepa lo que está haciendo. De todas formas, tales proposiciones pueden ser verdaderamente descriptivas y útiles para nuestra propia comprensión, con tal de que las tomemos simplemente con el ánimo en que fueron pronunciadas y no las consideremos como aserciones literales. Adoptaré una actitud similar respecto a las diversas afirmaciones de la IA de que las cualidades mentales podrían estar presentes en los dispositivos que hemos estado construyendo —*¡independientemente* del ánimo con que se planearon! Si acepto que se diga que la tortuga de Gray Walter puede estar hambrienta, es precisamente en este sentido medio jocoso. Si estoy dispuesto a utilizar términos como «dolor» o «placer» para la puntuación-pd de un dispositivo como se concibió más arriba, es porque encuentro estos términos útiles para mi comprensión de su comportamiento, debido a ciertas analogías con mis propios comportamientos y estados mentales. No quiero decir que estas analogías sean en realidad particularmente estrechas o, de hecho, que no haya otras cosas *inconscientes* que influyen en mi comportamiento de forma mucho *más* parecida.

Confío en que queda claro para el lector que, en mi opinión, hay mucho más que entender de las cualidades mentales de lo que puede obtenerse directamente de la IA. De todas formas, creo que la IA

* ¡A partir de 1989!

plantea un caso serio que debe ser respetado y sometido a consideración. Con esto no quiero decir que se haya conseguido mucho, si se ha conseguido algo, en la simulación de la inteligencia real. Pero hay que tener en cuenta que el tema es muy joven. Los computadores serán más rápidos, tendrán mayores memorias de acceso rápido, más unidades lógicas y podrán realizar un mayor número de operaciones en paralelo. Habrá mejoras en el diseño lógico y en la técnica de programación. Estas máquinas, los vehículos de la filosofía de la IA, serán enormemente mejoradas en sus capacidades técnicas. Además, la filosofía misma *no* es intrínsecamente absurda. Quizá la inteligencia humana pueda ser aproximadamente simulada por los computadores electrónicos —esencialmente los computadores actuales, basados en principios que ya son comprendidos, pero con la capacidad, velocidad, etc., mucho mayor que van a tener en los próximos años. Quizá, incluso, estos dispositivos *serán* realmente inteligentes; quizá pensarán, sentirán y tendrán mentes. O quizá no lo hagan, y se necesite algún principio nuevo del que por el momento no hay indicios. Esto es lo que está en discusión, y es algo que no puede despacharse a la ligera. Trataré de presentar las pruebas como mejor las veo. Finalmente expondré mis propias ideas.

IA fuerte y la habitación china de Searle

Existe un punto de vista, conocido como *IA fuerte* que adopta una posición más bien extrema sobre estas cuestiones ⁶. Según la IA fuerte, no sólo los dispositivos que acabamos de mencionar serían inteligentes y tendrían mentes, etc., sino que se puede atribuir un cierto tipo de cualidades mentales al funcionamiento lógico de *cualquier* dispositivo computacional, incluso los dispositivos mecánicos más simples tales como un termostato ⁷. La idea es que la actividad mental consiste simplemente en llevar a cabo alguna secuencia bien definida de operaciones, frecuentemente llamada *algoritmo*. Más adelante precisaré más lo que realmente es un algoritmo. Por el momento, será adecuado definir simplemente un algoritmo como cierto tipo de procedimiento de cálculo. En el caso de un termostato el algoritmo es extremadamente simple: el dispositivo registra si la temperatura es mayor o menor que la establecida, y a continuación dispone que el circuito se desconecte en el primer caso y se conecte en el segundo. Para cualquier tipo importante de actividad mental de un cerebro humano el algoritmo tendría que ser muchísimo más complicado pero,

según el punto de vista de la IA fuerte, un algoritmo en cualquier caso. Diferirá enormemente en grado del sencillo algoritmo del termostato, pero no habrá diferencia de principio. Así, según la IA fuerte, la diferencia entre el funcionamiento esencial de un cerebro humano (incluyendo todas sus manifestaciones conscientes) y el de un termostato radica sólo en su mucho mayor *complicación* (o quizá «mayor orden de estructura» o «propiedades autorreferentes», o algún otro atributo que pudieramos asignar a un algoritmo) en el caso de un cerebro. Y lo que es más importante, todas las cualidades mentales —pensamiento, sentimiento, inteligencia, comprensión, consciencia— deben ser consideradas, según este punto de vista, simplemente como aspectos de este funcionamiento complicado; es decir, son simplemente características del *algoritmo* a ejecutar por el cerebro.

La virtud de cualquier algoritmo específico residiría en su actuación, es decir, en la precisión de sus resultados, su amplitud, su economía y la velocidad con que puede ser ejecutado. Un algoritmo que pretende igualar lo que se presume que está operando en un cerebro humano tendría que ser algo prodigioso. Pero si existe un algoritmo de esta especie para el cerebro —y los defensores de la IA fuerte afirmarían ciertamente que sí existe— entonces podría en principio funcionar en un ordenador. De hecho podría funcionar en *cualquier* computador electrónico moderno de tipo general si no fuera por limitaciones de espacio de almacenamiento y velocidad de operación. (La justificación de este comentario vendrá más tarde, cuando consideremos la máquina universal de Turing.) Se prevé que cualquiera de estas limitaciones quedarán superadas en los grandes y rápidos computadores de un futuro no muy lejano. En esa eventualidad, un algoritmo así, si pudiera ser encontrado, superaría presumiblemente el test de Turing. Los defensores de la IA fuerte alegarán que, donde quiera que funcione, el algoritmo experimentará *autónomamente* sentimientos, tendrá una consciencia, será una mente.

No todos estarán de acuerdo, ni mucho menos, en que se puedan identificar así los estados mentales con los algoritmos. En particular, el filósofo americano John Searle (1980, 1987) se ha opuesto vigorosamente a esta idea. Él ha citado ejemplos en los que versiones simplificadas del test de Turing han sido *ya* superadas efectivamente por un computador adecuadamente programado, pero da fuertes argumentos que apoyan el punto de vista de que el atributo mental relevante de la «comprensión» está, en cualquier caso, totalmente ausente. Uno de estos ejemplos se basa en un programa de ordenador diseñado por Roger Schank (Schank y Abelson 1977). El propósito

del programa es proporcionar una simulación de la comprensión de historias sencillas como: «Un hombre entró en un restaurante y pidió una hamburguesa. Cuando llegó la hamburguesa estaba socarrada y el hombre salió vociferando furiosamente del restaurante, sin pagar la cuenta ni dejar propina». Un segundo ejemplo: «Un hombre entró en un restaurante y pidió una hamburguesa; cuando la hamburguesa llegó, le gustó mucho; y al salir del restaurante dio una buena propina al camarero antes de pagar su cuenta». Como un test de «comprensión» de las historias se le pregunta al computador si el hombre comió la hamburguesa en cada caso (un hecho que no se había mencionado explícitamente en ninguna de las dos historias). Para este tipo de historia sencilla y pregunta sencilla el computador puede dar respuestas que son esencialmente indistinguibles de las respuestas que daría un ser humano de habla española: «no», en el primer caso, y «sí», en el segundo, para estos ejemplos particulares. Entonces, en este sentido *muy* limitado, ¡la máquina ya ha superado un test de Turing!

La cuestión que debemos considerar es si un éxito de este tipo indica realmente una auténtica comprensión por parte del computador —o, quizá, por parte del propio programa. El argumento de Searle en favor de que *no* lo hace consiste en invocar su concepto de una «habitación china». En primer lugar, él imagina que las historias se cuentan en chino en lugar de español —ciertamente un cambio no esencial— y que todas las operaciones del algoritmo del ordenador para este ejercicio concreto se suministran (en español) como un conjunto de instrucciones para manipular fichas con símbolos chinos en ellas. Searle se imagina a *sí mismo* haciendo todas las manipulaciones en el interior de una habitación cerrada. Las secuencias de símbolos que representan las historias, y luego las preguntas, se introducen en la habitación a través de una pequeña ranura. No se permite ninguna otra información del exterior. Finalmente, cuando se han completado todas las manipulaciones, la secuencia resultante se extrae de nuevo a través de la ranura. Puesto que todas estas manipulaciones están simplemente ejecutando el algoritmo del programa de Schank, debe ocurrir que esta secuencia final resultante es simplemente el equivalente chino para «sí» o «no», según sea el caso, que da la respuesta correcta a la pregunta original en chino acerca de una historia en chino. Entonces Searle deja en claro que él no entiende una sola palabra de chino, de modo que no tiene la más remota idea de lo que cuentan las historias. De todas formas, llevando a cabo correctamente la serie de operaciones que constituyen el algoritmo de Schank (las instrucciones para este algoritmo le han sido dadas en español)

él sería capaz de contestar tan bien como lo haría una persona china que realmente entendiera las historias. El punto importante de Searle —y pienso que tiene bastante fuerza— es que la mera ejecución de un algoritmo correcto *no* implica en sí mismo que haya tenido lugar ninguna comprensión. El (imaginado) Searle, encerrado en su habitación china, ¡no comprenderá una sola palabra de ninguna de estas historias!

Se han levantado diversas objeciones contra el argumento de Searle *. Mencionaré sólo las que considero de seria importancia. En primer lugar, hay quizá algo más bien confuso en la frase «no entiendo una sola palabra» que se utilizó antes. La comprensión tiene tanto que ver con estructuras como con palabras individuales. Mientras ejecuta algoritmos de este tipo, uno podría perfectamente empezar a percibir algo de la estructura que forman los símbolos sin comprender el significado real de muchos de los símbolos individuales. Por ejemplo, el carácter chino para «hamburguesa» (si en verdad existe tal cosa) podría ser reemplazado por el de cualquier otro plato, pongamos por caso «chow mein», y las historias no se verían seriamente afectadas. De todas formas, me parece razonable suponer que (incluso considerando que tales reemplazamientos sean importantes) muy poco de los significados reales de las historias se concretaría efectivamente si simplemente se continuasen ejecutando los detalles de tal algoritmo.

En segundo lugar, debemos tener en cuenta el hecho de que incluso la ejecución de un programa de ordenador bastante sencillo sería normalmente algo extraordinariamente largo y tedioso si fuera realizada por seres humanos manipulando símbolos. (¡Después de todo, por esto tenemos ordenadores que hacen esas cosas para nosotros!) Si Searle tuviera que ejecutar realmente el algoritmo de Schank de la forma sugerida, necesitaría probablemente muchos días, meses o años de trabajo extremadamente molesto para responder sólo a una sencilla pregunta —¡lo que no es una actividad del todo plausible para un filósofo! Sin embargo, ésta no me parece una seria objeción puesto que aquí estamos interesados en cuestiones *de principio* y no en cuestiones prácticas. La dificultad es mayor con un supuesto programa de ordenador que se supone tiene suficiente complicación para igualar a un cerebro humano y, por lo tanto, superar *perfectamente* el test de

* El lector de habla española encontrará una discusión más reciente de las ideas de Searle en el número de «Investigación y Ciencia» correspondiente a marzo de 1990. (Traducción de «Scientific American», January 1990). [Nota del traductor.]

Turing. Cualquier programa semejante tendría que ser terriblemente complicado. Podemos imaginar que la operación de este programa, para dar la respuesta a incluso alguna pregunta bastante simple del test de Turing, podría suponer tantos pasos que no habría posibilidad de que ningún ser humano llevara a cabo el algoritmo a mano en el plazo de tiempo de una vida humana. En ausencia de tal programa es difícil decir si será así realmente⁸. Pero, en cualquier caso, esta cuestión de la extrema complicación no puede, en mi opinión, ser simplemente pasada por alto. Es cierto que aquí estamos interesados en cuestiones de principio, pero no me resulta inconcebible que pudiera haber algún grado «crítico» de complicación que fuera necesario alcanzar en un algoritmo para que éste muestre cualidades mentales. Quizá este valor crítico es tan grande que ningún algoritmo, hasta tal punto complicado, pudiera concebiblemente ser ejecutado a mano por ningún ser humano a la manera imaginada por Searle.

El propio Searle ha replicado a esta última objeción permitiendo que todo un equipo de seres humanos manipuladores de símbolos y que no-hablan-chino reemplace al anterior habitante único («él mismo») de su habitación china. Para tener números bastante altos él imagina incluso su habitación reemplazada por toda la India, con toda su población (¡excluyendo los que entienden chino!) ocupados ahora en la manipulación de símbolos. Aunque esto sería absurdo en la práctica, no es absurdo *en principio*, y el argumento es esencialmente el mismo que antes: los manipuladores de símbolos *no* comprenden la historia, pese a la afirmación de la IA fuerte de que la simple ejecución del algoritmo apropiado daría lugar a la cualidad mental de la «comprensión». Sin embargo, otra cuestión comienza ahora a cobrar importancia. ¿No son todos estos indios individuales más semejantes a las neuronas individuales del cerebro de una persona que al cerebro global? Nadie sugerirá que las neuronas, cuyas descargas constituyen aparentemente la actividad física de un cerebro en el acto de pensar, comprendan *ellas mismas* individualmente lo que la persona está pensando, así que ¿por qué esperar que los indios individuales comprendan las historias chinas? Searle replica a esta sugerencia señalando el aparente absurdo de la India, el país real, comprendiendo una historia que ninguno de sus habitantes individuales comprende. Un país, aduce, igual que un termostato o un automóvil, no está dedicado al «negocio de la comprensión» mientras que una persona individual sí lo está.

Este argumento tiene mucha menos fuerza que el anterior. Pienso que el argumento de Searle adquiere su mayor fuerza cuando hay sólo

un individuo ejecutando el algoritmo y restringimos nuestra atención al caso de un algoritmo que es suficientemente poco complicado para que una persona lo ejecute en un tiempo inferior a una vida humana. No considero que este argumento establezca *rigurosamente* que no hay ningún tipo de «entendimiento» incorpóreo asociado con las personas que ejecutan el algoritmo, y cuya presencia no choca en modo alguno con sus propias conciencias. Sin embargo, coincido con Searle en que esta posibilidad se ha hecho, cuando menos, bastante implausible. Pienso que el argumento de Searle tiene una fuerza considerable aunque no sea completamente concluyente. Es convincente sobre todo al demostrar que algoritmos con el tipo de complicación del programa de ordenador de Schank no pueden tener ninguna verdadera comprensión de cualquiera de las tareas que ejecutan: también *sugiere* (aunque sólo eso) que ningún algoritmo, por muy complicado que sea, puede por sí mismo incorporar alguna vez verdadera comprensión —en contraste con las tesis de la IA fuerte.

Existen, por lo que puedo ver, otras dificultades muy serias con el punto de vista de la IA fuerte. Según la IA fuerte, lo que cuenta es simplemente el algoritmo. No supone ninguna diferencia el que el algoritmo sea ejecutado por un cerebro, un computador electrónico, una nación entera de indios, un dispositivo mecánico de ruedas y engranajes, o un sistema de tuberías. La idea consiste en que es simplemente la estructura lógica del algoritmo lo significativo para el «estado mental» que se supone representa, siendo completamente irrelevante la encarnación física de dicho algoritmo. Como apunta Searle, esto entraña de hecho una forma de «dualismo». El *dualismo* es un punto de vista filosófico adoptado por el muy influyente filósofo y matemático del siglo XVII René Descartes, y afirma que hay dos tipos de sustancia separados: «sustancia mental» y materia ordinaria. El que uno de estos tipos de sustancia pueda o no afectar al otro, o de qué modo pueda hacerlo, es una cuestión adicional. El punto importante es que se supone que la sustancia mental no está compuesta de materia y puede existir independientemente de ella. La sustancia mental de la IA fuerte es la estructura lógica de un algoritmo. Como acabo de señalar, la encarnación física concreta de un algoritmo es algo totalmente irrelevante. El algoritmo tiene un tipo de «existencia» incorpórea que es ajena a cualquier realización de dicho algoritmo en términos físicos. Hasta qué punto debemos considerar seriamente este tipo de existencia es una cuestión sobre la que volveré en el próximo capítulo: es parte del problema general de la realidad platónica de los objetos matemáticos abstractos. Por el momento dejaré de lado este

tema general e indicaré simplemente que los defensores de la IA fuerte parecen estar tomando seriamente la realidad de los algoritmos ya que ellos creen que los «algoritmos» forman la substancia de sus pensamientos, sus sentimientos, su entendimiento, sus percepciones conscientes. Resulta curiosamente irónico, como Searle ha señalado, el hecho de que el punto de vista de la IA fuerte parezca llevarnos a una forma extrema de dualismo: ¡precisamente el punto de vista con el que menos desearían estar asociados los defensores de la IA fuerte!

El dilema yace entre bastidores en un argumento desarrollado por Douglas Hofstadter (1981) —un defensor importante de la IA fuerte— en un diálogo titulado «Una conversación con el cerebro de Einstein». Hofstadter imagina un libro, de proporciones absurdamente monstruosas, que se supone contiene una completa descripción del cerebro de Albert Einstein. Cualquier pregunta que uno pudiera plantear a Einstein podría ser respondida, exactamente igual que lo hubiera hecho Einstein en vida, simplemente hojeando el libro y siguiendo cuidadosamente todas las instrucciones detalladas que proporciona. Por supuesto, «simplemente» es una palabra totalmente inadecuada, como Hofstadter se cuida en señalar. Pero su tesis es que *en principio* el libro es exactamente equivalente, en el sentido operacional del test de Turing, a una ridículamente ralentizada versión del Einstein real. Así, según las opiniones de la IA fuerte, el libro pensaría, sentiría, entendería, sería consciente, justo como si fuera el propio Einstein, pero quizá viviendo a un ritmo tremendamente ralentizado (de modo que para el Einstein-libro el mundo externo parecería discurrir como una exhalación a un ritmo ridículamente acelerado). De hecho, ya que se supone que el libro es simplemente una particular encarnación del algoritmo que constituía el «yo» de Einstein, realmente *sería* Einstein.

Pero ahora se presenta una nueva dificultad. El libro podría no ser abierto nunca o, por el contrario, podría ser examinado continuamente por innumerables estudiantes e investigadores en pos de la verdad. ¿Cómo sabría el libro la diferencia? Tal vez el libro no necesitara ser abierto si su información fuera recuperada mediante tomografía de rayos X o cualquier otro prodigio técnico. ¿Estaría activa la consciencia de Einstein sólo cuando el libro está siendo examinado de esta forma? ¿Sería él dos veces consciente si dos personas decidiesen plantear al libro la misma pregunta en dos momentos diferentes? ¿O ello implicaría dos casos separados y totalmente diferentes del *mismo* estado de consciencia de Einstein? ¿Quizá esta consciencia se activa sólo si se *altera* el libro? Después de todo, cuando somos cons-

cientes de algo recibimos información del mundo externo que afecta a nuestras memorias, y los estados de nuestras mentes cambian ligeramente. Si es así, ¿significa esto que son los *cambios* (apropiados) en los algoritmos (y aquí incluyo la memoria de almacenamiento como parte del algoritmo) los que deben ser asociados con sucesos mentales en lugar de (o quizás además de) la *activación* de los algoritmos? ¿O permanecería el Einstein-libro completamente autoconsciente aun si nunca fuera examinado o perturbado por nadie o nada? Hofstadter toca alguna de estas cuestiones, pero no intenta responderlas ni llegar a una conclusión en la mayoría de ellas.

¿Qué significa activar un algoritmo, o encarnarlo en forma física? ¿Habría alguna diferencia entre cambiar un algoritmo y simplemente descartar un algoritmo y reemplazarlo por otro? ¿Qué demonios tiene todo esto que ver con nuestras sensaciones de consciencia? El lector (a menos que él o ella misma sea defensor de la IA fuerte) puede estar preguntándose por qué he dedicado tanto espacio a una idea tan patentemente absurda. En realidad, yo *no* considero la idea intrínsecamente absurda: ¡sólo especialmente errónea! De hecho, hay que reconocer cierta fuerza en el razonamiento que subyace a la IA fuerte, y esto es lo que trataré de explicar. Hay también, en mi opinión, un cierto atractivo en algunas de las ideas —si se modifican adecuadamente— como también trataré de mostrar. Además, en mi opinión, el punto de vista contrario expresado por Searle también contiene serias dificultades y absurdos aparentes, pero aun así ¡estoy en buena parte de acuerdo con él!

Searle, en su discusión, parece estar aceptando implícitamente que computadores electrónicos del tipo de los actuales, aunque con una velocidad de acción y un tamaño de memoria de acceso rápido considerablemente aumentados (y posiblemente con acción paralela), pueden ser perfectamente capaces de superar limpiamente el test de Turing en un futuro no muy lejano. Está dispuesto a aceptar la idea de la IA fuerte (y de muchos otros puntos de vista «científicos») de que «somos ejemplos concretos de cierto número de programas de ordenador». Además, él cede ante: «Por supuesto el cerebro es un computador digital. Puesto que cualquier cosa es un computador digital, los cerebros también lo son.»⁹ Searle mantiene que la diferencia entre la función de los cerebros humanos (que pueden tener mentes) y de los computadores electrónicos (que, según él, no pueden tenerlas), ambos de los cuales podrían estar ejecutando el mismo algoritmo, radica solamente en la construcción material de cada uno. Afirma, aunque por razones que no es capaz de explicar, que los objetos

biológicos (cerebros) pueden tener «intencionalidad» y «semántica», que él considera como características definitorias de la actividad mental, mientras que los electrónicos no pueden tenerlas. En sí mismo esto no me parece señalar el camino hacia una teoría de la mente científicamente útil. ¿Qué hay tan especial en los sistemas biológicos, aparte quizá de la forma «histórica» en que han evolucionado (y el hecho de que *nosotros* seamos uno de esos sistemas), que les discrimina como los únicos objetos a los que se permite alcanzar intencionalidad o semántica? La tesis se me presenta con el carácter sospechoso de una afirmación dogmática, ¡quizá no menos dogmática incluso que las afirmaciones de la IA fuerte que mantienen que la simple activación de un algoritmo puede producir un estado de conciencia!

En mi opinión Searle, y muchas otras personas, han sido confundidos por la gente de los computadores. Y ellos, a su vez, han sido confundidos por los físicos. (No es culpa de los físicos. ¡Ni siquiera ellos lo saben todo!) Parece ser una creencia muy extendida la de que «cualquier cosa es un computador digital». Mi intención en este libro es tratar de demostrar por qué, y quizá cómo, esto *no* tiene que ser así necesariamente.

Hardware y software

En la jerga de la ciencia de computadores se utiliza el término *hardware* para designar la maquinaria real de un ordenador (circuitos impresos, transistores, cables, memorias magnéticas, etc.), incluyendo la especificación completa del modo en que todo esta interconectado. En correspondencia, el término *software* se refiere a los diversos programas que pueden ser ejecutados en la máquina. Uno de los notables descubrimientos de Turing fue el de que, en realidad, cualquier máquina para la que el hardware ha alcanzado un cierto grado de complicación y flexibilidad, es *equivalente* a cualquier otra máquina semejante. Esta equivalencia debe tomarse en el sentido de que para dos de estas máquinas, A y B, existirá un software específico tal que si se le proporcionara a la máquina A, ésta actuaría exactamente como si fuera la máquina B; del mismo modo, existiría otro software que haría que la máquina B actuara exactamente como la máquina A. Utilizó aquí la palabra «exactamente» en referencia a la salida (output) real de las máquinas para cualquier entrada (input) dada (introducida después de que se haya introducido el software convertidor) y

no al tiempo que podría tardar cada máquina en producir el resultado de salida. También estoy admitiendo que, si una cualquiera de las máquinas agota en un momento dado el espacio de almacenamiento para sus cálculos, puede recurrir a algún suministro externo (en principio ilimitado) de «papel de borrador» en blanco —que podría tomar la forma de un cinta magnética, discos, tambores o cualquier otra cosa. Ciertamente, la diferencia de tiempo que necesitan las máquinas A y B para ejecutar alguna tarea podría ser una consideración importante. Pudiera ser el caso, por ejemplo, de que A sea más de mil veces más rápida que B en realizar alguna tarea particular. Podría darse también el caso de que, para las mismas máquinas, hay alguna otra tarea para la que B es mil veces más rápida que A. Además, estos tiempos podrían depender en gran medida de las elecciones concretas de software convertidor que se utilice. Ésta es sobre todo una discusión «de principios», en la que no estamos realmente interesados en cuestiones prácticas como la de conseguir los cálculos en un tiempo razonable. Precisaré más en la próxima sección los conceptos que utilizo aquí: las máquinas A y B son ejemplos de las llamadas *máquinas universales de Turing*.

En realidad, todos los modernos computadores de uso general son máquinas universales de Turing. Por lo tanto, todos los computadores de uso general son equivalentes en el sentido anterior: las diferencias entre ellos pueden ser subsumidas en el software, siempre que no estemos interesados en las diferencias en la velocidad de operación resultante y las posibles limitaciones de memoria. De hecho, la tecnología moderna ha hecho posible que los computadores actúen con tanta velocidad y con tan grandes capacidades de almacenamiento que, para la mayoría de los usos «cotidianos», ninguna de estas consideraciones prácticas representa limitación alguna seria para lo que se requiere normalmente *, de modo que esta equivalencia teórica entre computadores puede verse también a un nivel práctico. ¡La tecnología ha transformado, así lo parece, las discusiones totalmente académicas acerca de los dispositivos idealizados de computación en cuestiones que afectan directamente a nuestras vidas!

Por lo que puedo apreciar, uno de los factores más interesantes que subyace a la filosofía de la IA fuerte es esta equivalencia entre dispositivos de computación físicos. El hardware se considera realmente de poca importancia (quizá incluso totalmente irrelevante) y el

* Véase, sin embargo, la discusión sobre la teoría de la complejidad y los problemas NP al final del Capítulo 4.

software, esto es, el programa, o los algoritmos, se considera un ingrediente vital. Sin embargo, me parece que hay otros factores importantes subyacentes que proceden de la dirección de la física. Trataré de dar alguna indicación de cuáles son estos factores.

¿Qué es lo que da su identidad a una persona individual? ¿Son, hasta cierto punto, los mismos átomos que componen su cuerpo? ¿Depende su identidad de la particular elección de electrones, protones y otras partículas que componen estos átomos? Hay al menos dos razones por las que no puede ser así. En primer lugar, hay una continua renovación en el material del cuerpo de cualquier persona viva. Esto se aplica en particular a las células del cerebro de una persona, pese al hecho de que no se producen células cerebrales después del nacimiento. La inmensa mayoría de los átomos en cada célula viva (incluyendo cada célula del cerebro) —y, de hecho, virtualmente todo el material de nuestros cuerpos— han sido reemplazados muchas veces desde el nacimiento.

La segunda razón procede de la mecánica cuántica —y por una extraña ironía está, estrictamente hablando, en contradicción con la primera! Según la mecánica cuántica (y veremos más sobre esto en el Capítulo 6, p. 352) dos electrones cualesquiera deben ser por necesidad totalmente idénticos, y lo mismo sucede para dos protones y para dos partículas cualesquiera de cualquier tipo específico. Esto no quiere decir simplemente que no hay manera de distinguir las partículas: el enunciado es mucho más fuerte que todo eso. Si se intercambiara un electrón en el cerebro de una persona con un electrón en un ladrillo, entonces el estado del sistema sería *¡exactamente*¹⁰ *el mismo estado que antes*, y no simplemente indistinguible de él! Lo mismo sucede para protones o cualquier otro tipo de partícula, y para átomos enteros, moléculas, etc. Si todo el contenido material de una persona fuera intercambiado con las correspondientes partículas en los ladrillos de su casa entonces, en un sentido fuerte, no habría sucedido nada en absoluto. Lo que diferencia a la persona de su casa es la *pauta* con que están dispuestos su constituyentes, y no la individualidad de los propios constituyentes.

Hay en el nivel cotidiano algo quizá análogo a esto, que es independiente de la mecánica cuántica pero que se me hizo especialmente manifiesto cuando escribía esto gracias a la tecnología electrónica que me posibilita teclear en un procesador de textos. Si quiero cambiar una palabra, transformar, pongamos por caso, «casa» en «cosa», puedo hacerlo simplemente reemplazando la «a» por una «o», o puedo decidir en su lugar teclear de nuevo toda la palabra. Si hago esto

último, ¿es la «c» la misma «c» que antes, o la he reemplazado por una idéntica? ¿Qué sucede con la «s»? Incluso si simplemente reemplazo «a» por «o», en lugar de reescribir la palabra, hay un instante, justo entre la desaparición de «a» y la aparición de «o», cuando se cierra el hueco y hay (o, al menos, lo hay a veces) una onda de realineamiento de la página a medida que la colocación de cada letra sucesiva (incluyendo la «s») es re-calculada, y luego re-re-calculada cuando se inserta la «o». (¡Qué barato resulta el cálculo improvisado en esta edad moderna!) En cualquier caso, *todas* las letras que veo ante mí en la pantalla son simples huecos en el trazado de un haz electrónico que explora toda la pantalla cincuenta veces por segundo. Si tomo una letra cualquiera y la reemplazo por una idéntica, ¿la situación es la *misma* después del reemplazamiento, o es simplemente indistinguible de ella? Tratar de adoptar el segundo punto de vista (es decir, «simplemente indistinguible») como diferente del primero (es decir, «la misma») parece una pérdida de tiempo. Cuando menos parece razonable decir que la situación es la misma cuando las letras son las mismas. Y así sucede con la mecánica cuántica de partículas idénticas. Reemplazar una partícula por otra idéntica es no haber cambiado para nada el estado. La situación debe ser considerada la *misma* que antes. (Sin embargo, como veremos en el Capítulo 6, la diferencia *no* es nada trivial en un contexto mecano-cuántico.)

Los comentarios anteriores sobre la continua renovación de los átomos en el cuerpo de una persona se hicieron en el contexto de la física clásica más que de la cuántica. Los comentarios se expresaron como si tuviera algún significado mantener la individualidad de cada átomo. De hecho la física clásica es adecuada y no andamos totalmente errados, a este nivel de descripción, en considerar los átomos como objetos individuales. Siempre que los átomos estén razonablemente bien separados de sus homólogos idénticos cuando se mueven, *podemos* referirnos coherentemente a ellos como si mantuvieran sus identidades individuales, puesto que cada átomo puede ser rastreado continuamente de modo que podríamos pensar que no perdemos de vista a cada uno de ellos por separado. Desde el punto de vista de la mecánica cuántica sería sólo una conveniencia de lenguaje el referirnos a la individualidad de los átomos, pero es una descripción bastante consistente al nivel considerado.

Aceptemos que la individualidad de una persona no tiene nada que ver con ninguna individualidad que pudiéramos tratar de atribuir a sus constituyentes materiales. En lugar de ello, debe tener que ver con la *configuración*, en cierto sentido, de dichos constituyentes —di-

gamos la configuración espacial o espacio-temporal. (Más tarde diremos más sobre esto.) Pero los defensores de la IA fuerte van mucho más lejos. Ellos dirán que si el contenido de información de tal configuración puede ser traducido a otra forma desde la cual puede ser de nuevo recuperado el original, entonces la individualidad de la persona debe permanecer intacta. Es como las series de letras que acabo de teclear y veo ahora mostradas en la pantalla de mi procesador de textos. Si las borro de la pantalla todavía permanecen codificadas en forma de ciertos minúsculos desplazamientos de carga eléctrica, en una configuración que no tiene una forma geométrica parecida en modo alguno a las letras que acabo de teclear. Pero puedo devolverlas en cualquier momento a la pantalla y están allí, como si no hubiese tenido lugar ninguna transformación. Si decido guardar lo que acabo de escribir, puedo entonces transferir la información de las secuencias de letras a configuraciones de la magnetización de un disco que puedo extraer, y al desconectar a continuación la máquina neutralizo todos los (relevantes) minúsculos desplazamientos de carga en ella. Mañana puedo reinsertar el disco, restablecer los pequeños desplazamientos de carga y mostrar de nuevo la secuencia de letras en la pantalla, como si nada hubiera pasado. Para los defensores de la IA fuerte es «evidente» que la individualidad de una persona puede ser tratada del mismo modo. Estas personas dirán que, al igual que sucede con la secuencia de letras en mi pantalla, no se pierde nada de la individualidad de una persona —de hecho, nada en absoluto le habría sucedido— si su forma física hubiera sido traducida en algo bastante diferente, por ejemplo en campos de magnetización en un bloque de hierro. Incluso parecen querer decir que la consciencia de una persona persistiría mientras la «información» de la persona esté en otra forma. Desde este punto de vista, una «consciencia de persona» debe considerarse, en realidad, como un elemento de software, y su manifestación particular en un ser humano material debe considerarse como la ejecución de ese software por el hardware de su cerebro y su cuerpo.

Parece que la razón para estas afirmaciones es que, cualquiera que sea la forma material que tome el hardware —por ejemplo, algún dispositivo electrónico— se «plantearía» siempre cuestiones de software (a la manera del test de Turing) y estas respuestas serían idénticas a las que hubiera dado la persona en su estado normal, si suponemos que el hardware funciona satisfactoriamente al computar las respuestas a estas cuestiones. («¿Cómo te sientes esta mañana?» «Oh, muy bien, gracias, aunque tengo un ligero dolor de cabeza.» «No sientes, entonces, que hay... ejem... algo extraño en tu identidad per-

sonal... o algo parecido?» «No; ¿por qué dices eso? Me parece una pregunta bastante extraña para hacer?» «¿Entonces te sientes la misma persona que eras ayer?» «¡Por supuesto que sí!»)

Una idea frecuentemente discutida en este tipo de contexto es la *máquina de teleportación* de la ciencia ficción ¹¹. Se propone como un medio de «transporte» de, pongamos por caso, un planeta a otro, pero el objeto de la discusión es si realmente existirá tal máquina. En lugar de ser transportado físicamente por una nave espacial en la forma «normal», el hipotético viajero es explorado de arriba a abajo, registrando con todo detalle la localización exacta y la especificación completa de cada átomo y cada electrón de su cuerpo. Toda esta información es entonces emitida (a la velocidad de la luz), mediante una señal electromagnética, al lejano planeta de destino previsto. Allí, la información es recogida o utilizada como instrucciones para ensamblar un duplicado exacto del viajero junto con sus recuerdos, sus intenciones, sus esperanzas y sus sentimientos más profundos. Al menos eso es lo que se espera, pues cada detalle del estado de su cerebro ha sido fielmente registrado, transmitido y reconstruido. Suponiendo que el mecanismo ha funcionado, la copia original del viajero puede ser destruida «sin peligro». Por supuesto, la pregunta es: ¿es éste *realmente* un método de viajar de un lugar a otro, o es simplemente la construcción de un duplicado y el asesinato del original? ¿Estaría usted dispuesto a utilizar este método de «viaje» —suponiendo que el método se ha mostrado completamente fiable, dentro de sus límites previstos? Si la teleportación *no* es viajar, entonces ¿cuál es la diferencia *de principio* entre esto y caminar simplemente de una habitación a otra? En este último caso ¿no están los átomos de uno en un momento dado proporcionando simplemente la información para las localizaciones de los átomos en el instante siguiente? Después de todo, hemos visto que no hay significado en preservar la identidad de cualquier átomo particular. Ni siquiera tiene significado la cuestión de la identidad de cualquier átomo particular. ¿No constituye cualquier estructura de átomos en movimiento simplemente un tipo de onda de información que se propaga de un lugar a otro? ¿Dónde está la diferencia esencial entre la propagación de la onda que describe a nuestro viajero caminando normalmente de una habitación a otra y la que tiene lugar en el dispositivo teleportador?

Supongamos que es cierto que la teleportación «funciona» realmente, en el sentido de que la propia «consciencia» del viajero es reanimada en la copia de sí mismo en el planeta lejano (suponiendo que esta cuestión tenga un verdadero significado). ¿Qué sucedería si

la copia *original* del viajero no fuera destruida, como requieren las reglas del juego? ¿Estaría su «consciencia» en dos lugares a la vez? (Trate de imaginar su respuesta cuando le dicen lo siguiente: «¡Oh Dios mío!, ¿de modo que el efecto de la droga que le suministramos antes de colocarle en el Teleportador ha desaparecido prematuramente? Esto es un poco desafortunado, pero no importa. De todos modos le gustará saber que el otro usted —ejem, quiero decir el usted *real*, esto es— ha llegado a salvo a Venus, de modo que podemos, ejem, disponer de usted —ejem, quiero decir de la copia *redundante* que hay aquí. Será, por supuesto, totalmente indoloro.») La situación tiene un aire de paradoja. ¿Hay algo en las leyes de la física que pudiera hacer la teleportación imposible *en principio*? Quizá, en efecto, no haya nada en principio contra el hecho de transmitir una persona, y la consciencia de una persona, por tales medios; en tal caso, el proceso de «copiado» ¿destruiría inevitablemente el original? ¿Pudiera ser entonces que lo que es imposible en principio es el conservar *dos* copias viables? Creo que, a pesar de la extravagante naturaleza de estas consideraciones, *hay* quizá algo significativo que se puede extraer de ellas respecto a la naturaleza física de la consciencia y la individualidad. Creo que proporciona un indicador que señala un cierto papel esencial de la *mecánica cuántica* en la comprensión de los fenómenos mentales. Pero me estoy adelantando. Será necesario volver a estas cuestiones después de que hayamos examinado la estructura de la teoría cuántica en el Capítulo 6 (cf. p., 341).

Veamos qué relación tiene el punto de vista de la IA fuerte con la cuestión de la teleportación. Supongamos que en algún lugar entre los dos planetas hay una estación repetidora en la que se almacena temporalmente la información antes de ser retransmitida a su destino final. Por conveniencia, esta información no es almacenada en forma humana sino en algún dispositivo magnético o electrónico. ¿Estaría presente la «consciencia» del viajero en asociación con este dispositivo? Los defensores de la IA fuerte tendrán que hacernos creer que así debe ser. Después de todo, dicen ellos, cualquier pregunta que decidiésemos plantear al viajero podría ser respondida en principio por el dispositivo, «simplemente» estableciendo una simulación de la actividad apropiada de su cerebro. El dispositivo contendría toda la información necesaria; y el resto sería un asunto de computación. Puesto que el dispositivo respondería a todas las preguntas exactamente como si fuera el viajero, entonces (¡test de Turing!) *sería* el viajero. Esto nos lleva de nuevo a la concepción de la IA fuerte según la cual el hardware real no es importante con respecto a los fenóme-

nos mentales. Esta opinión me parece injustificada. Se basa en la presunción de que el cerebro (o la mente) es, en efecto, un computador digital. Supone que cuando pensamos no está siendo invocado ningún fenómeno físico concreto que pudiera requerir las estructuras físicas concretas (biológicas, químicas) que tienen realmente los cerebros.

Se alegrará sin duda (desde el punto de vista de la IA fuerte) que la única suposición que se está haciendo es que los efectos de cualquier fenómeno físico concreto que necesita ser considerado pueden siempre ser exactamente *modelizados* mediante cálculos digitales. Estoy totalmente seguro de que la mayoría de los físicos aducirían que ésta es una suposición natural en la que basar nuestra comprensión actual de la física. Presentaré las razones para mi propio punto de vista contrario en los últimos capítulos (donde también necesitaré preparar el terreno para decir por qué creo que hay siquiera una suposición apreciable que hacer). Pero, por el momento, aceptemos este punto de vista (mantenido normalmente) de que toda la física relevante *puede* ser siempre modelizada mediante cálculos digitales. Entonces, la única suposición real (aparte cuestiones de tiempo y espacio para cálculo) es la «operacional» de que si algo *actúa* completamente como una entidad consciente, entonces se debe también mantener que ese algo en sí mismo «siente» ser esa entidad.

El punto de vista de la IA fuerte sostiene que, tratándose «sólo» de una cuestión de hardware, cualquier física que deba ser invocada para el funcionamiento del cerebro puede ser simulada necesariamente mediante la introducción del software convertidor apropiado. Si aceptamos el punto de vista operacional, la cuestión descansa entonces en la equivalencia de las máquinas universales de Turing y en el hecho de que cualquier algoritmo puede ser ejecutado por tal máquina —junto con la presunción de que el cerebro actúa según algún tipo de acción algorítmica. Ha llegado el momento de ser más explícito sobre estos intrigantes e importantes conceptos.

2. Algoritmos y máquinas de Turing

Fundamentos del concepto de algoritmo

¿Qué es exactamente un algoritmo, o una máquina de Turing, o una máquina universal de Turing? ¿Por qué estos conceptos son tan cruciales para el punto de vista moderno acerca de lo que podría constituir un «dispositivo pensante»? ¿Existe alguna limitación absoluta para lo que un algoritmo pueda, en principio, hacer? Para tratar adecuadamente estas cuestiones necesitaremos examinar con cierto detalle la idea de algoritmo y de máquina de Turing.

En las discusiones que siguen tendré que utilizar en ocasiones expresiones matemáticas. Sé muy bien que algunos lectores pueden quedar desanimados con estas cosas o que tal vez se asusten ante ellas. ¡Si usted es uno de esos lectores pido su indulgencia y le recomiendo seguir el consejo que di en mi «Nota para el lector» en la página 15! Los argumentos que se dan aquí no requieren conocimientos matemáticos por encima de un nivel de escuela elemental, pero para seguirlos en detalle será necesaria una seria reflexión. De hecho, la mayor parte de las descripciones son bastante explícitas y se puede llegar a una buena comprensión siguiendo los detalles. Pero también se puede conseguir bastante si uno prescinde de los pasos detallados para sacar simplemente la idea general. Si, por el contrario, usted es un experto, pido de nuevo su indulgencia. Sospecho que aún le merecerá la pena hojear lo que tengo que decir, y quizá una o dos cosas atrapen su interés.

La palabra «algoritmo» procede del nombre del matemático persa del siglo IX Abu Ja'far Mohammed ibn Mûsa *al-Khowârizm*, autor de un influyente texto matemático, alrededor del año 825 d.C., titulado

«Kitab al jabr w'al-muqabala». El que hoy día se escriba «algoritmo», en lugar de la antigua y más aproximada palabra «algorism», parece debido a una asociación con la palabra «aritmética». (Es también digno de mención que la palabra «álgebra» procede del árabe «al jabr» que aparece en el título de su libro.)

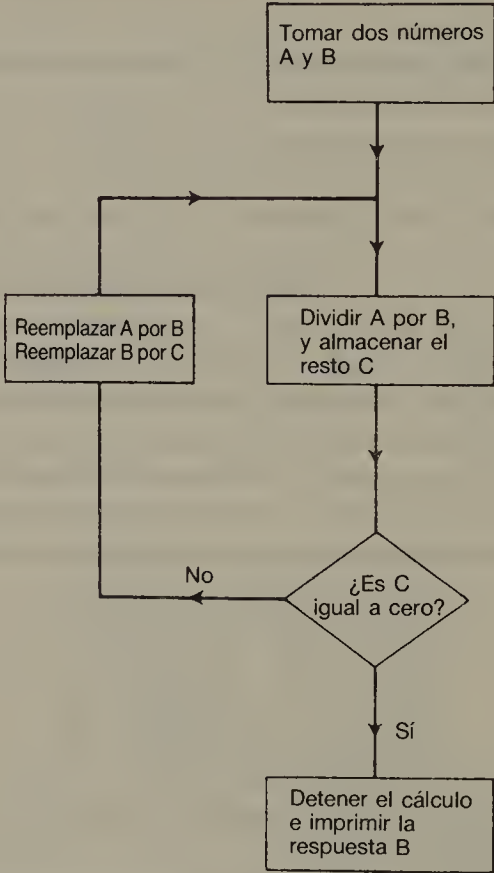
Sin embargo, mucho antes del libro de al-Khowârizm ya se conocían ejemplos de algoritmos. Uno de los más familiares, que data de la época griega (c. 300 a.C.) es el procedimiento hoy conocido como *algoritmo de Euclides* para encontrar el máximo común divisor de dos números. Veamos cómo funciona. Nos ayudará el considerar un par concreto de números, por ejemplo 1.365 y 3.654. El máximo común divisor es el mayor número entero que es divisor exacto de estos dos números. Para aplicar el algoritmo de Euclides dividimos uno de los dos números por el otro y tomamos el resto: 3.654 entre 1.365 da 2 con resto 924 ($= 3.654 - 2.730$). Ahora reemplazamos nuestro par de números originales por su resto, a saber 924, y el número por el que habíamos dividido, a saber 1.365, en este orden. Repetimos la operación que acabamos de hacer utilizando ahora este nuevo par de números: 1.365 dividido entre 924 da 1 con resto 441. Esto nos da un nuevo par, 441 y 924, y entonces dividimos 924 entre 441 obteniendo el resto 42 ($= 924 - 882$), y así sucesivamente hasta llegar a una división exacta. Escribiendo en orden todo esto tenemos:

$$\begin{array}{ll} 3.654 \div 1.365 & \text{da resto } 924 \\ 1.365 \div 924 & \text{da resto } 441 \\ 924 \div 441 & \text{da resto } 42 \\ 441 \div 42 & \text{da resto } 21 \\ 42 \div 21 & \text{da resto } 0 \end{array}$$

El último número por el que dividimos, es decir 21, es el máximo común divisor buscado.

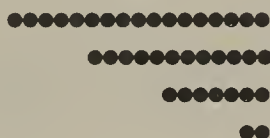
El algoritmo de Euclides es el *procedimiento sistemático* mediante el que encontramos este divisor. Hemos aplicado este procedimiento a un par de números particular, pero el procedimiento en sí se aplica con toda generalidad a números de cualquier magnitud. Para números muy grandes el procedimiento puede tardar bastante tiempo, y cuanto mayores sean los números más tiempo necesitará. Pero en cualquier caso *concreto* el procedimiento llegará a su final y se obtendrá una respuesta definida en un número finito de pasos. En cada paso está perfectamente claro qué operación debe realizarse, y también es perfectamente clara la decisión de cuándo debe darse por terminado todo el proceso. Además, la descripción del proceso total puede presentar-

se en términos finitos, a pesar de que se aplique a números naturales de tamaño ilimitado. (Los «números naturales» son simplemente los ordinarios no-negativos ¹ de números 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,...) De hecho, es fácil construir un «organigrama» (finito) para describir la operación lógica completa del algoritmo de Euclides.



Debería señalarse que este procedimiento no ha sido descompuesto todavía en sus partes más elementales, pues hemos supuesto implícitamente que ya «sabemos» cómo efectuar la necesaria operación básica para obtener el resto de una división entre dos números naturales **A** y **B** arbitrarios. Dicha operación es también algorítmica —realizada mediante el familiar procedimiento de la división que aprendemos en la escuela—. Este procedimiento es realmente bastante más complicado que el resto del algoritmo de Euclides pero de nuevo podemos construir un organigrama. La mayor complicación resulta del hecho de que (presumiblemente) utilizaríamos la notación «deci-

mal» estándar para los números naturales, de modo que necesitaríamos tener una lista de toda la tabla de multiplicar y preocuparnos de hacer estas mutiplicaciones, etc. Si hubiéramos utilizado simplemente una serie de n marcas de algún tipo para representar el número n —por ejemplo, ●●●●● para representar el cinco— entonces la formación de un resto se vería como una operación algorítmica muy elemental. Para obtener el resto de **A** dividido entre **B** borramos simplemente la serie de marcas que representa **B** de la serie que representa **A** hasta que ya no queden bastantes marcas para realizar de nuevo la operación. La serie de marcas que quede finalmente proporciona la respuesta buscada. Por ejemplo, para obtener el resto de la división de diecisiete entre cinco procedemos simplemente a borrar series de ●●●●● de la serie ●●●●●●●●●●●●●●●● de la forma siguiente:

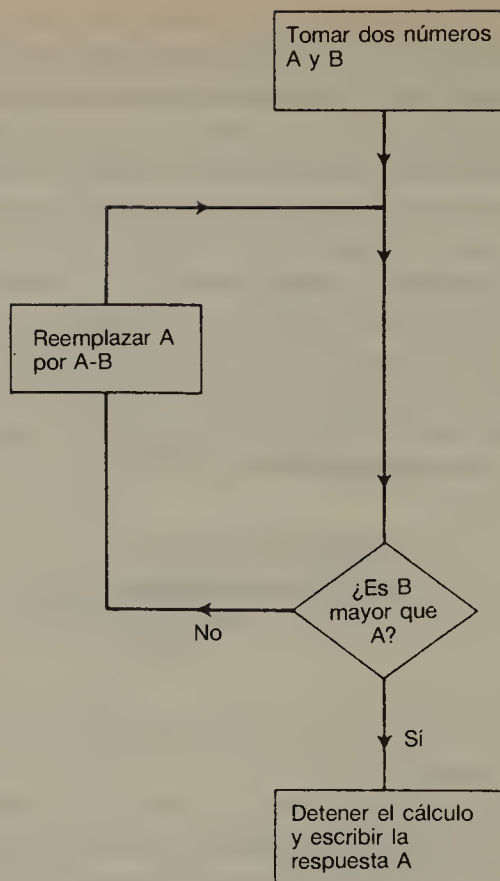


y evidentemente la respuesta es dos puesto que ya no podemos realizar la operación otra vez.

Un organigrama que encuentra el resto de una división, por medio de esta substracción repetida, puede ser el de la página siguiente.

Para completar el organigrama para el algoritmo de Euclides sustituimos el diagrama anterior para la formación del resto en la caja central derecha de nuestro organigrama original. Este tipo de sustitución de un algoritmo dentro de otro es un procedimiento normal en la programación de computadores. El algoritmo anterior para hallar un resto es un ejemplo de *subrutina*, es decir, un algoritmo (normalmente conocido con anterioridad) al que se llama y es utilizado por el algoritmo principal como parte de su operación.

Por supuesto, la representación del número n simplemente como n puntos resulta muy poco eficaz cuando intervienen números muy grandes, razón por la que normalmente utilizamos una notación más compacta como lo es la notación estándar (decimal). Sin embargo, aquí no estamos demasiado interesados por la *eficacia* de las operaciones o las notaciones. Más bien estamos interesados en la cuestión de acerca de qué operaciones pueden realizarse *en principio* algorítmicamente. Lo que es algorítmico si utilizamos una notación para los números es también algorítmico si utilizamos la otra. La única diferencia reside en el detalle y la complicación en los dos casos.



El algoritmo de Euclides es sólo uno entre los numerosos, a menudo clásicos, procedimientos algorítmicos que se encontrarán por todas las matemáticas. Pero es quizá destacable que, a pesar de los antiguos orígenes históricos de ejemplos concretos de algoritmos, la formulación precisa del *concepto general de algoritmo* data sólo de *este* siglo. De hecho se han dado varias descripciones alternativas de este concepto, todas ellas en los años 30. La más directa y convincente, y también la más importante históricamente, es en términos del concepto conocido como *máquina de Turing*. Será conveniente para nosotros examinar estas «máquinas» con algún detalle.

Una cosa a tener en cuenta a propósito de una «máquina» de Turing es que ésta es un elemento de «matemática abstracta» y no un objeto físico. El concepto fue introducido en 1935-1936 por el matemático inglés, extraordinario descifrador y pionero de los computadores, Alan Turing (Turing, 1937) para tratar un problema de muy

amplio alcance, conocido como el *Entscheidungsproblem*, parcialmente planteado por el gran matemático David Hilbert en 1900, en el Congreso Internacional de Matemáticos en París («décimo problema de Hilbert») y, de forma más completa, en el Congreso Internacional de Bolonia en 1928. Hilbert había pedido nada menos que un procedimiento algorítmico general para resolver cuestiones matemáticas —o, mejor dicho, una respuesta a la cuestión de si semejante procedimiento podía o no existir en principio. Hilbert tenía también un programa para situar las matemáticas sobre una base inatacable, con axiomas y reglas que quedaran establecidas de una vez por todas, pero para cuando Turing produjo su gran obra dicho programa ya había sufrido un golpe decisivo por parte de un sorprendente teorema demostrado en 1931 por el brillante lógico austriaco Kurt Gödel. Consideramos el teorema de Gödel y su significado en el Capítulo 4. El problema de Hilbert que interesaba a Turing (el *Entscheidungsproblem*) iba más allá de cualquier formulación concreta de las matemáticas en términos de sistemas axiomáticos. La pregunta era: ¿existe algún procedimiento mecánico general que pueda, *en principio*, resolver uno tras otro todos los problemas de las matemáticas (que pertenezcan a alguna clase apropiadamente bien definida)?

Parte de la dificultad para resolver esta cuestión consistía en decidir lo que se debe entender por «procedimiento mecánico». El concepto quedaba fuera de las ideas matemáticas comunes de la época. Para poder manejarse, Turing trató de imaginar cómo podría formalizarse el concepto de «máquina», descomponiendo su modo de operar en términos elementales. Parece claro que también Turing consideraba el cerebro humano como un ejemplo de «máquina» en este sentido, de modo que, cualesquiera que fueran las actividades que pudiera llevar a cabo un matemático humano cuando aborda sus problemas de matemáticas, éstas también tendrían que entrar en la etiqueta de «procedimientos mecánicos».

Aunque esta visión del pensamiento humano parece haber sido valiosa a Turing a la hora de desarrollar este concepto capital, no estamos obligados ni mucho menos a adherirnos a ella. En realidad, al precisar lo que se debía entender por procedimiento mecánico, Turing realmente mostró que existen algunas operaciones matemáticas perfectamente bien definidas que ¡no pueden ser llamadas mecánicas en ningún sentido ordinario! Hay quizá cierta ironía en el hecho de que este aspecto de la propia obra de Turing puede proporcionarnos ahora indirectamente una posible escapatoria a su propio punto de vista respecto a la naturaleza de los fenómenos mentales. Sin em-

bargo, esto no nos interesa por el momento. Primero necesitamos descubrir cuál es realmente el concepto de Turing de un procedimiento mecánico.

El concepto de Turing

Tratemos de imaginar un dispositivo para llevar a cabo un procedimiento de cálculo (definible en términos finitos). ¿Qué forma general tendría tal dispositivo? Debemos estar dispuestos a idealizar un poco y no preocuparnos demasiado por cuestiones prácticas: estamos pensando en una «máquina» idealizada matemáticamente. Queremos que nuestro dispositivo tenga un conjunto discreto de estados posibles diferentes, en número *finito* (aunque puede ser un número enorme). Éstos son los estados *internos* del dispositivo. Sin embargo no queremos limitar el tamaño de los cálculos que nuestro dispositivo puede realizar en principio. Recordemos el algoritmo de Euclides descrito anteriormente; en principio no hay límite para la magnitud de los números sobre los que el algoritmo puede actuar. El algoritmo —o el *procedimiento* general de cálculo— es el mismo independientemente de la magnitud de los números. Para números muy grandes el procedimiento puede tardar mucho tiempo y necesitar una gran cantidad de «papel de borrador» en el que realizar las operaciones. Pero el *algoritmo* es el mismo conjunto *finito* de instrucciones independientemente de lo grandes que sean los números.

Así, aunque tenga un número finito de estados internos, nuestro dispositivo debe poder manejar un *input* que no esté restringido en tamaño. Además, se le permitirá al dispositivo hacer uso de un espacio ilimitado de almacenamiento externo (nuestro «papel de borrador») para sus cálculos, y poder también producir un *output* de tamaño ilimitado. Puesto que nuestro dispositivo tiene sólo un número finito de estados internos distintos no cabe esperar «internar» todos los datos externos ni todos los resultados de sus propios cálculos. En lugar de ello deberá examinar sólo aquellas partes de los datos o cálculos previos que está manejando *inmediatamente*, y realizar entonces cualquier operación que sea necesario hacer con éstos. Puede anotar, quizá en el espacio de almacenamiento externo, los resultados relevantes de esta operación, y luego pasar de una manera exactamente determinada a la etapa siguiente de la operación. Es la naturaleza ilimitada del input, el espacio de cálculo y el output lo que nos dice que estamos considerando sólo una idealización matemática

en lugar de algo que pudiera ser realmente construido en la práctica (ver Fig. 2.1). Pero es una idealización de gran importancia. Las maravillas de la moderna tecnología de computadores nos han proporcionado dispositivos de almacenamiento electrónico que pueden considerarse como ilimitados para muchos propósitos prácticos.

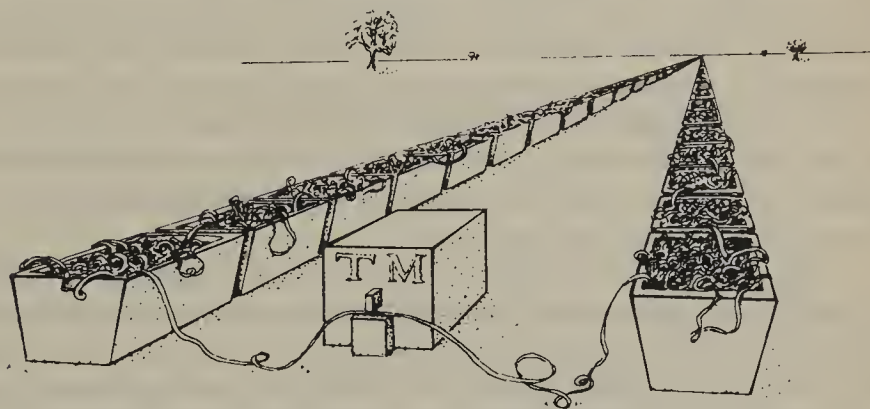


Fig. 2.1. ¡Una máquina de Turing estricta requiere una cinta infinita!

De hecho, el tipo de espacio de almacenamiento que hemos llamado «externo» en la discusión anterior podría ser considerado como una parte real del funcionamiento interno de un computador moderno. Es quizá una precisión técnica el que cierta parte del espacio de almacenamiento pueda considerarse como interna o externa. Una manera de referirse a esta división entre el «dispositivo» y la parte «externa» podría ser en términos de *hardware* y *software*. La parte interna sería entonces el hardware, y la parte externa, el software. No me atenderé necesariamente a esto pero, de cualquier forma que se considere, los computadores electrónicos actuales se aproximan de forma realmente notable a la idealización de Turing.

Turing representaba los datos externos y el espacio de almacenamiento en forma de una «cinta» sobre la que se hacen marcas. Esta cinta sería utilizada por el dispositivo y «leída» cuando fuera necesario, y el dispositivo podría, como parte de la operación, mover la cinta hacia delante o hacia atrás. También podría hacer nuevas marcas en los lugares de la cinta donde fuera necesario y podría borrar las viejas, permitiendo actuar a la *misma* cinta como almacenamiento externo (es decir, «papel de borrador») y como input. De hecho, resulta conveniente no establecer una clara distinción entre «almacenamiento

externo» e «input», ya que en muchas operaciones los resultados intermedios de un cálculo pueden jugar el papel de nuevos datos. Recuérdese que en el algoritmo de Euclides reemplazábamos nuestro input original (los números **A** y **B**) por los resultados de las diferentes etapas de cálculo. Análogamente, la misma cinta puede ser utilizada para el output final (es decir, la «respuesta»). La cinta seguirá pasando por el dispositivo hacia delante y hacia atrás mientras sea necesario hacer nuevos cálculos. Cuando el cálculo esté completo, el dispositivo se detiene y la respuesta se muestra en la parte de la cinta que queda a un lado del dispositivo. Supongamos, por ser concretos, que la respuesta se muestra siempre a la izquierda, mientras que todos los datos numéricos del input, junto con la especificación del problema a resolver, siempre quedan a la derecha.

Yo, personalmente, encuentro algo incómodo pensar en nuestro dispositivo finito que mueve hacia adelante y atrás a una cinta potencialmente infinita. ¡Por muy ligero que sea el material de que está hecha, una cinta *infinita* debería ser difícil de mover! En su lugar, prefiero pensar en la cinta como representación de un entorno a través del cual puede moverse nuestro dispositivo finito. (Por supuesto que con la electrónica moderna ni la «cinta» ni el «dispositivo» tienen que «moverse» realmente en el sentido físico ordinario, pero tal «movimiento» es una manera conveniente de representar las cosas.) Desde este punto de vista, el dispositivo recibe todo el input desde su entorno. Utiliza el entorno como su «papel de borrador». Y finalmente escribe su output en este mismo entorno.

En la imagen de Turing la «cinta» es una secuencia lineal de cuadrados que se considera infinita en ambas direcciones. Cada cuadrado de la cinta está en blanco o contiene una sola marca *. El uso de cuadrados marcados o sin marcar ilustra el hecho de que estamos admitiendo que nuestro entorno (es decir, la cinta) pueda ser descompuesta y descrita en términos de elementos *discretos* (en oposición a continuos). Esto parece razonable si queremos que nuestro dispositivo funcione de un modo fiable y perfectamente definido. Sin embargo, estamos admitiendo que este «entorno» sea (potencialmente) infinito como un rasgo de la idealización matemática que estamos

* En realidad, en su descripción original Turing permitía que su cinta estuviera marcada de maneras más complicadas, pero esto no supone ninguna diferencia real. Las marcas complejas podrían descomponerse siempre en series de marcas y espacios en blanco. Me tomaré otras libertades sin importancia respecto a las especificaciones originales de Turing.

utilizando, pero en cualquier caso *particular* el input, el cálculo y el output deben ser siempre *finitos*. De este modo, aunque la cinta se considera infinitamente larga, en ella debe haber sólo un número finito de marcas reales. Más allá de un cierto punto en cada dirección la cinta debe estar completamente en blanco.

Indicaremos un cuadrado en blanco mediante el símbolo «0» y uno marcado mediante el símbolo «1», v.g.:

..... 0 0 0 1 1 1 1 0 1 0 0 1 1 1 0 0 1 0 0 1 0 1 1 0 1 0 0

Necesitamos que nuestro dispositivo «lea» la cinta, y supondremos que lo hace *de cuadrado en cuadrado*, y después de cada operación se mueve sólo *un cuadrado más* a derecha o izquierda. No hay ninguna pérdida de generalidad en esto. Un dispositivo que leyera *n* cuadrados cada vez, o que se moviera *k* cuadrados cada vez, puede modelarse fácilmente mediante otro dispositivo que sólo lee y se mueve de uno en uno. A partir de *k* movimientos de un cuadrado puede construirse un movimiento de *k* cuadrados, y almacenando *n* lecturas de un cuadrado se comportaría como si leyera *n* cuadrados a un tiempo.

¿Qué puede hacer, en concreto, un dispositivo semejante? ¿Cuál es la forma más general en la que algo que describiríamos como «mecánico» podría funcionar? Recordemos que los *estados internos* de nuestro dispositivo son finitos en número. Todo lo que necesitamos saber, más allá de esta finitud, es que el comportamiento del dispositivo está completamente determinado por su estado interno y por el input. Hemos simplificado este input haciendo que sólo sea uno de los dos símbolos «0» ó «1». Dado este input y su estado inicial, el dispositivo actúa de una forma completamente determinista: cambia su estado interno a algún otro (o quizá el mismo) estado interno; reemplaza el 0 ó 1 que acaba de leer por el mismo o distinto símbolo 0 ó 1; se mueve un cuadrado hacia la derecha o la izquierda; finalmente, decide si continuar el cálculo o terminarlo y detenerse.

Para definir las operaciones de nuestro dispositivo de modo explícito *numeramos* en primer lugar los diferentes estados internos, por ejemplo, mediante las etiquetas 0, 1, 2, 3, 4, 5,...; la operación del dispositivo, o *máquina de Turing*, estará entonces totalmente especificada por alguna lista explícita de reemplazamientos, tal como:

00 → 00D
 01 → 13L

10	→	651D
11	→	10D
20	→	01D.STOP
21	→	661I
30	→	370D
.		.
.		.
.		.
2100	→	31I
.		.
.		.
.		.
2580	→	00D.STOP
2590.	→	971D
2591	→	00D.STOP

La cifra escrita en *grandes* caracteres en el lado izquierdo de la flecha es el símbolo que el dispositivo está actualmente leyendo sobre la cinta, y el dispositivo lo reemplaza por la cifra en grandes caracteres que figura en medio del lado derecho. D nos dice que el dispositivo va a moverse un cuadrado hacia la *derecha* a lo largo de la cinta e I nos dice que va a moverse un paso hacia la *izquierda*. (Si, como sucede en la descripción original de Turing, consideramos que lo que se mueve es la cinta en lugar del dispositivo, entonces debemos interpretar D como la instrucción para mover la *cinta* un cuadrado hacia la *izquierda* e I la instrucción para moverla a la *derecha*.) La palabra STOP indica que el cálculo ha sido completado y el dispositivo se detendrá. En concreto, la segunda instrucción 01 → 131I nos dice que si el dispositivo está en el estado interno 0 y lee 1 en la cinta, entonces debe cambiar al estado interno 13, dejar el 1 en la cinta sin cambio y moverse un cuadrado hacia la izquierda a lo largo de la cinta. La última instrucción 2591 → 00D.STOP nos dice que si el dispositivo está en el estado 259 y lee 1 en la cinta, entonces debe volver al estado 0, borrar el 1 y escribir un 0 en la cinta, moverse un cuadrado hacia la derecha a lo largo de la cinta y dar por terminado el cálculo.

En lugar de utilizar los caracteres numéricos 0, 1, 2, 3, 4, 5,... para etiquetar los estados internos, quedaría más en consonancia con la notación para las marcas en la cinta el utilizar símbolos contruidos sólo a base de 0s y 1s. Podríamos utilizar, si quisiéramos, simplemente una serie de n símbolos 1 para etiquetar al estado n , pero eso es poco

eficaz. En lugar de ello, utilicemos el sistema *binario* de numeración que actualmente resulta un modo familiar de notación:

0	→	0,
1	→	1,
2	→	10,
3	→	11,
4	→	100,
5	→	101,
6	→	110,
7	→	111,
8	→	1000,
9	→	1001,
10	→	1010,
11	→	1011,
12	→	1100, etc.

Aquí, el dígito final de la derecha se refiere a las «unidades» igual que en la notación estándar (decimal), pero el dígito inmediatamente anterior se refiere a «doses» en lugar de «dieces». El anterior se refiere a «cuatros» en lugar de «cientos» y el anterior a éste se refiere a «ochos» en lugar de «miles», y así sucesivamente, siendo el valor de cada dígito sucesivo, a medida que nos movemos hacia la izquierda, las sucesivas *potencias de dos*: 1, 2, 4(=2 × 2), 8(=2 × 2 × 2), 16(=2 × 2 × 2 × 2), 32(2 × 2 × 2 × 2 × 2), etc. (Para otros propósitos que surgirán posteriormente nos será también útil usar alguna otra base diferente de dos o de diez para representar los números naturales: por ejemplo, en base *tres*, el número decimal 64 se escribiría 2101, pues el valor de cada dígito es ahora una potencia de tres: 64=(2 × 3³) + 3² + 1; cf. Capítulo 4, p. 145, nota a pie de página.)

Utilizando la notación binaria para los estados internos, la especificación de la máquina de Turing anterior será ahora:

00	→	00D,
01	→	11011D,
10	→	10000011I
11	→	10D
100	→	01STOP
101	→	10000101I
110	→	1001010D

. . .
. . .
. . .

110100100 → 111

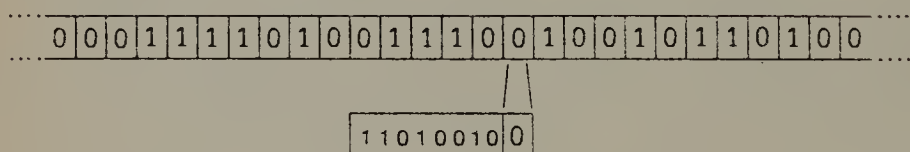
1000000101 → 00STOP

1000000110 → 11000011D

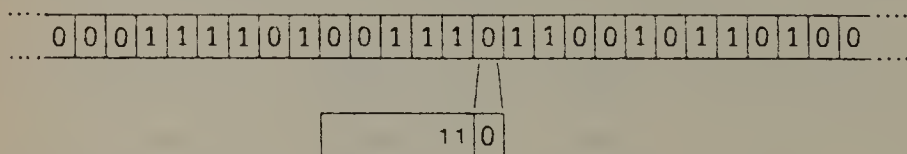
1000000111 → 00STOP

En esto último también he abreviado D.STOP por STOP, puesto que también podemos suponer que I.STOP nunca ocurre, de modo que el resultado del último paso del cálculo se muestra a la izquierda del dispositivo como parte de la respuesta.

Supongamos que nuestro dispositivo se encuentra en el estado interno representado por la secuencia binaria 11010010 y está a mitad de un cálculo para el que la cinta viene dada como en la p. 65, y apliquemos la instrucción 110100100 → 111:



El dígito de la cinta que está siendo leído (en este caso el dígito «0») está indicado en caracteres más grandes a la derecha de la cadena de símbolos que representa el estado interno. En el caso de una máquina de Turing como la que parcialmente se especifica arriba (y que yo he formado más o menos al azar), el «0» que está siendo leído será reemplazado por un «1» y el estado interno cambiará a «11»; luego el dispositivo se moverá un paso a la izquierda:



El dispositivo está ahora listo para leer otro dígito, de nuevo un «0». De acuerdo con la tabla, ahora deja este «0» inalterado, pero reemplaza el estado interno por «100101» y retrocede a lo largo de la cinta un paso hacia la derecha. Ahora lee «1», y en algún lugar más abajo en la tabla habrá una nueva instrucción sobre cómo debe reemplazar

su estado interno, si debe cambiar el dígito que está leyendo, y en qué dirección debe moverse. Continuará así hasta que llegue a un STOP, punto en el cual (tras moverse un paso más hacia la derecha) imaginamos una campana que suena para avisar al operador de que el cálculo ha concluido.

Supondremos que la máquina comienza siempre con estado interno «0» y que toda la cinta que queda a la izquierda del dispositivo de lectura está inicialmente en blanco. Todas las instrucciones y datos se introducen en la derecha. Como ya mencionamos, esta información que se introduce toma siempre la forma de una cadena *finita* de 0s y 1s, seguidos por cinta en blanco (es decir, todos 0s). Cuando la máquina llega a STOP, el resultado del cálculo aparece en la cinta a la izquierda del dispositivo de lectura.

Puesto que queremos poder incluir datos numéricos como parte de nuestro input, nos gustaría también tener una manera de describir números ordinarios (entendiendo aquí por ello los números naturales 0, 1, 2, 3, 4,...) como parte del input. Una manera de hacerlo sería simplemente usando una cadena de n símbolos 1 para representar el número n (aunque esto podría plantearnos alguna dificultad con el número natural cero):

$$1 \rightarrow 1, 2 \rightarrow 11, 3 \rightarrow 111, 4 \rightarrow 1111, 5 \rightarrow 11111, \text{ etc.}$$

Este primitivo sistema de numeración se llama (de forma no muy lógica) sistema *unario*. Entonces el símbolo «0» podría utilizarse como espacio para separar números diferentes. Es importante que dispongamos de este medio de separar números ya que muchos algoritmos actúan sobre *conjuntos* de números y no ya sobre un solo número. Por ejemplo, para el algoritmo de Euclides nuestro dispositivo necesitará actuar sobre el *par* de números A y B. Se pueden apuntar, sin gran dificultad, máquinas de Turing que ejecutan este algoritmo. Como ejercicio, los lectores curiosos podrían dedicarse a verificar que la siguiente descripción explícita de una máquina de Turing (que llamaré *EUC*) ejecuta realmente el algoritmo de Euclides cuando se aplica a un par de números unarios separados por un 0:

$$\begin{aligned} 00 &\rightarrow 00D, & 01 &\rightarrow 11I, & 10 &\rightarrow 101D, & 11 &\rightarrow 11I, & 100 &\rightarrow 10100D, \\ 101 &\rightarrow 110D, & 110 &\rightarrow 1000D, & 111 &\rightarrow 111D, & 1000 &\rightarrow 1000D, \\ 1001 &\rightarrow 1010D, & 1010 &\rightarrow 1110I, & 1011 &\rightarrow 1101I, & 1100 &\rightarrow 1100I, \\ 1101 &\rightarrow 11I, & 1110 &\rightarrow 1110I, & 1111 &\rightarrow 10001I, \\ 10000 &\rightarrow 10010I, & 10001 &\rightarrow 10001I, & 10010 &\rightarrow 100D, & 10011 &\rightarrow 11I, \\ 10100 &\rightarrow 00STOP, & 10101 &\rightarrow 10101D. \end{aligned}$$

Antes de embarcarse en esto, sin embargo, sería prudente que el lector empezara por algo mucho más simple, como la máquina de Turing $UN + 1$:

$$00 \rightarrow 00D, \quad 01 \rightarrow 11D, \quad 10 \rightarrow 01STOP, \quad 11 \rightarrow 11D.$$

que simplemente suma uno a un número unario. Para comprobar que $UN + 1$ hace precisamente eso, imaginemos que se aplica, por ejemplo, a la cinta

...00000111100000...,

que representa el número 4. Suponemos que el dispositivo está inicialmente en algún lugar a la izquierda de los 1s. Está en el estado interno 0 y lee un 0. De acuerdo con la primera instrucción, deja este 0 como está y se mueve un paso a la derecha permaneciendo en el estado interno 0. Sigue haciendo esto, moviéndose un paso a la derecha, hasta que encuentra el primer 1. Entonces entra en juego la segunda instrucción; deja el 1 como 1 y se mueve otra vez a la derecha, pero ahora en el estado interno 1. De acuerdo con la cuarta instrucción, permanece en el estado interno 1, conservando el 1, y moviéndose hacia la derecha hasta que llega al primer 0 que sigue a los 1s. La tercera instrucción le dice entonces que cambie ese 0 por un 1, se mueva un poco más a la derecha (recordemos que STOP representa D.STOP) y luego se detenga. De este modo, otro 1 se ha añadido a la cadena de 1s, y el 4 de nuestro ejemplo se ha cambiado por un 5, como se pretendía.

Como ejercicio algo más costoso se puede comprobar que la máquina $UN \times 2$ definida por

$$\begin{array}{llll} 00 \rightarrow 00D, & 01 \rightarrow 10D, & 10 \rightarrow 101I, & 11 \rightarrow 11D, \\ 100 \rightarrow 110D, & 101 \rightarrow 1000D, & 110 \rightarrow 01STOP, & 111 \rightarrow 111D, \\ 1000 \rightarrow 1011I, & 1001 \rightarrow 1001D, & 1010 \rightarrow 101L, & 1011 \rightarrow 1011I. \end{array}$$

duplica un número unario.

En el caso de **EUC**, para tener una idea de lo que supone, puede ensayarse con algún par explícito de números, digamos 6 y 8. El dispositivo de lectura se supone, como antes, en el estado 0 e inicialmente a la izquierda, y la cinta estará ahora marcada inicialmente de la forma:

...0000000000001111110111111100000...

Cuando, después de muchos pasos, la máquina de Turing se detenga tendremos una cinta marcada

...000011000000000000...

con el dispositivo de lectura a la derecha de los dígitos distintos de cero. De este modo obtenemos (correctamente) 2 para el máximo común divisor buscado.

La explicación completa del *por qué* EUC (o, en su caso, $UN \times 2$) hace realmente lo que se supone que hace entraña algunas sutilezas y resultaría más complicado de explicar de lo que es la propia máquina —¡una característica no poco frecuente de los programas de ordenador! (Entender completamente por qué un procedimiento algorítmico hace lo que se le supone requiere *intuiciones directas*. ¿Son algorítmicas las propias «intuiciones»? Ésta es una pregunta que más adelante tendrá importancia para nosotros.) No intentaré dar aquí una explicación para los ejemplos EUC o $UN \times 2$. El lector que los compruebe descubrirá que me he tomado una pequeña libertad con el auténtico algoritmo de Euclides para expresar las cosas de modo más conciso en el esquema deseado. La descripción de EUC es aún algo complicada pues comprende 22 instrucciones elementales para 11 estados internos distintos. Gran parte de la complicación es de tipo puramente organizativo. Se observará, por ejemplo, que de las 22 instrucciones ¡sólo 3 implican realmente una alteración de las marcas de la cinta! (Incluso para $UN \times 2$ he utilizado 12 instrucciones, la mitad de las cuales implican una alteración de las marcas.)

Codificación binaria de los datos numéricos

El sistema unario es extremadamente poco eficaz para la representación de números de gran tamaño. En consecuencia, utilizaremos normalmente el sistema *binario* de numeración que describimos anteriormente. No obstante, no podemos hacer esto de forma directa, intentando sencillamente leer la cinta como un número binario. Tal como están las cosas no habría modo de decir cuándo termina la representación binaria del número y comienza la sucesión infinita de 0s que representa la cinta en blanco a la derecha. Necesitamos alguna notación para dar por terminada la descripción binaria de un número. Más aún, a menudo tendremos que introducir *varios* números, como el *par* de números ² que requiere el algoritmo de Euclides. Tal como

están las cosas no podemos distinguir los *espacios* entre números de los 0s o cadenas de 0s que aparecen como parte de la representación binaria de un solo número. Además, tal vez nos gustaría introducir todo tipo de complicadas instrucciones en la cinta input tanto como números. Para superar estas dificultades adoptaremos un procedimiento que llamaré *contracción*, según el cual cualquier cadena de 0s y 1s (con un número total finito de 1s) *no* se lee simplemente como un número binario sino que es reemplazada por una cadena de 0s, 1s, 2s, 3s, etc. mediante una receta en la que cada dígito de la segunda secuencia es simplemente el número de 1s que hay entre 0s sucesivos de la primera secuencia. Así, por ejemplo, la secuencia

01000101101010110100011101010111100110

sería reemplazada según

010

0

010110101011010

0

01110101011110

0110

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

ésta:

1

0

0

1

2

1

1

2

1

0

0

3

1

1

4

0

2

Podemos ahora leer los números 2, 3, 4,... como marcadores o instrucciones de algún tipo. Así, consideremos 2 simplemente como una «coma» que indica el espacio entre dos números, mientras 3, 4, 5,... podrían, a nuestro gusto, representar varias instrucciones o notaciones de interés, tales como «menos», «más», «por», «saltar a la posición del número siguiente», «iterar la operación anterior el siguiente número de veces», etc. Ahora tenemos varias cadenas de 0s, y 1s que están separados por dígitos mayores. Las primeras representan números ordinarios escritos en la escala binaria. De este modo, la secuencia anterior se leerá (entendiendo «coma» por «2»):

(número binario 1001) coma (número binario 11) coma...

Utilizando la notación arábica estándar «9», «3», «4», «0» para los números binarios 1001, 11, 100, 0 respectivamente, obtenemos para toda la secuencia

9, 3, 4 (instrucción 3) 3 (instrucción 4) 0,

En particular, este procedimiento nos proporciona un medio para dar por terminada la descripción de un número (y por consiguiente distinguirlo de una extensión infinita de cinta en blanco a la derecha) utilizando simplemente una coma al final. Además, nos permite codificar cualquier secuencia finita de números naturales, escritos en

notación binaria, como una *única* secuencia de 0s y 1s, donde el uso de comas sirve para separar los números. Veamos cómo se hace en un caso específico. Considérese la secuencia

5, 13, 0, 1, 1, 4,

En notación binaria esto es

101, 1101, 0, 1, 1, 100,

que se codifica en la cinta por *expansión* (es decir, el inverso del anterior procedimiento de contracción) como

...000010010110101001011001101011010110100011000....

Para conseguir esta codificación de una forma sencilla y directa podemos hacer reemplazamientos en nuestra secuencia original de números binarios de la siguiente forma

0 → 0
1 → 10
, → 110

y luego añadir una cantidad ilimitada de 0s a ambos lados. Se verá más claro como se ha aplicado esto a la cinta anterior si la espaciamos

0000 10 0 10 110 10 10 0 10 110 0 110 10 110 10 110 10 0 0 110 00

Lamaré a esta notación para (conjuntos de) números notación *binaria expandida*. (Por ejemplo, la forma binaria expandida de 13 es 1010010.)

Hay que hacer una última puntualización sobre esta codificación. Es sólo una cuestión técnica pero necesaria para que todo esté completo ³. En la representación binaria (o decimal) de los números naturales existe una pequeña redundancia en el hecho de que los 0s colocados a la izquierda de una expresión no «cuentan» —y normalmente se omiten, v.g. 00110010 es el mismo número binario que 110010 (y 0050 es el mismo número decimal que 50). Esta redundancia se extiende al propio número cero que puede escribirse 000 ó 00 igual que 0. ¡De hecho, un espacio en blanco también debería lógicamente denotar cero! En notación ordinaria esto llevaría a gran confusión pero encaja muy bien en la notación que acabamos de describir. De este modo, un cero entre dos comas puede escribirse también como dos comas juntas (,,) lo que sería codificado en la cinta como dos pares 11 separados por un solo 0:

...001101100...

Entonces el conjunto anterior de seis números puede escribirse también en notación binaria de la forma

101,1101,,1,1,100,

y ser codificado en la cinta en forma binaria expandida como

...00001001011010100101101101011010100011000...

(donde hay un 0 menos que en la secuencia que teníamos antes).

Ahora podemos considerar una máquina de Turing para ejecutar, por ejemplo, el algoritmo de Euclides aplicándolo a pares de números escritos en notación binaria expandida. Por ejemplo, para el par de números 6 y 8 que consideramos previamente, en lugar de utilizar

000000000001111110111111100000...

como hicimos antes, consideraremos las representaciones binarias de 6 y 8, es decir, 110 y 1000 respectivamente. El *par* es

6,8, esto es, en notación binaria, 110,1000,

que, por expansión, se codifica como la cinta

...00000101001101000011000000...

Para este particular par de números no se gana en concisión con respecto a la forma unaria. Sin embargo, supongamos que tomamos, por ejemplo, los números (decimales) 1583169 y 8610. En notación binaria estos serían

110000010100001000001, 10000110100010,

de modo que tenemos el par codificado como la cinta

...001010000001001000001000000101101000001010010000100110...

que encaja en una línea, mientras que en notación unaria la cinta que representara «1583169,8610» ¡ocuparía más espacio que todo este libro!

Podríamos obtener, si quisiéramos, una máquina de Turing que ejecute el algoritmo de Euclides cuando los números se expresan en notación binaria expandida añadiendo simplemente a **EUC** un par apropiado de algoritmos «subrutina» que traduzcan entre unario y binario expandido. No obstante, esto sería extremadamente poco eficaz puesto que la ineficacia del sistema de numeración unario estaría aún presente «internamente» y quedaría de manifiesto en la lentitud

del dispositivo y en la desmedida cantidad de «papel de borrador» (que estaría en la parte izquierda de la cinta) que sería necesaria. También puede darse una máquina de Turing más eficiente para el algoritmo de Turing que opere íntegramente dentro del binario expandido, pero no sería especialmente ilustrativa para nosotros en este momento.

En lugar de ello, para ilustrar cómo se puede hacer una máquina de Turing que opere con números en expansión binaria ensayaremos algo mucho más sencillo que el algoritmo de Euclides, a saber: el simple proceso de *sumar uno* a un número natural. Esto puede ser efectuado por la máquina de Turing (que llamaré $XN + 1$):

$$\begin{array}{lllll} 00 \rightarrow 00D, & 01 \rightarrow 11D, & 10 \rightarrow 00D, & 11 \rightarrow 101D, & 100 \rightarrow 110I, \\ 101 \rightarrow 101D, & 110 \rightarrow 01STOP, & 111 \rightarrow 1000I, & 1000 \rightarrow 1011I, \\ 1001 \rightarrow 1001I, & 1010 \rightarrow 1100D, & 1011 \rightarrow 101D, & 1101 \rightarrow 1111D, \\ & 1110 \rightarrow 111D, & 1111 \rightarrow 1110D, \end{array}$$

Una vez más, el lector interesado podría dedicarse a comprobar que esta máquina hace realmente lo que se supone que debe hacer, aplicándola por ejemplo al número 167 cuya representación binaria es 10100111 y por lo tanto vendría dado por la cinta

...0000100100010101011000...

Para sumar uno a un número binario simplemente localizamos el último 0 y lo cambiamos por 1, y luego reemplazamos todos los 1 que siguen por 0, v.g., $167 + 1 = 168$ se escribe en la notación binaria

$$10100111 + 1 = 10101000$$

Así, nuestra máquina de Turing de «sumar-uno» debería reemplazar la cinta citada más arriba por

...0000100100100001100000...

que es lo que realmente hace.

Nótese que incluso la muy elemental operación de simplemente sumar uno es un poco complicada con esta notación: ¡utiliza quince instrucciones y ocho estados internos diferentes! Por supuesto, las cosas serían mucho más simples en notación unaria puesto que, en tal caso, «sumar-uno» significa simplemente extender la cadena de 1s con un 1 más, de modo que no es sorprendente que nuestra máquina $UN + 1$ sea más elemental. Sin embargo, para números muy grandes, $UN + 1$ sería extraordinariamente lenta debido a la desmedida longi-

tud de papel necesario, y la máquina más complicada $XN + 1$, que opera con la notación binaria expandida más compacta, sería mejor.

Como un inciso, señalo una operación para la que la máquina de Turing parece realmente más simple en la notación binaria expandida que en la notación unaria, a saber: *multiplicar por dos*. Aquí, la máquina de Turing $XN \times 2$, dada por

$$\begin{array}{llll} 00 \rightarrow 00D, & 01 \rightarrow 10D, & 10 \rightarrow 01D, & 11 \rightarrow 100D, \\ & 100 \rightarrow 111D, & 110 \rightarrow 01STOP, & \end{array}$$

lleva a cabo esto en binario expandido ¡mientras que la correspondiente máquina en notación unaria, $UN \times 2$, que se describió antes, es mucho más complicada!

Todo esto nos da alguna idea a un nivel muy básico de lo que las máquinas de Turing pueden hacer. Como es de esperar, ellas pueden complicarse mucho, y lo hacen, cuando hay que realizar operaciones de cierta complejidad. ¿Cuál es el alcance final de tales dispositivos? Consideraremos inmediatamente esta cuestión.

La tesis de Church-Turing

Una vez que nos hemos familiarizado algo con la construcción de máquinas de Turing sencillas resulta fácil convencerse de que las distintas operaciones aritméticas básicas, tales como la suma de dos números, o su multiplicación, o la elevación de uno a la potencia del otro, pueden ser todas efectuadas por máquinas de Turing concretas. No sería demasiado complicado dar explícitamente estas máquinas pero no me molestaré en hacerlo aquí. También pueden darse operaciones cuyo resultado es un *par* de números naturales, tales como la división con resto —u otras en las que el resultado es un conjunto finito pero arbitrariamente grande de números. Por otra parte, pueden construirse máquinas de Turing para las que no se especifica por adelantado qué operación matemática hay que realizar, sino que las instrucciones para ello están introducidas en la cinta. Quizá la operación concreta que haya que realizar en una etapa depende del resultado de algún cálculo que la máquina tiene que hacer en alguna etapa anterior. («Si la respuesta a ese cálculo era mayor que tal y tal, hacer esto; en caso contrario, hacer lo otro.»). Una vez que nos damos cuenta de que se pueden construir máquinas de Turing que realizan operaciones aritméticas, o simplemente lógicas, es más fácil imaginar de qué forma podrían construirse para realizar tareas más complicadas

de naturaleza algorítmica. Cuando uno ha jugado un rato con tales cosas se convence fácilmente de que puede construirse una máquina de este tipo para realizar ¡*cualquier operación mecánica*! Resulta razonable matemáticamente *definir* una operación mecánica como una operación que puede ser llevada a cabo por una máquina semejante. El sustantivo «algoritmo» y los adjetivos «computable», «recursivo», y «efectivo» son todos ellos usados por los matemáticos para denotar las operaciones mecánicas que pueden ser realizadas por máquinas teóricas de este tipo —las máquinas de Turing. Desde el momento que un procedimiento es suficientemente claro y mecánico, resulta razonable creer que se puede encontrar una máquina de Turing que realmente lo realice. Éste era, después de todo, el objetivo de nuestra (esto es, la de Turing) discusión introductoria que motivó el propio concepto de máquina de Turing.

Por otro lado, puede todavía parecer que el diseño de estas máquinas era quizá innecesariamente restrictivo. Imponer que el dispositivo sólo pueda leer un dígito binario (0 ó 1) cada vez, y moverse solamente un espacio cada vez a lo largo de una *única* cinta unidimensional parece limitador a primera vista. ¿Por qué no permitir cuatro o cinco, o quizá mil cintas separadas, con un gran número de dispositivos de lectura interconectados y funcionando todos a la vez? ¿Por qué no permitir todo un plano de cuadrados con 0s y 1s (o quizá una disposición tridimensional) en lugar de insistir en una cinta unidimensional? ¿Por qué no permitir otros símbolos de un alfabeto o un sistema de numeración más complicado? De hecho, ninguno de estos cambios supone la más mínima diferencia en cuanto a lo que puede conseguirse en principio, aunque alguno pueda suponer cierta diferencia en cuanto a la economía de las operaciones (como ciertamente sería el caso si permitiéramos más de una cinta). ¡La clase de operaciones realizadas, y que por consiguiente caen bajo el rótulo de «algoritmos» (o «computaciones» o «procedimientos efectivos» u «operaciones recursivas»), sería exactamente la misma que antes aunque ampliáramos la definición de nuestras máquinas con todas estas modificaciones a la vez!

Podemos ver que no hay *necesidad* de tener más de una cinta con tal de que el dispositivo siga encontrando en la cinta dada tanto espacio nuevo como necesite. Para ello puede que sea necesario cambiar datos de un lugar a otro de la cinta. Esto puede ser «ineficaz» pero no limita lo que en principio puede conseguirse ⁴. Análogamente, *no* se gana nada en principio (aunque, en ciertas circunstancias, puede conseguirse una mayor velocidad de acción) utilizando más de

un dispositivo de Turing en *acción paralela* —una idea que se ha puesto de moda en los últimos años en relación con los intentos de modelar más exactamente el cerebro humano. Teniendo dos dispositivos separados que no se comunican directamente no se consigue más que teniendo dos que *sí* comunican; ¡y *si* se comunican entonces son, de hecho, un solo dispositivo!

¿Qué sucede con la restricción de Turing a una cinta unidimensional? Si pensamos que esta cinta representa un «entorno», preferiríamos representarla, antes que como una cinta unidimensional, como una superficie plana, o quizá como un espacio tridimensional. Una superficie plana parecería estar más próxima a lo que se necesita para un «organigrama» (como en la descripción anterior de la operación del algoritmo de Euclides) de lo que lo está una cinta unidimensional *. No obstante, no hay dificultad en principio para escribir la operación de un organigrama en una forma «unidimensional» (v.g. mediante una descripción verbal del diagrama). La representación plana bidimensional se hace sólo en razón de nuestra propia conveniencia y facilidad de comprensión y no supone diferencia en lo que puede conseguirse en principio. Siempre es posible codificar el lugar de una marca o un objeto en un plano bidimensional, o incluso en un espacio tridimensional, en una forma simple sobre una cinta unidimensional. (De hecho, utilizar un plano bidimensional es completamente equivalente a utilizar *dos* cintas. Las dos cintas proporcionarían las dos «coordenadas» que serían necesarias para especificar un punto en un plano bidimensional; de modo análogo *tres* cintas pueden actuar como «coordenadas» de un punto en un espacio tridimensional.) Una vez más, esta codificación unidimensional puede ser «ineficaz» pero no limita lo que en principio puede conseguirse.

Pese a todo esto podríamos preguntarnos aún si el concepto de máquina de Turing engloba realmente *todas* las operaciones lógicas o matemáticas que llamaríamos «mecánicas». Cuando Turing escribió su artículo seminal esto era mucho menos claro de lo que es hoy, así que Turing creyó necesario presentar sus argumentos con gran detalle. Lo

* Tal como se han descrito las cosas aquí, este organigrama formaría él mismo parte del «dispositivo» antes que de la «cinta» entorno. Eran los verdaderos números A, B, A-B, etc. los que se representaban en la cinta. Sin embargo, también nos gustaría expresar las especificaciones del *dispositivo* en una forma lineal uni-dimensional. Como veremos más adelante, cuando tratemos la máquina *universal* de Turing, existe una íntima relación entre las especificaciones para un «dispositivo» particular y la especificación de posibles «datos» (o «programa») para un dispositivo dado. Es por ello conveniente tenerlas *ambas* en forma uni-dimensional.

que Turing argumentaba rigurosamente encontró apoyo adicional en el hecho de que, de forma independiente (y en realidad un poco antes), el lógico americano Alonzo Church (con la ayuda de S. C. Kleene) había desarrollado un esquema —el cálculo lambda— dirigido también a resolver el *Entscheidungsproblem* de Hilbert. Aunque este esquema mecánico omnicomprendivo no resultaba tan obvio como el de Turing, tenía algunas ventajas en la sorprendente economía de su estructura matemática. (Describiré el notable cálculo de Church al final de este capítulo.) También independientemente de Turing hubo aún otras propuestas para resolver el problema de Hilbert (ver Gandy, 1988), muy en particular la del lógico polaco-americano Emil Post (un poco después que Turing, pero con ideas mucho más afines a las de Turing que a las de Church). Pronto se demostró que todos estos esquemas eran completamente equivalentes. Esto reforzó considerablemente la idea, que se llegó a conocer como Tesis de Church-Turing, de que el concepto de máquina de Turing (o sus equivalentes) definía realmente lo que, en matemáticas, entendemos por procedimiento algorítmico (o efectivo o recursivo o mecánico). Ahora que los computadores electrónicos de alta-velocidad han llegado a ser una parte tan familiar de nuestras vidas, poca gente parece sentir la necesidad de cuestionarse esta tesis en su forma original. En lugar de ello, se ha dirigido la atención hacia el tema de si los sistemas *físicos* reales (incluyendo presumiblemente los cerebros humanos) —sujetos como están a leyes *físicas* precisas— son capaces de realizar más, menos o exactamente las mismas operaciones lógicas y matemáticas que las máquinas de Turing. Personalmente, acepto de buen grado la forma *matemática* original de la Tesis de Church-Turing. Por otro lado, su relación con el comportamiento de los sistemas físicos reales es un tema aparte que ocupará nuestra atención más adelante en este libro.

Números diferentes de los naturales

En la discusión anterior consideramos operaciones con *números naturales*, y señalamos el hecho notable de que simples máquinas de Turing pueden manejar números naturales de tamaño arbitrariamente grande a pesar de que cada máquina tiene un número *finito* fijo de estados internos diferentes. No obstante, a menudo tenemos que trabajar con tipos de números más complicados que aquéllos, tales como números negativos, fracciones, o números con infinitos decimales. Los

números negativos y las fracciones (v.g. números como $-597/26$) pueden ser fácilmente manejados por máquinas de Turing, y los numeradores y denominadores pueden ser tan grandes como se quiera. Todo lo que necesitamos es una codificación apropiada para los signos «-» y «/», y esto puede hacerse utilizando la notación binaria expandida descrita anteriormente (por ejemplo, «3» para «-» y «4» para «/» —codificados como 1110 y 11110, respectivamente, en notación binaria expandida). Los números negativos y fracciones se tratan entonces en términos de conjuntos finitos de números naturales, así que no aportan nada nuevo con respecto a las cuestiones generales de computabilidad.

Análogamente, las expresiones decimales *finitas* de cualquier longitud no nos aportan nada nuevo, ya que sólo son casos particulares de fracciones. Por ejemplo, la aproximación decimal finita al número irracional π , dada por 3.14159265, es simplemente la fracción $314159265/100000000$. Sin embargo, las expresiones decimales *infinitas*, tales como *toda* la interminable expansión

$$\pi = 3.14159265358979\dots$$

presentan ciertas dificultades. Ni el input ni el output de una máquina de Turing pueden ser, estrictamente hablando, números decimales infinitos. Se podría pensar que es posible encontrar una máquina de Turing que produzca uno tras otro *todos* los dígitos sucesivos, 3, 1, 4, 5, 9,... de la expansión anterior para π en el output de la cinta, y donde simplemente permitimos a la máquina que siga funcionando indefinidamente. Pero esto *no le está permitido* a una máquina de Turing. Debemos esperar a que la máquina se detenga (¡indicado por el sonido de la campana!) antes de poder examinar el output. Mientras la máquina no haya alcanzado una instrucción STOP, el output está sujeto a posibles cambios y, por consiguiente, no puede ser dado por válido. Por otro lado, una vez que se ha alcanzado el STOP, el output es necesariamente finito.

Existe, sin embargo, un procedimiento para hacer *legítimamente* que una máquina de Turing produzca dígitos uno tras otro de una forma muy parecida a ésta. Si queremos generar una expansión decimal infinita, por ejemplo la de π , tendríamos una máquina de Turing que produce la parte entera, 3, haciendo que la máquina actúe sobre 0; luego produciría la primera cifra decimal, 1, haciendo que la máquina actúe sobre 1; luego la segunda cifra decimal, 4, haciéndola actuar sobre 2; luego la tercera 1, haciéndola actuar sobre 3, y así

sucesivamente. De hecho *existe* una máquina de Turing para producir en *este* sentido la expansión decimal completa de π , aunque sería algo complicado definirla explícitamente. Una puntualización similar es aplicable a otros muchos números irracionales, tales como $\sqrt{2} = 1.414213562\dots$. Sucede, sin embargo, que algunos irracionales (curiosamente) no pueden ser producidos por ninguna máquina de Turing, como veremos en los próximos capítulos. Los números que pueden ser generados de esta forma se llaman *computables* (Turing, 1937). Los que no pueden (¡en realidad la inmensa mayoría!) se llaman *no-computables*. Volveré a este tema, y otros relacionados, en los últimos capítulos. Resultará de cierta importancia para nosotros en relación con la cuestión de si un *objeto físico real* (v.g. un cerebro humano) puede describirse adecuadamente, según nuestras teorías físicas, en términos de estructuras matemáticas computables.

El tema de la computabilidad es de importancia general en matemáticas. No deberíamos pensar en ello como algo que sólo concierne a los *números* como tales. Podemos tener máquinas de Turing que operen directamente sobre *fórmulas matemáticas* como, por ejemplo, expresiones algebraicas o trigonométricas, o que lleven a cabo las manipulaciones formales de cálculo. Todo lo que se necesita es una forma precisa de codificación en secuencias de 0s y 1s de todos los símbolos matemáticos involucrados y, a continuación, ya podemos aplicar el concepto de máquina de Turing. Después de todo, esto es lo que Turing tenía en mente cuando abordó el *Eintscheidungsproblem*, que reclamaba un procedimiento algorítmico para responder cuestiones matemáticas de naturaleza *general*. Volveremos a esto en breve.

La máquina universal de Turing

Todavía no he descrito el concepto de máquina *universal* de Turing. No es demasiado difícil enunciar el principio que hay detrás de ello, aunque los detalles son complicados. La idea básica consiste en codificar la lista de instrucciones para una máquina de Turing arbitraria T en una cadena de 0s y 1s que pueda ser representada en una cinta. Esta cinta se utiliza a continuación como parte inicial del input de alguna máquina de Turing *particular* U —que llamamos una máquina universal de Turing— que actúa entonces sobre el resto del input de la misma forma que lo hubiera hecho T . La máquina *universal* de Turing es un imitador universal. ¡La parte inicial de la cinta

proporciona a la máquina U toda la información que necesita para imitar exactamente a cualquier máquina T !

Para ver cómo funciona esto necesitamos en primer lugar un modo sistemático de *numerar* máquinas de Turing. Consideremos la lista de instrucciones que definen alguna máquina de Turing particular, por ejemplo una de las descritas más arriba. Debemos codificar esta lista en una cadena de 0s y 1s siguiendo un esquema preciso. Esto puede hacerse con la ayuda del procedimiento de «contracción» que adoptamos antes. En efecto, si representamos los símbolos D, I, STOP, la flecha ($\rightarrow$) y la coma respectivamente mediante los números 2, 3, 4, 5 y 6, podemos codificarlos como contracciones por 110, 1110, 11110, 111110 y 1111110. Entonces los dígitos 0 y 1, codificados como 0 y 10, respectivamente, pueden ser utilizados en las cadenas reales de estos símbolos que aparecen en la tabla. No necesitamos una notación diferente para distinguir las cifras de caracteres mayores 0 y 1 en la tabla para la máquina de Turing de las más pequeñas en negrita, ya que la posición de los dígitos grandes en el extremo de la numeración binaria es suficiente para distinguirlos de los demás. Así, por ejemplo, 1101 se leería con el número binario 1101 y se codificaría en la cinta como 1010010. En particular, 00 se leería como 00, que puede codificarse sin ambigüedad como 0, o como un símbolo omitido totalmente. Podemos economizar considerablemente si no nos molestamos en codificar ninguna flecha ni ninguno de los símbolos que las preceden inmediatamente, contando en su lugar con el ordenamiento numérico de las instrucciones para especificar cuáles deben ser estos símbolos —aunque para adoptar este procedimiento debemos asegurarnos de que no haya huecos en esta ordenación, añadiendo algunas órdenes «mudas» extra donde sea necesario. (Por ejemplo, la máquina de Turing $XN + 1$ no tiene ninguna orden que nos diga qué hacer con 1100, ya que nunca se da esta combinación en el funcionamiento de la máquina, de modo que debemos insertar una orden «muda», por ejemplo $1100 \rightarrow 00D$, que puede incorporarse a la lista sin cambiar nada. Analogamente deberíamos insertar $101 \rightarrow 00D$ en la máquina $XN \times 2$.) Sin estas órdenes «mudas», la codificación de las órdenes subsiguientes de la lista se echaría a perder. En realidad no necesitamos la coma al final de cada instrucción, como aparece, puesto que los símbolos I o D bastan para separar las instrucciones. En consecuencia, adoptamos sencillamente la siguiente codificación:

0 para 0 ó 0,	10 para 1 ó 1,	110 para D,	1110 para I,
	11110 para STOP		

Como ejemplo, vamos a codificar la máquina de Turing $XN + 1$ (con la instrucción $1100 \rightarrow 00D$ insertada). Prescindiendo de las flechas, de los dígitos que inmediatamente las preceden, y también de las comas tenemos

00D 11D 00D 101D 110I 101D 01STOP 1000I 1011I
1001I 1100D 101D 00D 1111D 111D 1110D.

Podemos mejorar esto prescindiendo de todos los 00 y reemplazando cada 01 por simplemente 1, según lo que hemos dicho antes, para obtener

D11DD101D110I101D1STOP1000I1011I1000D101DD1111D111D1110D

Esto se codifica como la secuencia en la cinta

1 1 0 1 0 1 0 1 1 0 1 1 0 1 0 0 1 0 1 1 0 1 0 1 0 0 1 1 1 0 1 0 0 1 0 1
1 0 1 0 1 1 1 1 0 1 0 0 0 0 1 1 1 0 1 0 0 1 0 1 0 1 1 1 0 1 0 0 0 1 0 1
1 1 0 1 0 1 0 0 0 1 1 0 1 0 0 1 0 1 1 0 1 1 0 1 0 1 0 1 0 1 0 1 1 0 1 0
1 0 1 0 1 1 0 1 0 1 0 1 0 0 1 1 0

Como dos nuevos ahorros menores podemos también borrar siempre el 110 inicial (junto con la extensión infinita de cinta en blanco que le precede) puesto que esto significa 00D, que representa la instrucción inicial $00 \rightarrow 00D$ que he estado suponiendo implícitamente que es común a *todas* las máquinas de Turing —de modo que el dispositivo puede comenzar arbitrariamente lejos a la izquierda de las marcas sobre la cinta e ir hacia la derecha hasta llegar a la primera marca— y podemos también borrar siempre el 110 final (y la implícita secuencia infinita de 0s que se supone le sigue) ya que todas las descripciones de máquinas de Turing deben acabar de esta forma (pues todas terminan con D, I o STOP). El *número binario* resultante es el *número* de la máquina de Turing, que en el caso de $XN + 1$ es:

1 0 1 0 1 1 0 1 1 0 1 0 0 1 0 1 1 0 1 0 1 0 0 1 1 1 0 1 0 0 1 0 1 1 0 1
0 1 1 1 1 0 1 0 0 0 0 1 1 1 0 1 0 0 1 0 1 0 1 1 1 0 1 0 0 0 1 0 1 1 1 0
1 0 1 0 0 0 1 1 0 1 0 0 1 0 1 1 0 1 1 0 1 0 1 0 1 0 1 0 1 1 0 1 0 1 0 1
0 1 1 0 1 0 1 0 1 0 0 .

En notación decimal estándar este concreto número es

450813704461563958982113775643437908

A veces nos referimos de forma un tanto imprecisa a la máquina de Turing cuyo número es n como la n -ésima máquina de Turing, denotada

por T_n . Así ¡ $XN + 1$ es la 45081370446156395892113775643437908-ésima máquina de Turing!

Parece un hecho sorprendente que tengamos que ir tan lejos en la «lista» de máquinas de Turing ¡antes de encontrar una que realice siquiera una operación tan trivial como la de sumar uno (en la notación binaria expandida) a un número natural! (No creo haber sido torpemente ineficaz en mi codificación, aunque veo lugar para algunas mejoras menores.) En realidad existen algunas máquinas de Turing con números más bajos que son de interés. Por ejemplo, $UN + 1$ tiene el número binario

101011010111101010

¡que es simplemente 177642 en notación decimal! En consecuencia, la particularmente trivial máquina de Turing $UN + 1$, que simplemente coloca un 1 adicional al final de una secuencia de 1s, es la 177642-ésima máquina de Turing. A modo de curiosidad podemos señalar que «multiplicar por dos» queda en alguna parte entre estas dos en la lista de máquinas de Turing, en cualquiera de las notaciones, pues encontramos que el número para $XN \times 2$ es 10389728107 mientras que el de $UN \times 2$ es 1492 923 420 919 872 026 917 547 669.

Quizá no sorprenda saber, a la vista de la magnitud de estos números, que la inmensa mayoría de los números naturales no dan máquinas de Turing que trabajen de forma alguna. Hagamos la lista de las trece primeras máquinas de Turing según esta numeración:

T_0 :	00 $\rightarrow$ 00D,	01 $\rightarrow$ 00D,	
T_1 :	00 $\rightarrow$ 00D,	01 $\rightarrow$ 00I,	
T_2 :	00 $\rightarrow$ 00D,	01 $\rightarrow$ 01D,	
T_3 :	00 $\rightarrow$ 00D,	01 $\rightarrow$ 00STOP,	
T_4 :	00 $\rightarrow$ 00D,	01 $\rightarrow$ 10D,	
T_5 :	00 $\rightarrow$ 00D,	01 $\rightarrow$ 01I,	
T_6 :	00 $\rightarrow$ 00D,	01 $\rightarrow$ 00D,	10 $\rightarrow$ 00D,
T_7 :	00 $\rightarrow$ 00D,	01 $\rightarrow$???,	
T_8 :	00 $\rightarrow$ 00D,	01 $\rightarrow$ 101D,	
T_9 :	00 $\rightarrow$ 00D,	01 $\rightarrow$ 10I,	
T_{10} :	00 $\rightarrow$ 00D,	01 $\rightarrow$ 11D,	
T_{11} :	00 $\rightarrow$ 00D,	01 $\rightarrow$ 01STOP,	
T_{12} :	00 $\rightarrow$ 00D,	01 $\rightarrow$ 00D,	10 $\rightarrow$ 00D.

De estas, T_0 simplemente se mueve hacia la derecha borrando todo lo que encuentra, sin detenerse nunca ni volver atrás. La má-

quina T_1 consigue finalmente el mismo efecto pero de una forma más torpe, saltando hacia atrás cada vez que borra una marca de la cinta. La máquina T_2 , al igual que la T_0 , también se mueve incesantemente hacia la derecha, pero es más respetuosa y deja todo tal como estaba antes en la cinta. Ninguna de estas sirve como máquina de Turing ya que ninguna de las dos se detiene nunca. T_3 es la primera máquina respetable: se detiene, modestamente, después de cambiar el primer 1 (el más a la izquierda) por un 0.

T_4 tropieza con un serio problema. Una vez que encuentra su primer 1 en la cinta entra en un estado interno para el que no hay listado, de modo que no tiene instrucciones sobre lo que debe hacer a continuación. T_8 , T_9 y T_{10} tropiezan con el mismo problema. La dificultad con T_7 es aún más básica. La cadena de 0s y 1s que la codifica incluye una secuencia de *cinco* 1 sucesivos: 110111110. No existe interpretación para semejante secuencia, así que se quedará bloqueada en cuanto encuentre su primer 1 en la cinta. (Llamaré a T_7 , o a cualquier otra máquina T_n para la que la expansión binaria de n contiene una secuencia de más de cuatro 1s, una máquina *no correctamente especificada*.) Las máquinas T_5 , T_6 y T_{12} tropiezan con problemas similares a los de T_0 , T_1 y T_2 . Sencillamente continúan indefinidamente sin detenerse nunca. ¡Todas las máquinas T_0 , T_1 , T_2 , T_4 , T_5 , T_6 , T_7 , T_8 , T_9 , T_{10} y T_{12} son inútiles! Sólo T_3 y T_{11} son máquinas de Turing que funcionan, y no muy interesantes por cierto. T_{11} es aún más modesta que T_3 : ¡se detiene al primer encuentro con un 1 y no cambia nada!

Señalemos que hay también una redundancia en nuestra lista. La máquina T_{12} es idéntica a T_6 , y también idéntica en actuación a T_1 , puesto que el estado interno 1 de T_6 y T_{12} nunca interviene. No tenemos que molestarnos por esta redundancia ni por la proliferación de máquinas de Turing inútiles en la lista. Sería desde luego posible mejorar nuestra codificación de modo que se eliminaran muchas de las inútiles y se redujera considerablemente la redundancia. Todo esto se haría a expensas de complicar nuestra pobre máquina universal de Turing que tiene que descifrar el código y tratar de ser la máquina de Turing T_n cuyo número n está leyendo. Merecería la pena hacerlo si pudiéramos eliminar *todas* las inútiles (o toda la redundancia). Pero, como veremos en un momento, ¡esto *no* es posible! Por consiguiente, dejemos nuestra codificación como está. Será conveniente interpretar una cinta con su serie de marcas, v.g.

...0001101110010000...

como la representación binaria de algún número. Recordemos que los 0s continúan indefinidamente en ambos extremos pero que hay sólo un número finito de 1s. Estoy suponiendo también que el número de 1s es *distinto de cero* (esto es, hay al menos un 1). Podríamos decidir leer la cadena finita de símbolos entre el primer y el último 1 (inclusive), que en el caso anterior es

110111001,

como la descripción binaria de un número natural (aquí 441, en notación decimal). Sin embargo, este procedimiento sólo nos daría números impares (números cuya representación binaria termina con un 1) y pretendemos representar *todos* los números naturales. En consecuencia, adoptamos el sencillo expediente de eliminar el 1 final (que se toma como un simple marcador que señala la terminación de la expresión) y leer lo que queda como un número binario ⁵. Así, para el ejemplo anterior, tenemos el número binario

11011100,

que en notación decimal es 220. Este método tiene la ventaja de que el cero también está representado por una cinta marcada, a saber

...0000001000000...

Consideremos la acción de la máquina de Turing T_n sobre alguna cadena (finita) de 0s y 1s en una cinta que alimentamos en la parte derecha. Será conveniente considerar también esta cinta como la representación binaria de algún número, digamos m , según el esquema dado más arriba. Supongamos que tras una serie de pasos la máquina T_n finalmente se detiene (es decir, alcanza STOP). La cadena de dígitos binarios que la máquina ha producido a la izquierda es la respuesta al cálculo. Leemos también esto, del mismo modo, como la representación binaria de un número, digamos p . Escribiremos esta relación, que expresa el hecho de que cuando la n -ésima máquina de Turing actúa sobre m produce p , como:

$$T_n(m) = p.$$

Consideremos ahora esta relación de una forma ligeramente diferente. Imaginemos que expresa una operación particular que se aplica al *par* de números n y m para dar lugar al número p . (Así: dados los *dos* números n y m podemos calcular a partir de ellos cuál es p viendo lo que la n -ésima máquina de Turing hace con m .) Esta operación

particular es un procedimiento totalmente algorítmico y, por lo tanto, puede ser llevado a cabo por *una particular* máquina de Turing U ; esto es, U actúa sobre el par (n, m) para producir p . Puesto que la máquina de Turing tiene que actuar sobre *ambos*, n y m , para producir el simple resultado p , necesitamos algún modo de codificar el par (n, m) en *la única* cinta. Para esto podemos suponer que n está escrito en la notación binaria ordinaria e inmediatamente a continuación terminado con la secuencia 11110. (Recordemos que el número binario de toda máquina de Turing correctamente especificada es una secuencia formada sólo a base de 0, 10, 110, 1110 y 11110, y por consiguiente no contiene ninguna secuencia de más de cuatro 1. De este modo, si T_n es una máquina correctamente especificada, la ocurrencia de 11110 significa en realidad que la descripción del número n termina con ella.) Todo lo que hay a continuación será simplemente la cinta representada por m de acuerdo con nuestra prescripción anterior (es decir, el número binario m seguido inmediatamente de 1000...). Entonces esta segunda parte es sencillamente la cinta sobre la que se supone que actúa T_n .

A modo de ejemplo, si tomamos $n = 11$ y $m = 6$, tenemos, para la cinta sobre la que U tiene que actuar, la secuencia de marcas.

...00010111111011010000....

Ésta está formada por:

- ...0000 (cinta en blanco inicial)
- 1011 (representación binaria de 11)
- 11110 (termina n)
- 110 (representación binaria de 6)
- 10000... (resto de la cinta)

Lo que tendría que hacer la máquina de Turing U , en cada paso sucesivo de la operación de T_n sobre m , sería examinar la estructura de la serie de dígitos en la expresión de n de modo que pueda hacerse la sustitución apropiada de los dígitos para m (esto es, la «cinta» de T_n). De hecho, no es difícil en principio (aunque sí indudablemente tedioso en la práctica) ver cómo se podría construir realmente una máquina semejante. Su propia lista de instrucciones nos estaría proporcionando simplemente un medio de leer la entrada apropiada en dicha «lista», entrada que está codificada en el número n , en cada paso de la aplicación a los dígitos de la «cinta», tal como figuran en m . Por descontado, habría una serie de fintas atrás y adelante entre

los dígitos de m y los de n , y el procedimiento tendería a ser enormemente lento. De todas formas, puede darse ciertamente una lista de instrucciones para una máquina semejante; y llamamos a esa máquina una máquina universal de Turing. Indicando la acción de esta máquina sobre el par de números n y m por $U(n,m)$, tenemos:

$$U(n,m) = T_n(m)$$

para cada (n,m) para el que T_n es una máquina de Turing correctamente especificada ⁶. La máquina U , cuando se alimenta en primer lugar con el número n , imita exactamente la n -ésima máquina de Turing!

Puesto que U es una máquina de Turing, ella misma tendrá un número; es decir, tenemos

$$U = T_u$$

para algún número u . ¿Cómo de grande es u ? De hecho podemos tomar *exactamente*

$u =$ 7 2 4 4 8 5 5 3 3 5 3 3 9 3 1 7 5 7 7 1 9 8 3 9 5 0 3 9 6 1 5 7 1
1 2 3 7 9 5 2 3 6 0 6 7 2 5 5 6 5 5 9 6 3 1 1 0 8 1 4 4 7 9 6 6 0 6 5
0 5 0 5 9 4 0 4 2 4 1 0 9 0 3 1 0 4 8 3 6 1 3 6 3 2 3 5 9 3 6 5 6 4 4
4 4 3 4 5 8 3 8 2 2 2 6 8 8 3 2 7 8 7 6 7 6 2 6 5 5 6 1 4 4 6 9 2 8 1
4 1 1 7 7 1 5 0 1 7 8 4 2 5 5 1 7 0 7 5 5 4 0 8 5 6 5 7 6 8 9 7 5 3 3
4 6 3 5 6 9 4 2 4 7 8 4 8 8 5 9 7 0 4 6 9 3 4 7 2 5 7 3 9 9 8 8 5 8 2
2 8 3 8 2 7 7 9 5 2 9 4 6 8 3 4 6 0 5 2 1 0 6 1 1 6 9 8 3 5 9 4 5 9 3
8 7 9 1 8 8 5 5 4 6 3 2 6 4 4 0 9 2 5 5 2 5 5 0 5 8 2 0 5 5 5 9 8 9 4
5 1 8 9 0 7 1 6 5 3 7 4 1 4 8 9 6 0 3 3 0 9 6 7 5 3 0 2 0 4 3 1 5 5 3
6 2 5 0 3 4 9 8 4 5 2 9 8 3 2 3 2 0 6 5 1 5 8 3 0 4 7 6 6 4 1 4 2 1 3
0 7 0 8 8 1 9 3 2 9 7 1 7 2 3 4 1 5 0 5 6 9 8 0 2 6 2 7 3 4 6 8 6 4 2
9 9 2 1 8 3 8 1 7 2 1 5 7 3 3 3 4 8 2 8 2 3 0 7 3 4 5 3 7 1 3 4 2 1 4
7 5 0 5 9 7 4 0 3 4 5 1 8 4 3 7 2 3 5 9 5 9 3 0 9 0 6 4 0 0 2 4 3 2 1
0 7 7 3 4 2 1 7 8 8 5 1 4 9 2 7 6 0 0 7 9 7 5 9 7 6 3 4 4 1 5 1 2 3 0
7 9 5 8 6 3 9 6 3 5 4 4 9 2 2 6 9 1 5 9 4 7 9 6 5 4 6 1 4 7 1 1 3 4 5
7 0 0 1 4 5 0 4 8 1 6 7 3 3 7 5 6 2 1 7 2 5 7 3 4 4 6 4 5 2 2 7 3 1 0
5 4 4 8 2 9 8 0 7 8 4 9 6 5 1 2 6 9 8 8 7 8 8 9 6 4 5 6 9 7 6 0 9 0 6
6 3 4 2 0 4 4 7 7 9 8 9 0 2 1 9 1 4 4 3 7 9 3 2 8 3 0 0 1 9 4 9 3 5 7
0 9 6 3 9 2 1 7 0 3 9 0 4 8 3 3 2 7 0 8 8 2 5 9 6 2 0 1 3 0 1 7 7 3 7
2 7 2 0 2 7 1 8 6 2 5 9 1 9 9 1 4 4 2 8 2 7 5 4 3 7 4 2 2 3 5 1 3 5 5
6 7 5 1 3 4 0 8 4 2 2 2 2 9 9 8 8 9 3 7 4 4 1 0 5 3 4 3 0 5 4 7 1 0 4
4 3 6 8 6 9 5 8 7 6 4 0 5 1 7 8 1 2 8 0 1 1 9 4 3 7 5 3 0 8 1 3 8 7 0
6 3 9 9 4 2 7 7 2 8 2 3 1 5 6 4 2 5 2 8 9 2 3 7 5 1 4 5 6 5 4 4 3 8 9

9 0 5 2 7 8 0 7 9 3 2 4 1 1 4 4 8 2 6 1 4 2 3 5 7 2 8 6 1 9 3 1 1 8 3
 3 2 6 1 0 6 5 6 1 2 2 7 5 5 5 3 1 8 1 0 2 0 7 5 1 1 0 8 5 3 3 7 6 3 3
 8 0 6 0 3 1 0 8 2 3 6 1 6 7 5 0 4 5 6 3 5 8 5 2 1 6 4 2 1 4 8 6 5 4 2
 3 4 7 1 8 7 4 2 6 4 3 7 5 4 4 4 2 8 7 9 0 0 6 2 4 8 5 8 2 7 0 9 1 2 4
 0 4 2 2 0 7 6 5 3 8 7 5 4 2 6 4 4 5 4 1 3 3 4 5 1 7 4 8 5 6 6 2 9 1 5
 7 4 2 9 9 9 0 9 5 0 2 6 2 3 0 0 9 7 3 3 7 3 8 1 3 7 7 2 4 1 6 2 1 7 2
 7 4 7 7 2 3 6 1 0 2 0 6 7 8 6 8 5 4 0 0 2 8 9 3 5 6 6 0 8 5 6 9 6 8 2
 2 6 2 0 1 4 1 9 8 2 4 8 6 2 1 6 9 8 9 0 2 6 0 9 1 3 0 9 4 0 2 9 8 5 7
 0 6 0 0 1 7 4 3 0 0 6 7 0 0 8 6 8 9 6 7 5 9 0 3 4 4 7 3 4 1 7 4 1 2 7
 8 7 4 2 5 5 8 1 2 0 1 5 4 9 3 6 6 3 9 3 8 9 9 6 9 0 5 8 1 7 7 3 8 5 9
 1 6 5 4 0 5 5 3 5 6 7 0 4 0 9 2 8 1 3 3 2 2 2 1 6 3 1 4 1 0 9 7 8 7 1
 0 8 1 4 5 9 9 7 8 6 6 9 5 9 9 7 0 4 5 0 9 6 8 1 8 4 1 9 0 6 2 9 9 4 4
 3 6 5 6 0 1 5 1 4 5 4 9 0 4 8 8 0 9 2 2 0 8 4 4 8 0 0 3 4 8 2 2 4 9 2
 0 7 7 3 0 4 0 3 0 4 3 1 8 8 4 2 9 8 9 9 3 9 3 1 3 5 5 2 6 6 8 8 2 3 4
 9 6 6 2 1 0 1 9 4 7 1 6 1 9 1 0 7 0 1 4 6 1 9 6 8 5 2 3 1 9 2 8 4 7 4
 8 2 0 3 4 4 9 5 8 9 7 7 0 9 5 5 3 5 6 1 1 0 7 0 2 7 5 8 1 7 4 8 7 3 3
 3 2 7 2 9 6 6 7 8 9 9 8 7 9 8 4 7 3 2 8 4 0 9 8 1 9 0 7 6 4 8 5 1 2 7
 2 6 3 1 0 0 1 7 4 0 1 6 6 7 8 7 3 6 3 4 7 7 6 0 5 8 5 7 2 4 5 0 3 6 9
 6 4 4 3 4 8 9 7 9 9 2 0 3 4 4 8 9 9 9 7 4 5 5 6 6 2 4 0 2 9 3 7 4 8 7
 6 6 8 8 3 9 7 5 1 4 0 4 4 5 1 6 6 5 7 0 7 7 5 0 0 6 0 5 1 3 8 8 3 9 9
 1 6 6 8 8 1 4 0 7 2 5 4 5 5 4 4 6 6 5 2 2 2 0 5 0 7 2 4 2 6 2 3 9 2 3
 7 9 2 1 1 5 2 5 3 1 8 1 6 2 5 1 2 5 3 6 3 0 5 0 9 3 1 7 2 8 6 3 1 4 2
 2 0 0 4 0 6 4 5 7 1 3 0 5 2 7 5 8 0 2 3 0 7 6 6 5 1 8 3 3 5 1 9 9 5 6
 8 9 1 3 9 7 4 8 1 3 7 5 0 4 9 2 6 4 2 9 6 0 5 0 1 0 0 1 3 6 5 1 9 8 0
 1 8 6 9 4 5 6 3 9 4 9 8

(o alguna otra posibilidad de al menos este orden de magnitud). ¡Sin duda, este número parece alarmantemente grande! En efecto *es* alarmantemente grande, pero yo no he sido capaz de ver cómo se hubiera podido reducir significativamente. Los procedimientos de codificación y las especificaciones que he dado para máquinas de Turing son bastante razonables y simples pero, a pesar de todo, uno es conducido inevitablemente a un número de este orden de magnitud para la codificación de una máquina universal de Turing real⁷.

He dicho que todos los computadores modernos de tipo general son, en efecto, máquinas universales de Turing. No quiero decir que el diseño lógico de estos ordenadores necesite parecerse mucho al tipo de descripción de una máquina universal de Turing que acabo de dar. El punto importante es simplemente que, si suministramos primero a cualquier máquina universal de Turing un programa apropiado (la parte inicial de la cinta input), ¡podemos hacer que imite el compor-

tamiento de una máquina de Turing cualquiera! En la descripción anterior, el programa toma simplemente la forma de un simple número (el número n), pero son posibles otros procedimientos y hay muchas variaciones sobre el tema original de Turing. De hecho, en mis propias descripciones me he desviado algo de las que dio Turing originalmente. Ninguna de estas diferencias es importante para nuestras actuales necesidades.

La insolubilidad del problema de Hilbert

Llegamos ahora al objetivo para el que Turing desarrolló originalmente sus ideas, la resolución del muy general *Entscheidungsproblem* de Hilbert: ¿existe algún procedimiento mecánico para responder a todas las cuestiones matemáticas pertenecientes a una clase amplia pero bien definida? Turing descubrió que podía enunciar su versión de la pregunta en términos del problema de decidir si la n -ésima máquina de Turing se *parará* o no cuando actúe sobre el número m . Este problema fue llamado *problema de la detención*. Resulta fácil construir una lista de instrucciones con arreglo a las cuales la máquina no se parará para *ningún* número m (por ejemplo, $n = 1$ ó 2 , como sucede más arriba, o cualquier otro caso en el que no haya ninguna instrucción STOP). Existen también muchas listas de instrucciones con arreglo a las cuales la máquina siempre se parará, cualquiera que sea el número dado (v.g. $n = 11$); y algunas máquinas se pararán para unos números pero no para otros. Sería correcto decir que un presunto algoritmo no es de mucha utilidad cuando sigue actuando indefinidamente sin pararse nunca. Eso no es en modo alguno un algoritmo. ¡Por ello, es una cuestión importante el poder decidir si T_n aplicada a m dará o no una respuesta! Si no lo hace (esto es, si el cálculo *no* se para) escribiremos

$$T_n(m) = \square$$

(Se incluirán en esta notación aquellas situaciones en las que la máquina de Turing abandona un problema en algún paso debido a que no encuentra ninguna instrucción adecuada que le diga lo que tiene que hacer —como sucede con las máquinas inútiles tales como T_4 y T_7 consideradas más arriba. También, desgraciadamente, nuestra máquina aparentemente adecuada T_3 debe ser ahora considerada inútil: $T_3(m) = \square$, debido a que el resultado de la acción de T_3 es siempre una cinta en blanco, ¡mientras que al menos necesitamos un 1 en el output para que sea asignado un número al resultado del cálculo! La

máquina T_{11} es, sin embargo, legítima puesto que produce un solo 1. Este output es la cinta numerada 0, de modo que tenemos $T_{11}(m) = 0$ para todo m .)

Sería importante en matemáticas el poder decidir cuándo se va a parar una máquina de Turing. Consideremos por ejemplo la ecuación:

$$(x + 1)^{w+3} + (y + 1)^{w+3} = (z + 1)^{w+3}$$

(Si las ecuaciones matemáticas son cosas que le molestan, ¡no se vaya! Esta ecuación se utiliza sólo como ejemplo y no hay necesidad de comprenderla en detalle). Esta ecuación concreta está relacionada con un famoso problema no resuelto en matemáticas —quizá el más famoso de todos. El problema es el siguiente: ¿existe *un* conjunto de números naturales w, x, y, z para el que esta ecuación se satisface? El famoso enunciado conocido como «último teorema de Fermat», escrito en un margen de la *Aritmética* de Diofanto por el gran matemático francés del siglo XVII Pierre de Fermat (1601-1665), afirma que la ecuación *nunca* se satisface ⁸ *. Aunque abogado de profesión (y contemporáneo de Descartes), Fermat era el mejor matemático de su época. El dijo tener «una demostración verdaderamente maravillosa» de su afirmación, que no cabía en el estrecho margen del libro; ¡pero hasta ahora nadie ha sido capaz de reconstruir tal demostración ni, tampoco, encontrar ningún contraejemplo de la afirmación de Fermat!

Es evidente que, una vez *dada* la cuadrupla de números (w, x, y, z) , es una simple cuestión de cálculo el decidir si se satisface o no la ecuación. Entonces podríamos imaginar un algoritmo para computador que recorra todas las cuadruplas de números una tras otra y se pare cuando se satisfaga la ecuación. (Hemos visto que hay modos de codificar conjuntos finitos de números, de una forma computable, en una sola cinta, esto es, simplemente como números individuales, de modo que podemos «recorrer» todas las cuadruplas siguiendo simplemente el orden natural de estos números individuales.) Si pudiéramos establecer que este algoritmo *no* se para, entonces tendríamos una demostración de la tesis de Fermat.

De modo análogo, es posible parafrasear en términos del problema de la detención de una máquina de Turing muchos otros proble-

* Recuérdesse que por números *naturales* entendemos 0, 1, 2, 3, 4, 5, 6, ... La razón de escribir la ecuación en términos de « $x + 1$ » y « $w + 3$ », etc., en lugar de la forma más familiar ($x^w + y^w = z^w$; $x, y, z > 0, w > 2$) del enunciado de Fermat, es que estamos admitiendo que x, w , etc. puedan ser *cualquier* número natural, comenzando por el cero.

mas matemáticos no resueltos. Un ejemplo es la «conjetura de Goldbach», que afirma que cualquier número par mayor que 2 es la suma de dos números primos *. Decidir si un número natural dado es primo o no, es un proceso algorítmico puesto que basta con comprobar su divisibilidad por números menores que *él mismo*, una cuestión que sólo requiere un cálculo *finito*. Podemos imaginar una máquina de Turing que recorra los números pares 6, 8, 10, 12, 14,... ensayando todas las formas diferentes de descomponerlos en pares de números impares.

$$\begin{array}{lll} 6 = 3 + 3, & 8 = 3 + 5, & 10 = 3 + 7 = 5 + 5, \\ 12 = 5 + 7, & 14 = 3 + 11 = 7 + 1, & \dots \end{array}$$

y verificando que *cada* uno de estos números pares se descompone en *algún* par para el que *ambos* miembros son primos. (Evidentemente no necesitamos comprobar pares de sumandos *pares*, excepto $2 + 2$, puesto que todos los primos salvo el 2 son impares.) Nuestra máquina se para sólo cuando llega a un número par para el que *ninguno* de los pares de números en los que se descompone consta de dos primos. En tal caso tendríamos un contraejemplo de la conjetura de Goldbach, a saber, un número par (mayor que 2) que *no* es la suma de dos primos. Por lo tanto, si pudiéramos decidir si esta máquina de Turing se para o no alguna vez tendríamos una forma de decidir también la verdad o falsedad de la conjetura de Goldbach.

Surge naturalmente una pregunta: ¿cómo decidir si una máquina de Turing particular (a la que se le introduce algún input específico) se parará o no alguna vez? Para muchas máquinas de Turing esto no sería difícil de responder; pero, en ocasiones, como hemos visto más arriba, la respuesta podría implicar la solución de un problema matemático extraordinario. Por consiguiente, ¿existe algún procedimiento algorítmico para responder a la cuestión general —el problema de la detención— de forma completamente automática? Turing demostró que no existe.

Su argumento era esencialmente el siguiente. Supongamos primero que, por el contrario, existe tal algoritmo **. En tal caso debe

* Recordemos que los números *primos* 2, 3, 5, 7, 11, 13, 17,... son aquellos números naturales que sólo son divisibles por sí mismos y por la unidad. Ni el 0 ni el 1 se consideran primos.

** Éste es el procedimiento matemático familiar —y potente— conocido como *reductio ad absurdum*, en el que se supone primero que lo que se trata de demostrar es falso; si de ello se deriva una contradicción, queda establecido que el resultado deseado es realmente *verdadero*.

haber alguna máquina de Turing H que «decide» si la n -ésima máquina de Turing, cuando actúa sobre el número m , se parará o no finalmente. Digamos que su output es la cinta numerada 0 si no se para y 1 si lo hace:

$$H(n; m) = \begin{cases} 0 & \text{si } T_n(m) = \square \\ 1 & \text{si } T_n(m) \text{ se para} \end{cases}$$

Aquí, podríamos hacer que la codificación del par (n,m) siguiera la misma regla que adoptamos para la máquina universal U . Sin embargo, esto tropezaría con el problema técnico de que para algunos números n (v.g. $n = 7$) T_n no está correctamente especificada y el marcador 11110 sería inadecuado para separar n de m en la cinta. Para evitar este problema, supongamos que n se codifica usando la notación binaria *expandida* en lugar de la simple notación binaria, con m en forma binaria ordinaria como antes. Entonces el marcador 110 será suficiente para separar n de m . El uso del *punto y coma* en $H(n;m)$, a diferencia de la *coma* en $U(n,m)$, indica este cambio.

Imaginemos ahora una matriz infinita que lista todos los outputs de todas las posibles máquinas de Turing que actúan sobre todos los diferentes inputs posibles. La n -ésima fila de la matriz muestra el output de la n -ésima máquina de Turing cuando se aplica a los diversos inputs 0, 1, 2, 3, 4,...

$m \rightarrow$	0	1	2	3	4	5	6	7	8	...
n ↓										
0	□	□	□	□	□	□	□	□	□	...
1	0	0	0	0	0	0	0	0	0	...
2	1	1	1	1	1	1	1	1	1	...
3	0	2	0	2	0	2	0	2	0	...
4	1	1	1	1	1	1	1	1	1	...
5	0	□	0	□	0	□	0	□	0	...
6	0	□	1	□	2	□	3	□	4	...
7	0	1	2	3	4	5	6	7	8	...
8	□	1	□	□	1	□	□	□	1	...
.	.				.			.	...	
.	.				.			.		
.	.				.			.		
197	2	3	5	7	11	13	17	19	23	...
.	.				.			.		
.	.				.			.		
.	.				.			.		

En la tabla superior he hecho una pequeña trampa y no he listado las máquinas de Turing tal como están numeradas *realmente*. De haberlo hecho así hubiera dado lugar a una lista con la que resultaría muy engorroso trabajar puesto que todas las máquinas para las que n es menor que 11 no dan nada más que $\square$, y para $n = 11$ sólo dan 0. Para hacer la lista más interesante desde el principio he supuesto que se ha conseguido una codificación mucho más eficaz. En realidad he formado las entradas aleatoriamente sólo para dar una impresión de cuál podría ser su apariencia general.

No estoy pidiendo que realmente tengamos que *calcular* esta matriz mediante algún algoritmo. (De hecho no existe tal algoritmo, como veremos en un momento.) Simplemente se nos pide que *imaginemos* que la *verdadera* lista ha sido desplegada ¡tal vez por Dios! ante nosotros. Lo que nos causaría dificultades, si intentásemos calcular la matriz, sería la aparición de $\square$ pues no podríamos saber con certeza cuando colocar un $\square$ en alguna posición ya que esos cálculos ¡continuarían indefinidamente!

Sin embargo, *podríamos* dar un procedimiento de cálculo para generar la tabla si se nos permitiera utilizar nuestra H supuesta, pues H nos diría entonces dónde aparecen realmente los $\square$. En lugar de esto, utilicemos H para *eliminar* todos los $\square$ reemplazando cada aparición con un 0. Esto se consigue haciendo preceder la actuación de T_n sobre m por el cálculo $H(n;m) = 1$ (esto es, sólo si el cálculo $T_n(m)$ da realmente una respuesta), y escribiendo simplemente 0 si $H(n;m) = 0$ (esto es, si $T_n(m) = \square$). Podemos escribir nuestro nuevo procedimiento (es decir, el obtenido precediendo $T_n(m)$ por la actuación de $H(n;m)$) en la forma

$$T_n(m) \times H(n;m)$$

(Aquí estoy usando un convenio matemático común sobre el orden de las operaciones matemáticas: la que figura a la *derecha* debe ser la *primera* en realizarse. Nótese que, simbólicamente, tenemos $\square \times 0 = 0$.)

Nótese que, al suponer que H existe, las filas de la tabla de la página siguiente constan de *secuencias computables*. (Entiendo por secuencia computable una secuencia infinita cuyos sucesivos valores pueden ser generados por un algoritmo; es decir, existe alguna máquina de Turing que, cuando se aplica uno tras otro a los números naturales $m = 0, 1, 2, 3, 4, 5, \dots$, da los números sucesivos de la secuencia.) Ahora tomemos nota de dos hechos acerca de esta tabla. En primer lugar,

La tabla para esto se lee ahora

$m \rightarrow$	0	1	2	3	4	5	6	7	8	...
n ↓										
0	0	0	0	0	0	0	0	0	0	...
1	0	0	0	0	0	0	0	0	0	...
2	1	1	1	1	1	1	1	1	1	...
3	0	2	0	2	0	2	0	2	0	...
4	1	1	1	1	1	1	1	1	1	...
5	0	0	0	0	0	0	0	0	0	...
6	0	0	1	0	2	0	3	0	4	...
7	0	1	2	3	4	5	6	7	8	...
8	0	1	0	0	1	0	0	0	1	...
.	.				.				.	
.	.				.				.	
.	.				.				.	

toda secuencia computable de números naturales debe aparecer en alguna parte (quizá muchas veces) entre sus filas. Esta propiedad era ya válida para la tabla original con sus $\square$. Simplemente hemos *añadido* algunas filas para reemplazar las máquinas de Turing «inútiles» (esto es, las que producen al menos un $\square$). En segundo lugar, al hacer la suposición de que la máquina de Turing H existe realmente, la tabla ha sido *generada computablemente* (esto es, generada mediante algún algoritmo definido), a saber, mediante el procedimiento $T_n(m) \times H(n;m)$. Es decir, existe alguna máquina de Turing Q que cuando actúa sobre el par de números (n,m) produce la entrada apropiada en la tabla. Para esto, podemos codificar n y m en la cinta de Q de la misma forma que lo hacíamos para H , y tenemos

$$Q(n;m) = T_n(m) \times H(n;m).$$

Ahora aplicamos una variante de un ingenioso y potente artificio, el «corte diagonal» de Georg Cantor. (Veremos la versión original del corte diagonal de Cantor en el próximo capítulo.) Consideremos los elementos de la diagonal principal, marcados ahora con guarismos en **negrita**:

0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1
0	2	0	2	0	2	0	2	0
1	1	1	1	1	1	1	1	1
0	0	0	0	0	0	0	0	0
0	0	1	0	2	0	3	0	4
0	1	2	3	4	5	6	7	8
0	1	0	0	1	0	0	0	1
.				.				.
.				.				.
.				.				.

Estos elementos proporcionan cierta secuencia 0, 0, 1, 2, 1, 0, 3, 7, 1,... a cada uno de cuyos términos sumamos 1:

$$1, 1, 2, 3, 2, 1, 4, 8, 2, \dots$$

Éste es claramente un procedimiento computable y, dado que nuestra tabla estaba generada computablemente, nos proporciona una nueva secuencia computable, de hecho la secuencia $1 + Q(n;n)$, esto es,

$$1 + T_n(n) \times H(n;n)$$

(puesto que la diagonal se obtiene haciendo m igual a n). Ahora bien, nuestra tabla contiene *todas* las secuencias computables, de modo que nuestra nueva secuencia debe estar en alguna parte de la lista. ¡Pero esto es imposible! En efecto, nuestra nueva secuencia difiere de la primera fila en la primera entrada, de la segunda fila en la segunda entrada, de la tercera fila en la tercera entrada, y así sucesivamente. Esto es una contradicción manifiesta. Es esta contradicción la que establece lo que hemos estado tratando de demostrar, a saber: ¡que en realidad no existe la máquina de Turing H ! *No existe un algoritmo universal para decidir si una máquina de Turing se parará o no.*

Otra manera de parafrasear este argumento consiste en señalar que, suponiendo que H exista, hay algún número de máquina de Turing, digamos k , para el algoritmo $1 + Q(n;n)$, de modo que tenemos

$$1 + T_n(n) \times H(n;n) = T_k(n).$$

Pero si sustituimos $n = k$ en esta relación (¡proceso diagonal!) obtenemos

$$1 + T_k(k) \times H(k;k) = T_k(k).$$

Esto es una contradicción porque si $T_k(k)$ se para obtenemos la relación imposible

$$1 + T_k(k) = T_k(k)$$

(puesto que $H(k;k) = 1$), mientras que si $T_k(k)$ no se para (de modo que $H(k;k) = 0$) tenemos la igualmente inconsistente

$$1 + 0 = \square$$

La pregunta de si una máquina de Turing concreta se para o no, es un problema matemático perfectamente bien definido (y ya hemos visto que, recíprocamente, varias cuestiones matemáticas importantes pueden ser enunciadas en términos de parada de máquinas de Turing). De este modo, al demostrar que no existe ningún algoritmo para decidir la cuestión de la parada de las máquinas de Turing, Turing demostró (como lo hiciera Church utilizando su propio tipo de aproximación bastante diferente) que no puede haber un algoritmo general para decidir cuestiones matemáticas. ¡El *Entscheidungsproblem* de Hilbert no tiene solución!

Esto no quiere decir que en algún caso *individual* no seamos capaces de decidir la verdad, o la falsedad, de alguna cuestión matemática concreta; o decidir si una máquina de Turing dada se parará o no. Ejercitando el ingenio, o incluso el simple sentido común, podemos decidir semejante cuestión en un caso dado. (Por ejemplo, si una lista de instrucciones de una máquina de Turing no contiene una orden STOP, o contiene solamente órdenes STOP, entonces ¡el solo sentido común es suficiente para decirnos si se parará o no!) Pero no hay ningún algoritmo que funcione para *todas* las cuestiones matemáticas, ni para *todas* las máquinas de Turing y todos los números sobre los que podría actuar.

Pudiera parecer que hemos establecido con esto que hay al menos *algunas* cuestiones matemáticas indecidibles. ¡Nada de eso! No hemos demostrado que exista alguna tabla de máquina de Turing especialmente complicada para la que, en un sentido absoluto, sea imposible decidir si la máquina se para o no cuando se le introduce algún número especialmente complicado; más bien, casi hemos demostrado lo contrario, como veremos en un momento. No hemos dicho nada sobre la insolubilidad de problemas *individuales* sino sólo sobre la insolubilidad *algorítmica* de *familias* de problemas. En cualquier caso individual la respuesta es «sí» o «no», de modo que ciertamente *hay* un algoritmo para decidir este caso concreto, a saber: ¡el algoritmo que

simplemente dice «sí» cuando se le plantea el problema, o el que simplemente dice «no», según sea el caso! La dificultad es, por supuesto, que no podemos saber cuál de estos algoritmos usar. Éste es un problema de decidir la verdad matemática de un enunciado individual, y no de la decisión sistemática para una familia de enunciados. Es importante darse cuenta de que los algoritmos no deciden, por sí mismos, la verdad matemática. La *validez* de un algoritmo debe establecerse siempre por medios externos.

¿Cómo superar a un algoritmo?

Volveremos más tarde a la cuestión de decidir la verdad de enunciados matemáticos, cuando tratemos el teorema de Gödel (ver Capítulo 4). De momento quiero señalar que el argumento de Turing es en verdad mucho más constructivo y menos negativo de lo que haya podido parecer hasta ahora. Ciertamente no hemos mostrado una máquina de Turing específica para la que, en algún sentido absoluto, sea indecidible si se para o no. ¡En realidad, si examinamos cuidadosamente el argumento, descubrimos que nuestro propio procedimiento nos ha dado implícitamente la respuesta para las, en apariencia, «especialmente complicadas» máquinas que construimos utilizando el procedimiento de Turing!

Veamos de donde viene esto. Supongamos que tenemos algún algoritmo que es a veces efectivo para decirnos cuándo no se parará una máquina de Turing. El procedimiento de Turing, como se esbozó más arriba, mostrará *explícitamente* un cálculo de una máquina de Turing para el que ese particular algoritmo no es capaz de decidir si el cálculo se para o no. ¡Sin embargo, al hacerlo así, realmente *nos* permite ver la respuesta en este caso! El cálculo de la máquina de Turing particular que mostramos *no* se parará.

Para ver cómo surge esto en detalle supongamos que tenemos semejante algoritmo que es a veces efectivo. Como antes, denotamos este algoritmo por H , pero ahora permitimos que el algoritmo pueda no estar siempre en lo cierto al decirnos que una máquina de Turing no se parará:

$$H(n;m) = \begin{cases} 0 \text{ o } \square & \text{si } T_n(m) = \square \\ 1 & \text{si } T_n(m) \text{ se para} \end{cases}$$

de modo que existe la posibilidad $H(n;m) = \square$ cuando $T_n(m) = \square$. Existen muchos de estos algoritmos $H(n;m)$. Por ejemplo, $H(n;m)$

podría simplemente producir un 1 en cuanto $T_n(m)$ se pare, ¡aunque éste particular algoritmo apenas sería de uso práctico!)

Podemos completar el procedimiento de Turing en detalle igual que antes, excepto que en lugar de reemplazar *todos* los $\square$ por 0 dejamos ahora algunos $\square$. Como antes, nuestro procedimiento diagonal nos ha dado

$$1 + T_n(n) \times H(n;n)$$

como el n -ésimo término de la diagonal. (Obtendremos un $\square$ cuandoquiera que $H(n;n) = 0$. Nótese que $\square \times \square = \square$, $1 + \square = \square$.) Ésta es una computación perfectamente válida, de modo que es lograda por alguna máquina de Turing, digamos la k -ésima, y así *tenemos*

$$1 + T_n(n) \times H(n;n) = T_k(n).$$

Consideremos el k -ésimo término diagonal, esto es $n = k$, y obtenemos

$$1 + T_k(k) \times H(k;k) = T_k(k).$$

Si la computación $T_k(k)$ se para tenemos una contradicción (puesto que se supone que $H(k;k)$ es 1 cuandoquiera que $T_k(k)$ se para, y la ecuación da entonces una inconsistencia: $1 + T_k(k) = T_k(k)$). Así que $T_k(k)$ no puede pararse, es decir

$$T_k(k) = \square.$$

Pero el algoritmo no puede «saber» esto, ya que si diera $H(k;k) = 0$, tendríamos de nuevo una contradicción (simbólicamente, tendríamos la relación no válida: $1 + 0 = \square$).

Por consiguiente, si podemos hallar k sabremos cómo reconstruir nuestro cálculo específico para derrotar al algoritmo ¡pero para el que *nosotros* sabemos la respuesta! ¿Cómo hallar k ? Ésta es difícil tarea. Lo que tenemos que hacer es observar en detalle la construcción de $H(n;m)$ y de $T_n(m)$ y luego ver en detalle cómo actúa $1 + T_n(n) \times H(n;n)$ como máquina de Turing. Hallamos el número de esta máquina de Turing, que es k . Esto sería ciertamente complicado de llevar a cabo en detalle pero puede hacerse *. Debido a la complejidad no estaríamos en absoluto interesados en el cálculo $T_k(k)$ ¡de no ser por el hecho de que lo hemos producido especialmente para derrotar al algoritmo H ! Lo importante es que tenemos un pro-

* De hecho, ya se ha conseguido lo más difícil mediante la construcción anterior de la máquina universal de Turing U , puesto que ello nos posibilita escribir $T_n(n)$ como una máquina de Turing actuando sobre n .

cedimiento bien definido, cualquiera que sea el H que nos dan, para hallar el correspondiente k para el que *nosotros* sabemos que $T_k(k)$ derrota a H , y para el que podemos superar al algoritmo. ¡Quizá nos consuele un poco el pensar que somos mejores que los simples algoritmos!

De hecho, el procedimiento está tan bien definido que podríamos hallar un *algoritmo* para generar k , dado H . Por lo tanto, antes de quedarnos demasiado satisfechos, tenemos que darnos cuenta de lo que *este* algoritmo puede mejorar ⁹ en H , ya que, en efecto, él «sabe» que $T_k(k) = \square$ —¿o no lo sabe? Nos ha sido de ayuda en la descripción anterior el utilizar el término antropomorfo «saber» con referencia a un algoritmo. Sin embargo, ¿no somos *nosotros* quienes estamos «sabiendo», mientras que el algoritmo tan sólo sigue las reglas que le hemos dicho que siga? ¿O estamos nosotros mismos simplemente siguiendo reglas para cuyo seguimiento estamos programados por la construcción de nuestros cerebros y de nuestro entorno? Este asunto no es simplemente una cuestión de algoritmos sino también una cuestión acerca de cómo juzgamos lo que es cierto y lo que es falso. Éstos son temas capitales a los que tendré que volver más adelante. La cuestión de la verdad matemática (y su naturaleza no algorítmica) será estudiada en el Capítulo 4. Cuando menos, ahora tendremos alguna noción sobre los *significados* de los términos «algoritmo» y «computabilidad», y una comprensión de algunos de los temas relacionados.

El cálculo lambda de Church

El concepto de computabilidad es una idea matemática muy importante y bella. Es también notablemente reciente —si se tiene en cuenta su naturaleza fundamental para las matemáticas— ya que fue desarrollado por primera vez en los años 30. Es una idea que atraviesa *todas* las áreas de la matemática (aunque es bastante cierto que la mayoría de los matemáticos, por ahora, no se preocupan muy a menudo por cuestiones de computabilidad). La potencia de la idea reside en parte en el hecho de que algunas operaciones bien definidas en matemáticas *no* son computables (como la parada, o no, de una máquina de Turing; veremos otros ejemplos en el Capítulo 4). En efecto, si no existieran estas cosas no computables el concepto de computabilidad no tendría mucho interés matemático. Los matemáticos, después de todo, aman los rompecabezas. Puede ser un rompecabezas intrigante para ellos el decidir, en relación con alguna opera-

ción matemática, si es computable o no. Es especialmente intrigante debido a que ¡la solución general a *dicho* rompecabezas es en sí mismo no computable!

Una cosa debe quedar clara. El concepto de computabilidad es un concepto matemático verdaderamente «absoluto». Es una idea abstracta que queda más allá de cualquier realización concreta en términos de máquinas de Turing tal como las he descrito. Como he resalado antes, no necesitamos ligar ningún significado concreto a las «cintas» y «estados internos», etc., que caracterizan la ingeniosa aunque particular aproximación de Turing. Existen también otras formas de expresar la idea de computabilidad, de las que la primera históricamente fue el notable «cálculo lambda» del lógico americano Alonzo Church con la ayuda de Stephen C. Kleene. El método de Church era bastante diferente e indudablemente más abstracto que el de Turing. En efecto, en la forma que Church estableció sus ideas, apenas existe una conexión obvia entre estas y cualquier cosa que pudiéramos llamar «mecánica». La idea clave que subyace tras el método de Church es verdaderamente *abstracta* en su misma esencia: una operación matemática que Church, de hecho, llamaba «abstracción».

Creo que merece la pena dar una breve descripción del esquema de Church no sólo porque hace énfasis en que la computabilidad es una idea matemática, independiente de cualquier concepto particular sobre una máquina de computar, sino también porque ilustra la potencia de las ideas abstractas en matemáticas. El lector que no esté versado en ideas matemáticas, ni intrigado por propio gusto por tales cosas, puede, en este punto, saltar al siguiente capítulo sin que haya una pérdida significativa en el curso del argumento. De todas formas, creo que tales lectores podrían obtener provecho acompañándome un poco más y siendo testigos de la mágica economía del esquema de Church (ver Church 1941).

En este esquema estamos interesados en un «universo» de objetos, denotados, por ejemplo, por

$$a, b, c, d, \dots, z, a', b', \dots, z', a'', b'', \dots, a''', \dots, a'''' , \dots$$

cada uno de los cuales representa una operación matemática o *función*. (La razón para las letras con prima es simplemente permitir una cantidad ilimitada de símbolos para denotar tales funciones.) Los «argumentos» de estas funciones —es decir, las cosas sobre las que estas funciones actúan— son otras cosas del mismo tipo, esto es, también funciones. Además, el resultado (o «valor») de una función actuando sobre otra es de nuevo una función. (Hay ciertamente una maravillosa

economía de conceptos en el sistema de Church.) Así, cuando escribimos *

$$a = bc$$

queremos decir que el resultado de la función b actuando sobre la función c es otra función a . No hay ninguna dificultad para expresar la idea de una función de dos o más variables en este esquema. Si queremos pensar en f como una función de dos variables p y q , por ejemplo, podemos escribir simplemente

$$(fp) q$$

(que es el resultado de la función fp aplicada a q). Para una función de tres variables consideramos

$$((fp)q)r,$$

y así sucesivamente.

Ahora llega la poderosa operación de *abstracción*. Para ésta utilizamos la letra griega λ (lambda) seguida inmediatamente por una letra que representa una de las funciones de Church, digamos x , que consideramos como una «variable muda». Toda aparición de la variable x en la expresión entre corchetes que sigue inmediatamente es considerada simplemente como un «hueco» en el que se puede sustituir cualquier cosa que siga a la expresión completa. Así, si escribimos

$$\lambda x.[fx],$$

queremos expresar la función que cuando actúa sobre, pongamos por caso, a produce el resultado fa . Es decir

$$(\lambda x.[fx])a = fa.$$

En otras palabras, $\lambda x.[fx]$ es simplemente la función f , es decir

$$\lambda x.[fx] = f.$$

Esto requiere un poco de reflexión. Es una de esas sutilezas matemáticas que parece tan pedante y trivial a primera vista que uno tiende a pasar por alto el punto esencial. Consideremos un ejemplo tomado de las matemáticas familiares de la escuela. Sea la función f

* Una forma más familiar de notación hubiera consistido en escribir $a = b(c)$, por ejemplo, pero estos paréntesis no son realmente necesarios y es mejor que los omitamos. El incluirlos de forma consistente nos llevaría a fórmulas bastante farragosas como $(f(p))(q)$ y $((f(p))(q))(r)$, en lugar de $(fp)q$ y $((fp)q)r$ respectivamente.

la operación matemática de tomar el seno de un ángulo, de modo que la función abstracta «sen» se define por

$$\lambda x.[\text{sen } x] = \text{sen}.$$

(No se preocupe por cómo la «función» x puede ser considerada como un ángulo. Pronto veremos algo sobre cómo los números pueden ser tomados como funciones; y un ángulo es sólo un tipo de número.) Hasta aquí, esto *es* más bien trivial. Pero imaginemos que la notación «sen» no hubiera sido inventada, aunque conocemos el desarrollo en serie de potencias para $\text{sen } x$:

$$x - \frac{1}{6} x^3 + \frac{1}{120} x^5 - \dots$$

Entonces podríamos definir

$$\text{sen} = \lambda x.[x - \frac{1}{6} x^3 + \frac{1}{120} x^5 - \dots].$$

Nótese que, de forma aún más simple, podríamos definir, pongamos por caso, la operación de «un sexto del cubo» para la que no hay notación «funcional» estándar:

$$Q = \lambda x.[\frac{1}{6}x^3]$$

y encontrar por ejemplo

$$Q(a + 1) = \frac{1}{6}(a + 1)^3 = \frac{1}{6} a^3 + \frac{1}{2} a^2 + \frac{1}{2} a^2 + \frac{1}{6}$$

Más pertinentes para la discusión actual serían expresiones formadas simplemente a partir de las operaciones funcionales elementales de Church, tales como

$$\lambda f.[f(fx)].$$

Ésta es la función que, cuando actúa sobre otra función, digamos g , produce g iterada dos veces actuando sobre x , es decir

$$(\lambda f.[f(fx)])g = g(gx)$$

También podríamos haber «abstraído fuera» primero la x para obtener

$$\lambda f. [\lambda x. [f(f\ x)]]$$

que podemos abreviar

$$\lambda fx. [f(fx)].$$

Ésta es la operación que, cuando actúa sobre g , produce la función « g iterada dos veces». De hecho ésta precisamente es la función que Church identifica con el número natural 2:

$$\mathbf{2} = \lambda fx. [f(fx)],$$

de modo que $(\mathbf{2}g)y = g(gy)$. Análogamente define:

$$\mathbf{3} = \lambda fx. [f(f(fx))], \quad \mathbf{4} = \lambda fx. [f(f(f(fx)))], \quad \text{etc.}$$

junto con

$$\mathbf{1} = \lambda fx. [fx], \quad \mathbf{0} = \lambda fx. [x]$$

Realmente el « $\mathbf{2}$ » de Church se parece más a «doble» y su « $\mathbf{3}$ » a «triple», etc. Así, la acción de $\mathbf{3}$ sobre una función f , a saber $\mathbf{3}f$, es la operación «iterar f tres veces». La acción de $\mathbf{3}f$ sobre y sería entonces $(\mathbf{3}f)y = f(f(f(y)))$.

Veamos ahora cómo puede expresarse en el sistema de Church una operación aritmética muy sencilla, a saber, la operación de añadir $\mathbf{1}$ a un número. Definamos

$$S = \lambda abc. [b((ab)c)]$$

Para ilustrar el hecho de que S añade simplemente $\mathbf{1}$ a un número descrito en la notación de Church, comprobémoslo en:

$$\begin{aligned} S\mathbf{3} &= \lambda abc. [b((ab)c)] \quad \mathbf{3} = \lambda bc. [b((\mathbf{3}\ b)c)] \\ &= \lambda bc. [b(b(b(bc)))] = \mathbf{4} \end{aligned}$$

puesto que $(\mathbf{3}b)c = b(b(bc))$. Evidentemente esto también se aplica a cualquier otro número natural. (De hecho $\lambda abc. [(ab)(bc)]$ hubiera hecho lo mismo que S .)

¿Cómo sería la multiplicación de un número por dos? Esta duplicación se lograría mediante

$$D = \lambda abc. [(ab)((ab)c)],$$

que de nuevo se ilustra por su actuación sobre **3**:

$$\begin{aligned} D &= \lambda abc.[(ab)((ab)c)] \mathbf{3} = \lambda bc.[(\mathbf{3}b)((\mathbf{3}b) c)] \\ &= \lambda bc.[(\mathbf{3}b)(b(b(bc))))] = \lambda bc.[b(b(b(b(b(bc))))))] = \mathbf{6} \end{aligned}$$

De hecho, las operaciones aritméticas básicas de adición, multiplicación y elevación a una potencia pueden definirse, respectivamente, por:

$$\begin{aligned} A &= \lambda fgxy.[((fx) (gx))y], \\ M &= \lambda fgx.[f(gx)], \\ P &= \lambda fg.[fg]. \end{aligned}$$

La lectora o el lector pueden ocuparse en convencerse por sí mismos —o por cualquier otro en quien tengan plena confianza— de que efectivamente

$$(A\mathbf{m})\mathbf{n} = \mathbf{m} + \mathbf{n}, \quad (M\mathbf{m})\mathbf{n} = \mathbf{m} \times \mathbf{n}, \quad (P\mathbf{m})\mathbf{n} = \mathbf{n}^{\mathbf{m}},$$

donde **m** y **n** son funciones de Church para dos números naturales, **m** + **n** es la función de Church para su suma, y así para las demás. La última de éstas es la más sorprendente. Comprobémosla para el caso **m** = **2**, **n** = **3**:

$$\begin{aligned} (P\mathbf{2})\mathbf{3} &= ((\lambda fg.[fg])\mathbf{2})\mathbf{3} = (\lambda g.[\mathbf{2}g])\mathbf{3} \\ &= (\lambda g.[\lambda fx.[f(fx)]g])\mathbf{3} = \lambda gx.[g(gx)]\mathbf{3} \\ &= \lambda x.[\mathbf{3}(\mathbf{3}x)] = \lambda x.[\lambda fy[f(f(fy))](\mathbf{3}x)] \\ &= \lambda xy.[(\mathbf{3}x)((\mathbf{3}x)((\mathbf{3}x)y))] \\ &= \lambda xy.[(\mathbf{3}x)((\mathbf{3}x)(x(x(xy))))] \\ &= \lambda xy.[(\mathbf{3}x)(x(x(x(x(xy)))))] \\ &= \lambda xy.[x(x(x(x(x(x(xy))))))] = \mathbf{9} = \mathbf{3}^2 \end{aligned}$$

Las operaciones de substracción y división no se definen tan fácilmente (y necesitamos efectivamente algún convenio sobre lo que hacer con «**m** − **n**» cuando **m** es menor que **n** y con **m** ÷ **n** cuando **m** no es divisible por **n**). ¡De hecho, un hito importante en la materia sucedió a comienzos de los años 30 cuando Kleene descubrió la forma de expresar la operación de substracción dentro del esquema de Church! A continuación siguieron otras operaciones. Finalmente, en 1937, Church y Turing, independientemente, demostraron que cualquier operación computable (o algorítmica) —ahora en el sentido de las máquinas de Turing— puede lograrse en términos de una de las expresiones de Church (y viceversa).

Éste es un hecho verdaderamente notable, y sirve para subrayar

el carácter fundamentalmente objetivo y matemático de la noción de computabilidad. La noción de computabilidad de Church tiene, a primera vista, muy poco que ver con las máquinas computadoras. Pese a todo tiene, en cualquier caso, ciertas relaciones fundamentales con la computación práctica. En particular, el potente y flexible lenguaje LISP incorpora, de un modo esencial, la estructura básica del cálculo de Church.

Como indiqué antes, hay también otras maneras de definir la noción de computabilidad. El concepto de Post de máquina computadora estaba muy próximo al de Turing y fue presentado independientemente y casi al mismo tiempo. Existía también una definición bastante más manejable de computabilidad (recursividad) debida a J. Herbrand y Godel. H. B. Curry en 1929, y también M. Schonfinkel en 1924, tuvieron un poco antes una aproximación diferente, a partir de la que se desarrolló en parte el cálculo de Church. (Ver Gandy 1988). Las aproximaciones modernas a la computabilidad (tales como la de una *máquina de registro ilimitado*, descrita en Cutland 1980) difieren considerablemente en los detalles respecto a la original de Turing, y son bastante más prácticas. No obstante, el *concepto* de computabilidad sigue siendo el mismo, cualquiera que sea la aproximación que se adopte.

Al igual que sucede con muchas otras ideas matemáticas, en especial las más profundamente bellas y fundamentales, la idea de computabilidad parece tener una especie de *realidad platónica* autónoma. En los dos próximos capítulos tendremos que volver a esta misteriosa cuestión de la realidad platónica de los conceptos matemáticos en general.

3. Matemáticas y realidad

La tierra de Tor'Bled-Nam

Imaginemos que hemos realizado un largo viaje a algún mundo lejano. Llamaremos a este mundo Tor'Bled-Nam. Nuestros sensores de detección remota han captado una señal que se muestra ahora en una pantalla frente a nosotros. La imagen se focaliza y vemos (Fig. 3.1.):



Fig. 3.1. Una primera impresión de un mundo extraño.

¿Qué puede ser? ¿Es algún insecto de extraña apariencia? Tal vez, en lugar de ello, sea un lago de color oscuro con muchas cadenas montañosas que se adentran en él. ¿O podría ser una inmensa ciudad extraterrestre con una forma extraña y con carreteras que salen en varias direcciones hacia las pequeñas ciudades y pueblos vecinos? Quizá sea una isla; tratemos entonces de descubrir si existe algún continente próximo al que está asociada. Podemos hacer esto «retrocediendo», reduciendo la amplificación de nuestro dispositivo sensor en un factor lineal de aproximadamente quince. Dicho y hecho, el mundo completo surge a la vista (Fig. 3.2.):

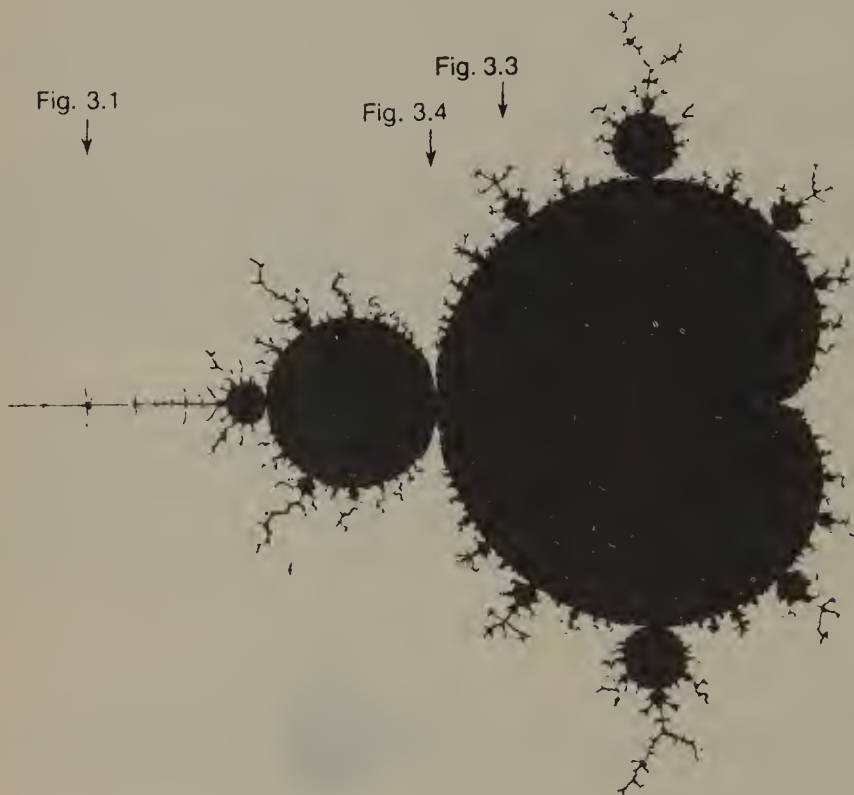


Fig. 3.2. «Tor'Bled-Nam» en su totalidad. Las localizaciones de las ampliaciones que se muestran en las Figuras 3.1, 3.3. y 3.4. son las indicadas bajo las flechas.

Nuestra «isla» se ve como un punto pequeño indicado bajo «Figura 3.1.» en la Figura 3.2. Todos los filamentos (¿ríos, carreteras, puentes?) que parten de la isla original tienen un extremo final, con la

excepción del que sale del interior de la grieta de su lado derecho que acaba por unirse a un objeto mucho mayor que vemos representado en la Figura 3.2. Este objeto mayor es claramente similar a la isla que vimos primero, aunque no es exactamente igual. Si examinamos con mayor atención lo que parece ser la línea costera de este objeto vemos innumerables protuberancias —redondeadas, si bien ellas mismas poseen sus propias protuberancias similares. Cada pequeña protuberancia parece estar unida a otra mayor en alguna región minúscula, dando lugar a una serie de verrugas sobre otras verrugas. A medida que la imagen se hace más nítida, vemos miríadas de pequeñísimos filamentos que emanan de la estructura. Los propios filamentos se ramifican en varios lugares y a menudo serpentean de forma imprevisible. En ciertos puntos de los filamentos parecen apreciarse pequeños nudos de más complejidad que nuestro dispositivo sensor, con su amplificación actual, no puede resolver. Evidentemente el objeto no es una isla o continente real, ni ningún tipo de paisaje. Tal vez, después de todo, estamos viendo algún monstruoso escarabajo, y lo primero que vimos era parte de su prole, unida todavía a él por algún tipo de cordón umbilical.

Tratemos de examinar la naturaleza de una verruga de nuestra criatura, aumentando la amplificación de nuestro dispositivo sensor en un factor lineal de aproximadamente diez (Fig. 3.3. —cuya localización está indicada bajo «Fig. 3.3» en la Fig. 3.2.). La propia verruga tiene un fuerte parecido con la criatura global —excepto sólo en el punto de unión. Nótese que hay varios lugares en la Figura 3.3 en donde se juntan cinco filamentos. Hay quizá una cierta «quincuenería» en esta verruga particular (como habría una «ternariedad» en la verruga superior). De hecho, si examináramos la siguiente verruga de tamaño apreciable, hacia la parte inferior izquierda en la Figura 3.2, descubriríamos una «septenariedad» en ella; y una «nonanariedad» en la siguiente, y así sucesivamente. A medida que nos adentramos en la grieta entre las dos regiones mayores de la Figura 3.2, encontramos verrugas a la derecha caracterizadas por números impares que incrementan de dos en dos. Examinemos más profundamente esta grieta, aumentando la ampliación de la Figura 3.2 en un factor de unos diez (Fig. 3.4). Vemos otras numerosas verrugas minúsculas y también una agitada actividad. A la derecha, podemos distinguir algunas pequeñas «colas de caballo de mar» espirales —en un área que llamaremos «valle de los caballos de mar». Aquí encontraremos, si la amplificación se aumenta lo suficiente, varias «anémonas de mar» o regiones con una apariencia floral diferente. Tal vez, después de



Fig. 3.3. Una verruga con una «quinquenariedad» en sus filamentos.

todo, se trate realmente de alguna exótica línea costera —quizá algún arrecife coralino, con vida de todo tipo. Lo que pudiera haber parecido una flor revelará, en una posterior ampliación, estar compuesto de miríadas de pequeñísimas aunque increíblemente complicadas estructuras, cada una con numerosos filamentos y colas espirales retorcidas. Examinemos con algún detalle una de las mayores colas de caballo de mar, a saber, la que es discernible donde se indica como «Figura 3.5» en la Figura 3.4 (que está unida a una verruga con una ¡«29-riedad»!). Con una posterior ampliación de 250 aumentos se nos presenta la espiral mostrada en la Figura 3.5. Descubrimos que ésta no es una cola ordinaria sino que ella misma está formada por más complicados remolinos hacia uno u otro lado, con innumerables espirales minúsculas y regiones que semejan pulpos y caballos de mar.

En muchos lugares la estructura está entrelazada precisamente donde se juntan dos espirales. Examinemos uno de estos lugares (indica-

do bajo «Fig. 3.6» en la Fig. 3.5) incrementando nuestra amplificación en un factor de aproximadamente treinta. ¡Ya está!: ¿distinguimos un objeto extraño, aunque ahora familiar, en el centro? Una amplificación posterior en un factor de seis (Fig. 3.7) revela una minúscula criatura bebé —¡casi idéntica a la estructura global que hemos estado examinando! Si miramos más de cerca vemos que los filamentos que emanan de ella difieren un poco de los de la estructura principal, y se retuercen y extienden hasta distancias relativamente mucho mayores. Pero la propia criatura minúscula apenas parece diferir en absoluto de su progenitor, hasta el grado de poseer incluso su propia prole en posiciones estrechamente análogas. Podríamos examinarlas una vez más si siguiésemos aumentando la amplificación. Los nietos se parecerán también a su ancestro común —y ya nos sentimos dispuestos a creer que esto continúa indefinidamente. Podemos explorar



← Fig.
3.5

Fig. 3.4. La grieta principal. El «valle de los caballos de mar» es apenas perceptible en la parte inferior derecha.

Fig. 3.6



Fig. 3.5. Un primer plano de una cola de caballo de mar.

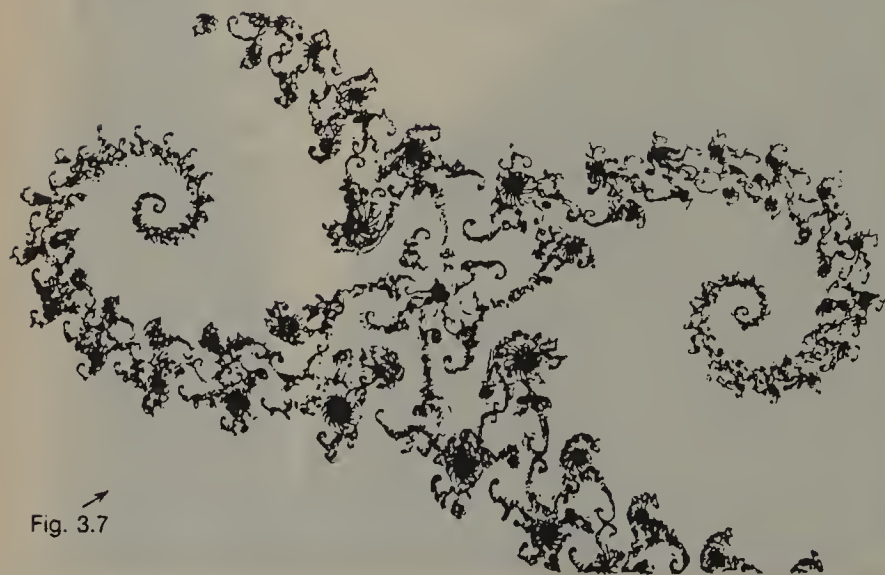


Fig. 3.7

Fig. 3.6. Una ampliación adicional de un punto de unión en el que se juntan dos espirales. Un pequeño bebé es apenas visible en el punto central.

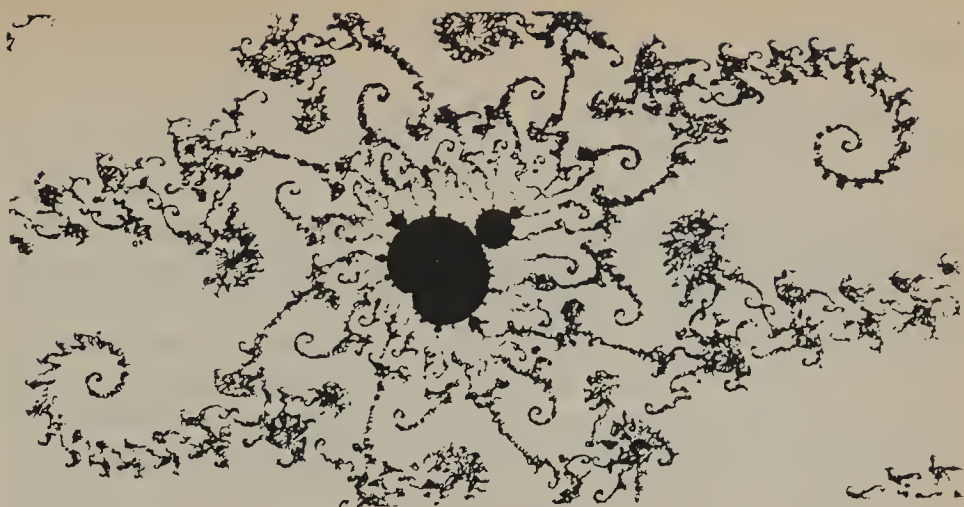


Fig. 3.7. En una nueva ampliación se aprecia que el bebé se asemeja mucho al mundo global.

este extraordinario mundo de Tor'Bled-Nam hasta donde deseemos, ajustando nuestro dispositivo sensor a mayores y mayores grados de amplificación. Encontramos una variedad sin fin: no hay dos regiones que sean exactamente iguales, pero hay un aire general al que pronto nos acostumbramos. La ahora familiar criatura escarabajoides emerge en escalas más y más pequeñas. En todo momento las estructuras filamentosas vecinas difieren de las que habíamos visto antes, y se nos presentan fantásticas escenas nuevas de increíble complicación.

¿Qué es esta extraña, variada y maravillosamente intrincada tierra con la que hemos topado? Sin duda muchos lectores ya lo sabrán. Pero otros no. Este mundo no es más que un objeto de la matemática abstracta: el conjunto conocido como conjunto de Mandelbrot ¹. Es complicado, sin duda; ¡pero está generado por una regla de notable simplicidad! Para explicar la regla adecuadamente tendré que explicar antes lo que es un *número complejo*. Eso es lo que voy a hacer ahora. Necesitaremos los números complejos más adelante. Son absolutamente fundamentales para la estructura de la mecánica cuántica y son, por consiguiente, básicos para el funcionamiento del propio mundo en que vivimos. También constituyen uno de los Grandes Milagros de la Matemática. Para explicar lo que es un número complejo necesitare antes recordar al lector lo que significa el término «número real». ¡Será útil, también, señalar la relación entre dicho concepto y la propia realidad del «mundo real»!

Números reales

Recordemos que los números *naturales* son las cantidades enteras:

$$0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, \dots$$

Son los más elementales y básicos entre los diferentes tipos de números. Cualquier entidad discreta puede cuantificarse mediante el uso de números naturales: podemos hablar de veintisiete ovejas en un prado, de dos relámpagos, doce noches, mil palabras, cuatro conversaciones, cero ideas nuevas, un error, seis ausentes, dos cambios de dirección, etc. Los números naturales pueden sumarse o multiplicarse para dar nuevos números naturales. Son los objetos de nuestra discusión general de los algoritmos, como la que se dio en el capítulo precedente.

Sin embargo, algunas operaciones importantes pueden llevarnos fuera del dominio de los números naturales; la resta es la más sencilla de éstas. Para definir la resta de una forma sistemática necesitamos los números *negativos*; podemos establecer, para este propósito, el sistema global de los *enteros*

$$\dots, -6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7, \dots$$

Ciertas cosas, como la carga eléctrica, los balances bancarios o las fechas * se cuantifican mediante números de este tipo. Pese a todo, estos números son aún de alcance demasiado limitado puesto que nos quedaríamos bloqueados cuando tratáramos de *dividir* un número por otro. En consecuencia, necesitaremos las *fracciones*, o *números racionales* como son llamados.

$$0, 1, -1, 1/2, -1/2, 2, -2, 3/2, -3/2, 1/3, \dots$$

Esto es suficiente para las operaciones de la aritmética finita, pero para muchos otros propósitos necesitamos ir más allá e incluir operaciones infinitas o de paso al límite. Por ejemplo, la familiar —y de gran importancia en matemáticas— cantidad π aparece en muchas de estas expresiones infinitas. En particular, tenemos

$$\pi = 2 \left\{ (2/1) (2/3) (4/3) (4/5) (6/5) (6/7) (8/7) (8/9) \dots \right\}$$

* En realidad, las convenciones usuales sobre fechas no se ajustan totalmente a esto ya que se omite el año cero.

y

$$\pi = 4(1 - 1/3 + 1/5 - 1/7 + 1/9 - 1/11 + \dots).$$

(Éstas son expresiones famosas, habiendo sido encontrada la primera por el matemático, gramático y experto criptógrafo inglés John Wallis en 1665; y la segunda por el matemático y astrónomo escocés (e inventor del primer telescopio reflector) James Gregory en 1671.) Como sucede con π , los números definidos de esta forma *no* deben ser racionales (esto es, no de la forma n/m , en donde n y m son enteros con m distinto de cero). El sistema de números necesita ser *ampliado* para poder incluir tales cantidades.

Este sistema ampliado de números se denomina sistema de los números «reales» —aquellos números familiares que pueden representarse como *expansiones decimales* infinitas, tales como:

$$-583.70264439121009538\dots$$

En términos de una representación semejante tenemos la bien conocida expresión para π :

$$\pi = 3.14159265358979323846\dots$$

Entre los tipos de números que también pueden representarse de este modo están las raíces cuadradas (o las raíces cúbicas o las raíces cuartas, etc.) de números racionales positivos, tales como

$$\sqrt{2} = 1.41421356237309504\dots$$

o, de hecho, la raíz cuadrada (o raíz cúbica, etc.) de cualquier número real positivo, como sucede con la expresión para π encontrada por el gran matemático suizo Leonhard Euler:

$$\pi = \sqrt{\{ 6(1 + 1/4 + 1/9 + 1/25 + 1/36 + \dots) \}}.$$

Los números reales son, en efecto, el tipo familiar de números con los que tenemos que trabajar en la vida de cada día, aunque normalmente estamos interesados sólo en aproximaciones a tales números y preferimos trabajar con expansiones que incluyen sólo algunas cifras decimales. Sin embargo, en las proposiciones matemáticas los números reales tienen que ser especificados *exactamente*, y necesitamos algún tipo de descripción infinita como pueda ser una completa expansión decimal infinita, o quizá alguna otra expresión matemática infinita como las fórmulas anteriores para π dadas por Wallis, Gregory o Euler. (Normalmente utilizaré expansiones decimales en

mis descripciones, pero sólo porque resultan más familiares. Para un matemático hay varias maneras bastante más satisfactorias de presentar los números reales pero no necesitamos preocuparnos por eso ahora.)

Podría dar la impresión de que es imposible considerar una expansión infinita *completa* pero no es así. Un ejemplo sencillo en el que podemos considerar claramente la secuencia completa es

$$1/3 = 0.3333333333333333...$$

donde los puntos nos indican que la serie de los 3s continúa indefinidamente. Para considerar esta expansión, todo lo que necesitamos saber es que la expansión continúa de la misma forma indefinidamente con 3s. Todo número racional tiene una expansión decimal repetida (o finita), tal como

$$93/74 = 1.2567567567567567567...$$

donde la secuencia 567 se repite indefinidamente, y ésta también puede ser considerada en su totalidad. También, la expresión

$$0.22000222200000222222000000022222220...$$

que define un número *irracional*, puede ser considerada en su totalidad (las cadenas de 0s y 2s incrementan su longitud en uno cada vez) y pueden darse muchos ejemplos similares. En cada caso nos damos por satisfechos cuando conocemos una regla con arreglo a la que se construye la expansión. Si hay algún algoritmo que genera los sucesivos dígitos, el conocimiento de dicho algoritmo nos proporciona una forma de considerar toda la expansión decimal infinita. Los números reales cuya expansión puede ser generada mediante algoritmos se llaman números *computables* (ver también pp. 80-81). (El uso de una expansión decimal en lugar de, pongamos por caso, una binaria no tiene importancia aquí. Los números que son «computables» en este sentido son los mismos cualquiera que sea la base utilizada para la expansión.) Los números reales π y $\sqrt{2}$ que hemos estado considerando son ejemplos de números computables. En ambos casos sería un poco complicado establecer la regla en detalle, pero no hay dificultad de principio.

Sin embargo, existen también muchos números reales que *no* son computables en este sentido. Hemos visto en el último capítulo que existen secuencias no-computables que están, a pesar de todo, perfectamente bien definidas. Por ejemplo, podríamos tomar la expan-

sión decimal cuyo n -ésimo dígito es 1 ó 0 según se detenga o no la n -ésima máquina de Turing actuando sobre el número n . En general sólo pedimos que para un número real haya *cierta* expansión decimal infinita. No pedimos que exista un algoritmo para generar el n -ésimo dígito, ni siquiera que conozcamos algún tipo de regla que, en principio, defina cuál es realmente el n -ésimo dígito ². El trabajo con los números computables es cosa muy difícil. No se puede hacer que todas las operaciones sean computables aun cuando se trabaje sólo con números computables. Por ejemplo, ¿ni siquiera es un asunto computable el decidir, en general, si dos números computables son iguales o no! Por razones de este tipo, es preferible trabajar, en su lugar, con *todos* los números reales, para los que las expansiones decimales pueden ser cualesquiera y no necesitan ser, pongamos por caso, una secuencia computable.

Señalaré finalmente que existe una identificación entre un número real cuya expansión decimal termina con una sucesión infinita de 9s y otro cuya expansión termina con una sucesión infinita de 0s; por ejemplo

$$-27.1860999999... = -27.1861000000...$$

¿Cuántos números reales hay?

Hagamos una pequeña pausa para apreciar el alcance de la generalización que se ha conseguido al pasar de los números racionales a los números reales.

A primera vista se podría pensar que el número de enteros es mayor que el número de números naturales: todo número natural es un entero mientras que algunos enteros (a saber, los negativos) no son números naturales; y, análogamente, se podría pensar que el número de fracciones es mayor que el de enteros. Sin embargo, esto no es así. Según la potente y bella teoría de los números infinitos o transfinitos propuesta a finales del siglo XIX por el muy original matemático ruso-germano Georg Cantor, el número total de fracciones, el número total de enteros y el número total de números naturales son todos ellos el *mismo* número transfinito, denotado por $\aleph_0$ («aleph sub-cero»). (Curiosamente una idea de este tipo había sido parcialmente anticipada unos 260 años antes, a comienzos del siglo XVII, por el gran físico y astrónomo italiano Galileo Galilei. Recordaremos algunos otros logros de Galileo en el Capítulo 5.) Se puede ver que el

número de enteros es el mismo que el de naturales si establecemos una «correspondencia uno-a-uno» de la siguiente forma:

Enteros		Números naturales
0	$\leftrightarrow$	0
-1	$\leftrightarrow$	1
1	$\leftrightarrow$	2
-2	$\leftrightarrow$	3
2	$\leftrightarrow$	4
-3	$\leftrightarrow$	5
3	$\leftrightarrow$	6
-4	$\leftrightarrow$	7
.	.	.
.	.	.
.	.	.
$-n$	$\leftrightarrow$	$2n - 1$
n	$\leftrightarrow$	$2n$
.	.	.
.	.	.
.	.	.

Nótese que cada entero (en la columna izquierda) y cada número natural (en la columna derecha) aparecen una y sólo una vez en la lista. La existencia de una correspondencia uno-a-uno como ésta es lo que establece, en la teoría de Cantor, que el número de objetos en la columna izquierda es el *mismo* que el número de objetos en la columna derecha. Por consiguiente, el número de enteros es el mismo que el de números naturales. En este caso el número es infinito, pero no importa. (La única peculiaridad que ocurre con los números infinitos es que podemos olvidar algunos de los miembros de una lista ¡y todavía encontramos una correspondencia uno-a-uno entre las dos listas!) De manera análoga, aunque algo más complicada, podemos establecer una correspondencia uno-a-uno entre las fracciones y los enteros. (Para ello podemos adoptar una de las formas de representar *pares* de números naturales, numerador y denominador, como números naturales simples; ver Capítulo 2, p. 72.) Los conjuntos que pueden ser puestos en correspondencia uno-a-uno con los números naturales se llaman *numerables*, de modo que los conjuntos infinitos numerables tienen $\aleph_0$ elementos. Hemos visto así que los enteros son numerables y también lo son todas las fracciones.

¿Existen conjuntos que sean *no* numerables? Aunque hemos ampliado el sistema, pasando primero de los números naturales a los

enteros, y luego a los números racionales, no hemos incrementado realmente el número de objetos con los que trabajamos. Hemos visto que el número de objetos es numerable en cada caso. Quizá el lector sacó de esto la impresión de que *todos* los conjuntos infinitos son numerables. No es así; la situación es muy diferente al pasar a los números reales. Uno de los logros notables de Cantor fue demostrar que realmente hay *más* números reales que racionales. El argumento que utilizó Cantor fue el del «corte diagonal» a que nos referimos en el Capítulo 2 y que Turing adaptó en su argumento para demostrar que el problema de la detención de las máquinas de Turing es insoluble. El argumento de Cantor, como el posterior de Turing, procede por *reductio ad absurdum*. Supongamos que el resultado que estamos tratando de establecer es falso, es decir, que el conjunto de todos los números reales es numerable. Entonces, los números reales comprendidos entre 0 y 1 son ciertamente numerables, y tendremos *alguna* lista que proporciona una correspondencia uno-a-uno que empareja todos estos números con los números naturales, tal como:

Números naturales		Números reales
0	$\leftrightarrow$	0. 1 0357627183. . .
1	$\leftrightarrow$	0.1 4 329806115. . .
2	$\leftrightarrow$	0.02 1 66095213. . .
3	$\leftrightarrow$	0.430 0 5357779. . .
4	$\leftrightarrow$	0.9255 0 489101. . .
5	$\leftrightarrow$	0.59210 3 43297. . .
6	$\leftrightarrow$	0.636679 1 0457. . .
7	$\leftrightarrow$	0.8705007 4 193. . .
8	$\leftrightarrow$	0.04311737 8 04. . .
9	$\leftrightarrow$	0.78635081 1 50. . .
10	$\leftrightarrow$	0.409167388 9 1. . .
.	.	.
.	.	.
.	.	.

He marcado en **negrita** los dígitos de la diagonal. Para este listado particular, estos dígitos son

1, 4, 1, 0, 0, 3, 1, 4, 8, 5, 1,..

y el procedimiento del corte diagonal consiste en construir un número real (entre 0 y 1) cuya expansión decimal (tras el punto decimal) difiere de estos dígitos en cada uno de los lugares correspondientes.

Para poner un ejemplo concreto, digamos que el dígito será 1 dondequiera que el dígito de la diagonal es diferente de 1, y será 2 dondequiera que el dígito de la diagonal es 1. Por lo tanto, en este caso tenemos el número real

0.21211121112...

Este número real no puede figurar en nuestro listado puesto que difiere del primer número en la primera cifra decimal, del segundo número en la segunda cifra, del tercer número en la tercera cifra, etc. Existe una contradicción ya que se suponía que nuestra lista contiene *todos* los números reales entre 0 y 1. Esta contradicción establece lo que estamos tratando de probar, a saber: que *no* existe correspondencia uno-a-uno entre los números reales y los números naturales y, en consecuencia, que el número de números reales es realmente *mayor* que el número de números racionales y *no* es numerable.

El número de números reales es el número transfinito llamado C . (C significa en este caso el *continuo*, otro nombre para el sistema de los números reales.) Podríamos preguntar por qué este número no es llamado $\aleph_1$, por ejemplo. En realidad, el símbolo $\aleph_1$ representa el siguiente número transfinito mayor que $\aleph_0$, y el decidir si efectivamente $C = \aleph_1$ constituye un famoso problema no resuelto, la llamada *hipótesis del continuo*.

Puede señalarse que, por el contrario, los números *computables* son numerables. Para numerarlos basta con hacer una lista, en orden numérico, de las máquinas de Turing que generan números reales (es decir, que producen los sucesivos dígitos de los números reales). Nos gustaría tachar de la lista cualquier máquina de Turing que genere un número real que ya haya aparecido antes en la lista. Puesto que las máquinas de Turing son numerables debe darse ciertamente el caso de que los números reales computables son numerables. ¿Por qué no podemos utilizar el corte diagonal en esta lista y producir un nuevo número computable que *no* esté en la lista? La respuesta está en el hecho de que, en general, no podemos decidir computablemente si una máquina de Turing estará o no realmente en la lista. Eso supondría, en efecto, que fuéramos capaces de resolver el problema de la detención. Algunas máquinas de Turing pueden empezar a producir los dígitos de un número real, y luego quedarse en suspenso y no producir ningún otro dígito (debido a que «no se paran»). No hay medio computable de decidir qué máquinas de Turing se bloquearán de esta forma. Éste es básicamente el problema de la detención. Por tanto, aunque nuestro procedimiento del corte diagonal producirá al-

gún número real, ese número no será un número computable. En realidad, podríamos haber utilizado este argumento para *demostrar* la existencia de números no computables. El argumento de Turing para demostrar la existencia de clases de problemas que no pueden resolverse algorítmicamente, como fue expuesto en el último capítulo, sigue precisamente esta línea de razonamiento. Veremos más tarde otras aplicaciones del corte diagonal.

«Realidad» de los números reales

Dejando aparte la cuestión de la computabilidad, los números reales se llaman «reales» debido a que parecen proporcionar las magnitudes necesarias para la medida de distancias, ángulos, tiempo, energía, temperatura u otras numerosas cantidades geométricas y físicas. Sin embargo, la relación entre los números «reales» estrictamente definidos y las cantidades físicas no es tan nítida como uno pudiera imaginar. Los números reales se refieren a una *idealización matemática* más que a cualquier cantidad física real objetiva. El sistema de los números reales tiene, por ejemplo, la propiedad de que entre dos cualesquiera de ellos, por muy próximos que estén, hay un tercero. No es en absoluto evidente que las distancias o los tiempos físicos posean esa propiedad. Si continuamos dividiendo la distancia física entre dos puntos alcanzaremos finalmente escalas tan pequeñas que el mismo concepto de distancia, en sentido ordinario, deja de tener significado. Se espera que éste sería el caso en la escala de la «gravitación cuántica» de una 10^{20} -ava parte del tamaño * de una partícula subatómica. Pero para reflejar los números reales tendríamos que ir a escalas infinitamente más pequeñas que ésta: 10^{200} -avo, 10^{2000} -avo, o $10^{10^{200}}$ -avo del tamaño de una partícula, por ejemplo. No es en absoluto evidente que tales escalas absurdamente ínfimas tengan ningún significado físico. Un comentario similar sería también válido para los correspondientemente minúsculos intervalos de tiempo.

El sistema de los números reales se escoge en física por su utilidad, simplicidad y elegancia *matemática*, junto con el hecho de que concuerda, en un rango muy amplio, con los conceptos físicos de distancia y tiempo. No se ha escogido porque sepamos que está de acuerdo con estos conceptos físicos en *todos* los rangos. Se podría esperar que

* Recuérdese que la notación « 10^{20} » representa el número 10000000000000000000, en donde el 1 está seguido de veinte 0s.

no exista tal acuerdo a escalas muy pequeñas de distancia o tiempo. Es práctica común utilizar reglas para la medida de distancias simples, pero esas mismas reglas tendrán una naturaleza granular cuando descendamos a la escala de sus propios átomos. Esto, en sí mismo, no nos impide seguir utilizando los números reales de una forma aproximada, pero se necesita una sofisticación mucho mayor para la medida de distancias aún más pequeñas. Deberíamos tener al menos alguna sospecha de que pudiera haber en último término alguna dificultad de tipo fundamental para distancias en la escala más ínfima. Resulta que la Naturaleza se muestra muy amable con nosotros y parece que los mismos números reales que nos hemos acostumbrado a utilizar para la descripción de las cosas a una escala cotidiana o mayor conservan su utilidad a escalas mucho más pequeñas que los átomos —con certeza por debajo de una centésima del diámetro «clásico» de una partícula subatómica, por ejemplo un electrón o un protón— y aparentemente por debajo de la «escala de la gravitación cuántica», ¡veinte órdenes de magnitud más pequeña que el tamaño de dicha partícula! Ésta es una extraordinaria extrapolación a partir de la experiencia. El concepto familiar de distancia como número real parece ser válido también para el cuásar más remoto y aún más allá, lo que equivale a un rango global de al menos 10^{42} , y quizá 10^{60} o más. De hecho, esta adecuación del sistema de los números reales no se cuestiona normalmente. ¿Por qué se confía tanto en estos números para la exacta descripción de la física, cuando nuestra experiencia inicial de la importancia de tales números se reduce a un rango relativamente limitado? Esta confianza —quizá inmerecida— debe descansar (aunque este hecho no se reconoce a menudo) en la elegancia lógica, consistencia y potencia matemática del sistema de los números reales junto con una creencia en la profunda armonía matemática de la Naturaleza.

Números complejos

El sistema de los números reales no tiene el monopolio de la potencia y elegancia matemática. Existe todavía una dificultad, por ejemplo, en el hecho de que sólo se pueden tomar raíces cuadradas de los números positivos (o cero) y no de los negativos. Desde el punto de vista matemático —y dejando de lado, por el momento, cualquier cuestión acerca de la conexión directa con el mundo físico— resulta extraordinariamente conveniente poder extraer raíces cuadra-

das de números negativos tanto como de números positivos. Postulemos simplemente, o «inventemos», una raíz cuadrada para el número -1 . La denotaremos por el símbolo « i », de modo que tenemos

$$i^2 = -1$$

La cantidad i no puede ser, por supuesto, un número real puesto que el producto de un número real por sí mismo es siempre positivo (o cero, si el propio número es el cero). Por esta razón se ha aplicado convencionalmente el término imaginario a los números cuyos cuadrados son negativos. Sin embargo, es importante resaltar el hecho de que estos números «imaginarios» no son menos reales que los números «reales» a los que estamos acostumbrados. Como he señalado antes, la relación entre tales números «reales» y la realidad *física* no es tan directa o irresistible como pudiera parecer a primera vista, al implicar, como lo hace, una idealización matemática de resolución infinita para la que no hay *a priori* una clara justificación en la Naturaleza.

Teniendo una raíz cuadrada para -1 no supone ahora gran esfuerzo proporcionar raíces cuadradas para *todos* los números reales. En efecto, si a es un número real positivo, entonces la cantidad

$$i \times \sqrt{a}$$

es una raíz cuadrada del número real negativo $-a$. (Hay también otra raíz cuadrada, a saber, $-i \times \sqrt{a}$.) ¿Qué sucede con el propio i ? ¿Tiene una raíz cuadrada? Seguro que la tiene. Se comprueba fácilmente que la cantidad

$$(1 + i)/\sqrt{2}$$

(y también el negativo de esa cantidad) elevada al cuadrado da i . ¿Tiene *este* número una raíz cuadrada? De nuevo la respuesta es sí; el cuadrado de

$$\sqrt{\frac{(1 + 1/\sqrt{2})}{2}} + i \sqrt{\frac{(1 - 1/\sqrt{2})}{2}}$$

o de su negativo es en efecto $(1 + i)/\sqrt{2}$.

Nótese que al formar tales cantidades nos hemos permitido sumar números reales con números imaginarios, tanto como multiplicar nuestros números por números reales arbitrarios (o dividirlos por números reales diferentes de cero, que es lo mismo que multiplicar por sus

recíprocos). Los objetos resultantes son los llamados *números complejos*. Un número complejo es un número de la forma

$$a + ib$$

donde a y b son números reales llamados *parte real* y *parte imaginaria*, respectivamente, del número complejo. Las reglas para sumar y multiplicar tales números se siguen de las reglas ordinarias del álgebra, con la regla añadida de que $i^2 = -1$:

$$\begin{aligned}(a + ib) + (c + id) &= (a + c) + i(b + d) \\ (a + ib) \times (c + id) &= (ac - bd) + i(ad + bc)\end{aligned}$$

¡Ahora sucede algo notable! Nuestro motivo para introducir este sistema de números había sido proporcionar la posibilidad de que siempre se pudieran tomar raíces cuadradas. El sistema consigue este cometido, aunque esto no es en sí mismo todavía obvio. Pero consigue mucho más: pueden obtenerse sin problemas raíces cúbicas, raíces quintas, raíces novenas, raíces de orden π , raíces de orden $(1 + i)$, etc. (como fue capaz de demostrar el gran matemático del siglo XVIII Leonhard Euler). Como un ejemplo más de la magia de los números complejos examinemos las fórmulas trigonométricas de apariencia algo complicada que aprendimos en la escuela: los senos y cosenos de la suma de dos ángulos

$$\begin{aligned}\text{sen}(A + B) &= \text{sen } A \cos B + \cos A \text{sen } B \\ \cos(A + B) &= \cos A \cos B - \text{sen } A \text{sen } B\end{aligned}$$

son simplemente las partes imaginaria y real respectivamente de la ecuación compleja mucho más simple (¡y mucho más fácil de recordar!) *

$$e^{iA + iB} = e^{iA} e^{iB}.$$

Todo lo que necesitamos saber aquí es la «fórmula de Euler» (aparentemente también obtenida muchos años antes que Euler por el famoso matemático inglés del siglo XVI Roger Cotes)

$$e^{iA} = \cos A + i \text{sen } A,$$

* La cantidad $e = 2.7182818285...$ (la base de los logaritmos naturales, y un número irracional de una importancia en matemáticas comparable a la de π) viene definida por:

$$e = 1 + 1/1 + 1/(1 \times 2) + 1/(1 \times 2 \times 3) + \dots$$

y e^z significa la z -ésima potencia de e , donde tenemos, como resultado

$$e^z = 1 + z/1 + z^2/(1 \times 2) + z^3/(1 \times 2 \times 3) + \dots$$

que sustituimos en la expresión anterior. La expresión resultante es $\cos(A + B) + i \sin(A + B) = (\cos A + i \sin A)(\cos B + i \sin B)$

y llevando a cabo la multiplicación en el segundo miembro obtenemos las relaciones deseadas.

Más aún, cualquier ecuación algebraica

$$a_0 + a_1z + a_2z^2 + a_3z^3 + \dots + a_nz^n = 0$$

(en la que $a_0, a_1, \dots, a_n$ son números complejos, con $a_n \neq 0$) puede satisfacerse siempre para un cierto z . Por ejemplo, existe un número complejo z que satisface la ecuación

$$z^{102} + 999z^{33} - \pi z^2 = -417 + i,$$

¡aunque esto no es en modo alguno obvio! El hecho general es denominado a veces «teorema fundamental del álgebra». Varios matemáticos del siglo XVIII lucharon por demostrar este resultado. Ni siquiera Euler encontró un argumento general satisfactorio. Más tarde, en 1831, el gran matemático y científico Carl Friedrich Gauss dio una línea de argumentación sorprendentemente original y obtuvo la primera demostración general. Un ingrediente clave de esta demostración consistía en representar *geométricamente* los números complejos y, una vez hecho esto, utilizar un argumento topológico*.

En realidad, Gauss no fue el primero en utilizar una descripción geométrica de los números complejos. Wallis lo había hecho, de forma más tosca, unos doscientos años antes, aunque no la había utilizado con un efecto tan poderoso como lo hizo Gauss. El nombre que va asociado normalmente a esta representación geométrica de los números complejos es el de Jean Robert Argand, un contable suizo, que la describió en 1806, aunque el topógrafo noruego Caspar Wessel había dado ya una descripción muy completa nueve años antes. De acuerdo con esta terminología convencional (aunque no totalmente exacta históricamente) me referiré a la representación geométrica estándar de los números complejos como el *plano de Argand*.

El plano de Argand es un plano euclídeo ordinario con coordenadas cartesianas estándar x e y , donde x indica la distancia horizontal (positiva hacia la derecha y negativa hacia la izquierda) y donde y

* La palabra «topológico» se refiere al tipo de geometría —denominada a veces «geometría de la lámina elástica»— en la que no se consideran las distancias reales, y sólo tienen relevancia las propiedades de continuidad de los objetos.

indica la distancia vertical (positiva hacia arriba y negativa hacia abajo). El número complejo

$$z = x + iy$$

viene representado entonces por el punto del plano de Argand cuyas coordenadas son

$$(x, y)$$

(ver Fig. 3.8).

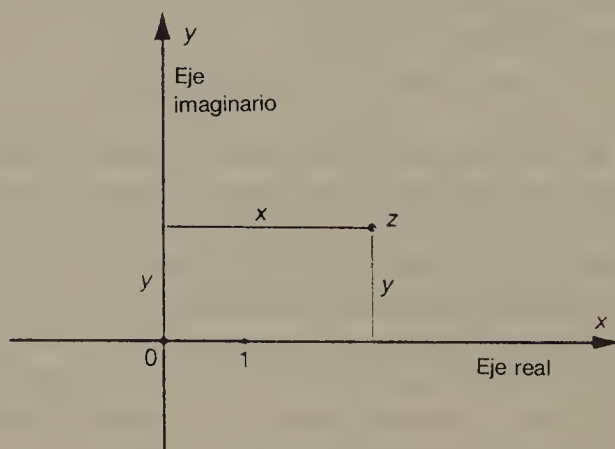


Fig. 3.8. Un número complejo $z = x + iy$ representado en el plano de Argand.

Nótese que 0 (considerado como un número complejo) viene representado por el origen de coordenadas, y 1 viene representado por un punto en el eje x .

El plano de Argand proporciona simplemente un modo de organizar nuestra familia de números complejos en una imagen geométricamente útil. Cosas de este tipo no son realmente nuevas para nosotros. Ya estamos familiarizados con el modo en que se pueden organizar los números *reales* en una imagen geométrica: la imagen de una línea recta que se extiende indefinidamente en ambas direcciones. Un punto particular de la línea se etiqueta como 0 y otro punto se etiqueta como 1. El punto 2 está colocado de modo que su desplazamiento respecto a 1 es el mismo que el desplazamiento de 1 respecto a 0; el punto $1/2$ es el punto medio entre 0 y 1; el punto -1 está

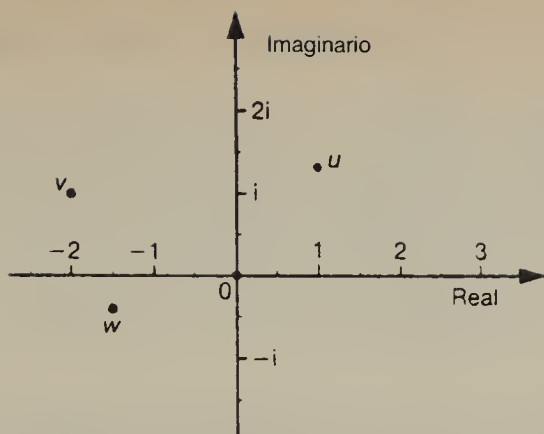


Fig. 3.9. Localizaciones en el plano de Argand de $u = 1 + i 1.3$, $v = -2 + i$, y $w = -1.5 - i 0.4$.

situado de modo que 0 esté en el punto medio entre -1 y 1 , etc. El conjunto de los números reales representado de esta forma se conoce como la *recta real*. Para los números complejos tenemos que utilizar dos números reales como coordenadas, a saber a y b , para el número complejo $a+ib$. Estos dos números nos dan las coordenadas de un punto en un plano: el plano de Argand. Como ejemplo, he indicado en la Figura 3.9 dónde estarían situados aproximadamente los números complejos

$$u = 1 + i1.3, \quad v = -2 + i, \quad w = -1.5 - i 0.4$$

Las operaciones algebraicas básicas de la suma y multiplicación de números complejos encuentran ahora una forma geométrica clara. Consideremos primero la suma. Supongamos que u y v son dos números complejos representados en el plano de Argand de acuerdo con el esquema anterior. Entonces su suma $u+v$ viene representada como la «suma vectorial» de los dos puntos; es decir, el punto $u+v$ está en el lugar que completa el paralelogramo formado por u , v y el origen 0. No es muy difícil ver que esta construcción da efectivamente la suma (ver Fig. 3.10), pero omito aquí el argumento.

El producto uv tiene también una interpretación geométrica clara (ver Fig. 3.11), que es quizá un poco más difícil de ver. (De nuevo omito el argumento.) El ángulo subtendido en el origen entre 1 y uv es la suma de los ángulos entre 1 y u y entre 1 y v (todos los ángulos se miden en sentido contrario a las agujas del reloj) y la distancia

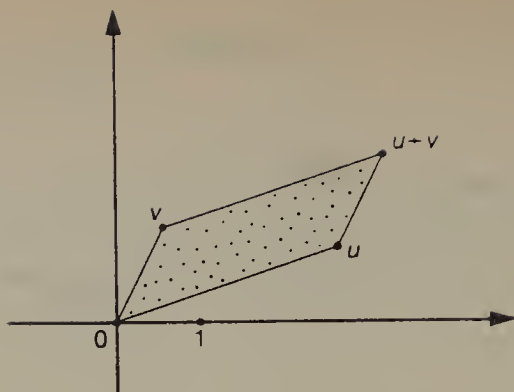


Fig. 3.10. La suma $u + v$ de los dos números complejos u y v se obtiene mediante la ley del paralelogramo.

desde el origen a uv es el producto de las distancias desde el origen a u y a v . Esto equivale a decir que el triángulo formado por 0 , v y uv es semejante (y orientado de la misma forma) al triángulo formado por 0 , 1 y u .

(El lector animoso que no esté familiarizado con estas construcciones puede ocuparse en comprobar que se siguen directamente de las reglas algebraicas para la suma y la multiplicación de números complejos que se dieron antes, junto con las identidades trigonométricas arriba expresadas).

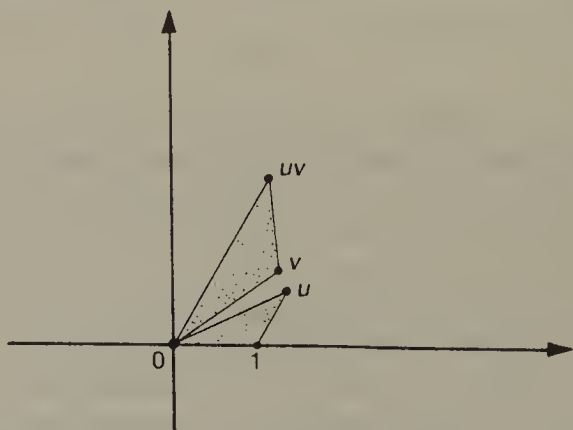


Fig. 3.11. El producto uv de dos números complejos u y v es tal que el triángulo formado por 0 , v y uv es semejante al formado por 0 , 1 y u . O lo que es equivalente: la distancia desde 0 a uv es el producto de las distancias desde 0 a u y a v , y el ángulo que forma uv con el eje real (horizontal) es la suma de los ángulos que forman u y v con dicho eje.

Construcción del conjunto de Mandelbrot

Estamos ahora en disposición de ver cómo se define el conjunto de Mandelbrot. Sea z algún número complejo escogido arbitrariamente. Cualquiera que sea este número complejo estará representado por algún punto en el plano de Argand. Consideremos ahora la *aplicación* según la cual z es reemplazado por un *nuevo* número complejo dado por

$$z \rightarrow z^2 + c$$

donde c es otro número complejo *fijo* (es decir, dado). El número $z^2 + c$ estará representado por algún nuevo punto del plano de Argand. Por ejemplo, si se da c como $1.63 - i 4.2$, entonces z se aplicará según

$$z \rightarrow z^2 + 1.63 - i 4.2$$

de modo que, en particular, 3 será reemplazado por

$$3^2 + 1.63 - i 4.2 = 9 + 1.63 - i 4.2 = 10.63 - i 4.2$$

y $-2.7 + i 0.3$ será reemplazado por

$$\begin{aligned} & (-2.7 + i 0.3)^2 + 1.63 - i 4.2 = \\ & = (-2.7)^2 - (0.3)^2 + 1.63 + i \{ 2(-2.7)(0.3) - 4.2 \} = \\ & = 8.83 - i 5.82 \end{aligned}$$

Cuando tales números se complican es mejor realizar los cálculos con un computador electrónico.

Ahora, cualquiera que sea c , el número particular 0 se reemplaza, según este esquema, por el número dado c . ¿Qué ocurre con el propio c ? Este debe ser reemplazado por el número $c^2 + c$. Supongamos que continuamos este proceso y aplicamos el reemplazamiento al número $c^2 + c$; entonces obtenemos

$$(c^2 + c)^2 + c = c^4 + 2c^3 + c^2 + c.$$

Iteremos de nuevo el reemplazamiento, aplicándolo ahora al número anterior para obtener

$$\begin{aligned} (c^3 + 2c^3 + c^2 + c)^2 + c &= c^8 + 4c^7 + 6c^6 + 6c^5 + 5c^4 \\ &+ 2c^3 + c^2 + c \end{aligned}$$

y luego otra vez a este número y así sucesivamente. Obtenemos una secuencia de números complejos, empezando con 0:

$$0, c, c^2 + c, c^4 + 2c^3 + c^2 + c, \dots$$

Ahora bien, si hacemos esto con *ciertas* elecciones del número complejo c , la secuencia de números que obtenemos así nunca se aleja mucho del origen del plano de Argand; dicho con mayor exactitud, la secuencia permanece *acotada* para tales elecciones de c , lo que equivale a decir que todo número de la secuencia está dentro de algún *círculo fijo* centrado en el origen (ver Fig. 3.12). Un buen ejemplo en donde ocurre esto es el caso $c=0$, ya que en este caso todos los miembros de la secuencia resultan ser 0. Otro ejemplo de comportamiento acotado ocurre para $c=-1$, pues entonces la secuencia es: 0, -1, 0, -1, 0, -1,...; y un ejemplo más ocurre para $c=i$, siendo entonces la secuencia 0, i , $i-1$, $-i$, $i-1$, $-i$, $i-1$, $-i$,.... Sin embargo, para otros números complejos c la secuencia se aleja más y más hasta una distancia infinita del origen; es decir, la secuencia está *no acotada*, y no puede estar contenida dentro de ningún círculo fijo. Un ejemplo de este último comportamiento ocurre para $c=1$, pues entonces la secuencia es 0, 1, 2, 5, 26, 677, 458330,...; también sucede esto

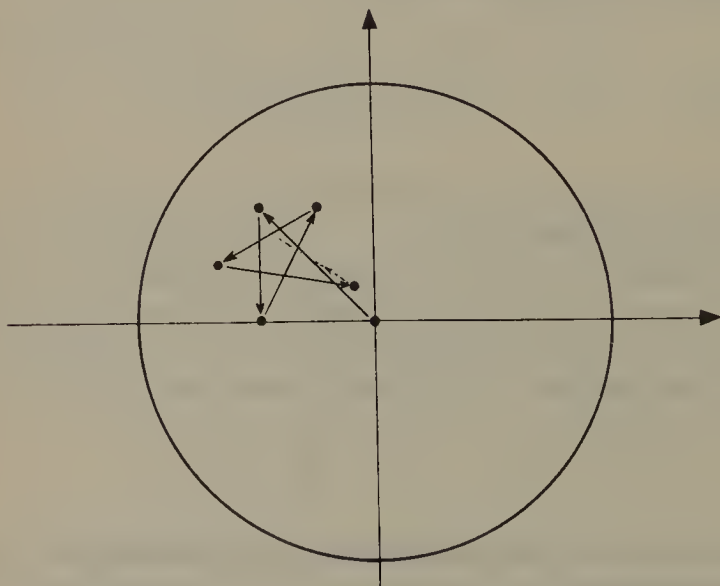


Fig. 3.12. Una secuencia de puntos en el plano de Argand está *acotada* si existe algún círculo fijo que contiene todos los puntos. (Esta iteración particular comienza con cero y tiene $c = -\frac{1}{2} + \frac{1}{2}i$.)

para $c = -3$, siendo este caso la secuencia $0, -3, 6, 33, 1086, \dots$; y también para $c = i-1$, siendo la secuencia $0, i-1, -i-1, -1+3i, -9-i5, 55+i91, -5257+i10011, \dots$

El *conjunto de Maldelbrot*, es decir, la región *en negro* de nuestro mundo de Tor'Bled-Nam, es precisamente la región del plano de Argand que consta de los puntos c para los que la secuencia permanece acotada. La región *blanca* consta de los puntos c para los que la secuencia es no acotada. Las imágenes detalladas que vimos antes estaban dibujadas a partir de datos de salida de computadores. El computador recorrerá sistemáticamente las posibles elecciones del número complejo c ; para cada elección de c calculará la secuencia $0, c, c^2+c, \dots$ y decidirá, según algún criterio apropiado, si la secuencia permanece acotada o no. Si *está* acotada, el computador dispone que un punto negro aparezca en la pantalla en el punto correspondiente a c . Si no está acotada, el computador dispone que aparezca un punto blanco. En definitiva, para cada punto de la pantalla en el rango considerado el ordenador decidirá si el punto debe estar coloreado blanco o negro.

La complejidad del conjunto de Mandelbrot es muy notable, sobre todo a la vista de que la definición de este conjunto es, como definición matemática, sorprendentemente simple. Sucede también que la estructura general de este conjunto no es muy sensible a la forma algebraica exacta de la aplicación $z \rightarrow z^2+c$ que hemos escogido. Muchas otras aplicaciones iteradas complejas (v.g. $z \rightarrow z^3+iz^2+c$) darán estructuras extraordinariamente similares (siempre que escojamos un número apropiado para empezar —quizá no 0, sino un número cuyo valor esté caracterizado por una regla matemática clara para cada elección apropiada de la aplicación). Existe, en efecto, una especie de carácter absoluto o universal para estas estructuras «de Mandelbrot», con respecto a aplicaciones iteradas complejas. El estudio de estas estructuras constituye por sí mismo una rama de las matemáticas conocida como *sistemas dinámicos complejos*.

¿Realidad platónica de los conceptos matemáticos?

¿Hasta qué punto son «reales» los objetos del mundo del matemático? Desde un cierto punto de vista parece que no puede haber nada real en ellos. Los objetos matemáticos son sólo conceptos; son idealizaciones mentales que hacen los matemáticos, a menudo estimulados por el orden aparente de ciertos aspectos del mundo que nos

rodea, pero idealizaciones mentales en cualquier caso. ¿Pueden ser algo más que meras construcciones arbitrarias de la mente humana? Al mismo tiempo parece que existe alguna realidad profunda en estos conceptos matemáticos que va más allá de las elucubraciones mentales de un matemático particular. En lugar de ello, es como si el pensamiento matemático estuviese guiado por alguna verdad exterior eterna —una verdad que tiene realidad por sí misma y que sólo se nos revela parcialmente a uno cualquiera de nosotros.

El conjunto de Mandelbrot proporciona un ejemplo sorprendente. Su estructura maravillosamente elaborada no fue la invención de ninguna persona, ni el diseño de un equipo de matemáticos. El propio Benoit Mandelbrot, el matemático polaco-americano (protagonista de la teoría fractal) que primero estudió el conjunto no tenía ninguna concepción previa acerca de la fantástica elaboración inherente al mismo, aunque sabía que estaba en la pista de algo muy interesante. ¡De hecho, cuando empezaron a surgir sus primeras imágenes de ordenador, él tuvo la impresión de que las estructuras rizadas que estaba viendo eran el resultado de un mal funcionamiento del computador (Mandelbrot 1986)! Sólo más tarde llegó a convencerse de que estaban realmente en el propio conjunto. Además, los detalles completos de la compleja estructura del conjunto de Mandelbrot no pueden ser aprehendidos realmente por ninguno de nosotros, ni pueden ser completamente revelados por un computador. Parecería que esta estructura no es sólo parte de nuestras mentes sino que tiene una realidad autónoma. Cualquiera que sea el entusiasta matemático o computador que decida examinar el conjunto, encontrará aproximaciones a la *misma* estructura matemática fundamental. No hay ninguna verdadera diferencia que dependa del computador que se utilice para hacer los cálculos (siempre que el computador tenga una precisión suficiente), aparte del hecho de que las diferencias en velocidad y memoria del computador, y capacidades de representación gráfica, puedan conducir a diferencias en los detalles finos que saldrán a la luz y en la velocidad con que se produce este detalle. El computador está siendo utilizado esencialmente de la misma forma en que el físico experimental utiliza un aparato experimental para explorar la estructura del mundo físico. El conjunto de Mandelbrot no es una invención de la mente humana; fue un descubrimiento. Al igual que el Monte Everest ¡el conjunto de Mandelbrot está *ahí*!

De modo análogo, el propio sistema de los números complejos tiene una realidad profunda e intemporal que va bastante más allá de las construcciones mentales de cualquier matemático particular. Los

comienzos de una apreciación de los números complejos procedían de la obra de Girolamo Cardano. Éste fue un italiano que vivió entre 1501 y 1576, médico de formación, jugador y confeccionador de horóscopos (en cierta ocasión hizo un horóscopo de Cristo), y que escribió un importante e influyente tratado de álgebra, el «Ars Magna» en 1545. En éste desarrolló la primera expresión completa para la solución (en términos de radicales, esto es, raíces n -ésimas) de una ecuación cúbica general *. Él había notado, sin embargo, que en cierto tipo de casos —los llamados «irreducibles», en los que la ecuación tiene tres soluciones reales— se veía obligado a tomar, en cierto paso de su expresión, la *raíz cuadrada de un número negativo*. Aunque esto era un enigma para él, se dio cuenta de que si se permitía *tomar* esa raíz cuadrada, y *sólo* entonces, podía expresar la respuesta completa (siendo la respuesta final siempre real). Más tarde, en 1572, Raphael Bombelli, en una obra titulada «L'Algebra», extendió el trabajo de Cardano y comenzó el estudio del actual álgebra de los números complejos.

Aunque al principio puede parecer que la introducción de tales raíces cuadradas de números negativos es sólo un artificio —una invención matemática diseñada para conseguir un determinado propósito— se hizo claro más adelante que estos objetos consiguen mucho más que aquello para lo que fueron diseñados originalmente. Como mencioné arriba, aunque el objetivo original de la introducción de los números complejos era poder tomar raíces cuadradas sin problemas, al introducir tales números nos encontramos, como premio añadido, con la potencialidad de tomar cualquier otro tipo de raíz o resolver cualquier ecuación algebraica. Más adelante encontraremos muchas otras propiedades mágicas que poseen estos números complejos, propiedades que no sospechábamos al principio. Estas propiedades están *ahí*. No fueron puestas por Cardano, ni por Bombelli, ni Wallis, ni Cotes, ni Euler, ni Wessel, ni Gauss, a pesar de la indudable perspicacia de estos, y otros, grandes matemáticos; dicha magia era inherente a la propia estructura que estaban descubriendo gradualmente. Cuando Cardano introdujo sus números complejos no podía sospechar las muchas propiedades mágicas que se iban a seguir de ello —propiedades que aparecen bajo nombres diversos, como la fórmula integral de Cauchy, el teorema de la aplicación de Riemann, la propiedad de extensión de Lewy. Estos, y muchos otros hechos notables,

* Basado en parte en la obra previa de Scipione del Ferro y de Tartaglia.

son propiedades de los mismos números, sin ninguna modificación adicional, que Cardano encontró por primera vez alrededor de 1539.

¿Es la matemática invención o descubrimiento? Cuando los matemáticos obtienen sus resultados ¿están produciendo solamente elaboradas construcciones mentales que no tienen auténtica realidad, pero cuyo poder y elegancia basta simplemente para engañar incluso a sus inventores haciéndoles creer que estas construcciones mentales son «reales»? ¿O están descubriendo realmente verdades que estaban ya «ahí», verdades cuya existencia es independiente de las actividades de los matemáticos? Creo que, por ahora, debe quedar muy claro para el lector que me adhiero al segundo punto de vista, más que al primero, al menos con respecto a estructuras como los números complejos y el conjunto de Mandelbrot.

Pero quizá la cuestión no sea tan sencilla como esto. Como ya he dicho, hay cosas en las matemáticas, tales como los ejemplos que acabo de citar, para las que el término «descubrimiento» es mucho más apropiado que «invención». Existen los casos en que sale de la estructura mucho más de lo que se introdujo al principio. Podríamos adoptar el punto de vista de que en tales casos los matemáticos han tropezado con «obras de Dios». Sin embargo, existen otros casos en los que la estructura matemática no tiene esa compulsiva unicidad; por ejemplo, cuando en medio de la demostración de algún resultado el matemático encuentra necesario introducir alguna construcción artificial, y de ningún modo única, para conseguir algún fin muy específico. En tales casos, no es probable que se obtenga nada más de la construcción que lo que se puso al principio, y la palabra «invención» parece más apropiada que «descubrimiento». Éstas son «obras del hombre». Desde este punto de vista, los auténticos descubrimientos matemáticos serían considerados, de forma general, como consecuencias o aspiraciones más altas que lo que serían las «meras» invenciones.

Tales clasificaciones no son muy diferentes de las que podríamos utilizar en las artes o la ingeniería. Las grandes obras de arte están «más cerca de Dios» que las obras menores. Es un sentir no poco común entre artistas el que en sus obras más grandes están revelando verdades eternas que tienen algún tipo de existencia etérea previa *, mientras que sus obras menores podrían ser más arbitrarias, de la naturaleza de las meras construcciones mortales. De modo análogo, una innovación de bella sencillez en ingeniería, con la que abre una

* Como dijo el famoso escritor argentino Jorge Luis Borges: «... un buen poeta es un descubridor antes que un inventor...».

enorme perspectiva para la aplicación de alguna idea simple e inesperada, podría ser descrita con propiedad como un descubrimiento más que una invención.

No obstante, después de hacer estos comentarios no puedo evitar el sentimiento de que, en el caso de las matemáticas, la creencia en algún tipo de existencia etérea y eterna, al menos para los conceptos más profundamente matemáticos, es mucho más fuerte que en los otros casos. Hay en tales ideas matemáticas una compulsiva unicidad y universalidad que parecen ser de un orden diferente del que se pudiera esperar en las artes o la ingeniería. El punto de vista de que los conceptos matemáticos podrían existir en ese sentido etéreo e intemporal fue planteado en tiempos antiguos (c. 360 a.C.) por el gran filósofo griego Platón. En consecuencia, este punto de vista es calificado a veces de platonismo matemático. Tendrá gran importancia para nosotros más tarde.

En el Capítulo 1 discutí con cierta amplitud el punto de vista de la *IA fuerte*, que supone que los fenómenos mentales encuentran su existencia dentro de la idea matemática de un algoritmo. En el Capítulo 2 subrayé el punto de que el concepto de un algoritmo es una idea profunda e «infusa». En este capítulo he argumentado que tales ideas matemáticas «infusas» tendrían algún tipo de existencia intemporal, independientes de las nuestras terrenales. ¿No da este punto de vista algún crédito al punto de vista de la *IA fuerte*, al proporcionar la posibilidad de un tipo de existencia etérea para los fenómenos mentales? Es perfectamente razonable —y aún especularé más tarde en favor de un punto de vista no muy diferente de éste; pero si los fenómenos mentales pueden encontrar un hogar de este tipo general, no creo que pueda ser en el concepto de un algoritmo. Sería necesario algo mucho más sutil. El hecho de que las cosas algorítmicas constituyen una parte muy estrecha y limitada de las matemáticas será un aspecto importante de las discusiones siguientes. Empezaremos a ver algo del alcance y sutileza de las matemáticas no algorítmicas en el próximo capítulo.

4. Verdad, demostración e intuición directa

El programa de Hilbert para las matemáticas

¿Qué es la verdad? ¿Cómo formamos nuestros juicios acerca de lo que es verdadero y lo que es falso en el mundo? ¿Estamos siguiendo simplemente un *algoritmo* —sin duda favorecido por el poderoso proceso de selección natural sobre otros posibles algoritmos menos eficaces? ¿O podría existir algún otro camino, posiblemente no algorítmico —tal vez perspicacia, instinto o intuición directa— para descubrir la verdad? Ésta parece una pregunta difícil. Nuestros juicios se basan en complejas e interconectadas combinaciones de datos sensoriales, razonamientos y conjeturas. Además, en muchas situaciones ordinarias puede no haber un acuerdo general sobre lo que realmente *es* verdadero y lo que *es* falso. Para simplificar la pregunta, consideremos sólo la verdad *matemática*. ¿Cómo formamos nuestros juicios —quizá incluso nuestro conocimiento «seguro»— respecto a las cuestiones matemáticas? Aquí, por lo menos, las cosas deberían ser más nítidas. No debería haber ningún problema sobre lo que realmente es verdadero y lo que realmente es falso, ¿o debería haberlo? ¿Qué *es*, realmente, la verdad matemática?

La cuestión de la verdad matemática es muy vieja, pues se remonta al menos a la época de los primeros filósofos y matemáticos griegos —y, sin duda, aun más atrás. No obstante, en los últimos cien años, más o menos, se ha clarificado mucho y se han conseguido intuiciones *nuevas* y sorprendentes. Son estos nuevos desarrollos los que trataremos de comprender. Los resultados son muy fundamentales y afectan a la cuestión misma de si nuestros procesos de pensamiento pueden

ser de naturaleza completamente algorítmica. Es importante para nosotros que entendamos esto.

A finales del siglo XIX la matemática había hecho grandes progresos, debido en parte al desarrollo de métodos de demostración matemática cada vez más potentes. (David Hilbert y George Cantor, a quienes ya hemos encontrado antes, y el gran matemático francés Henri Poincaré, a quien encontraremos más adelante, estaban en la vanguardia de estos desarrollos.) En consecuencia, los matemáticos habían adquirido confianza en el uso de estos potentes métodos. Muchos de estos métodos implicaban el considerar conjuntos* con un infinito número de miembros, y las demostraciones tenían a menudo éxito precisamente porque era posible considerar tales conjuntos como «objetos» en acto —totalidades completas existentes, con una existencia más allá de la simple existencia en potencia. Muchas de estas poderosas ideas habían brotado del muy original concepto de *números infinitos o transfinitos* de Cantor, que éste había desarrollado coherentemente utilizando conjuntos infinitos. (Vimos algo de esto en el capítulo anterior.)

Sin embargo, esta confianza quedó hecha añicos cuando en 1902 el lógico y filósofo británico Bertrand Russell presentó su ahora célebre paradoja (ya anticipada por Cantor, y una descendiente directa del argumento del «corte diagonal».) Para comprender el argumento de Russell necesitamos primero alguna noción de lo que está implícito en la consideración de los conjuntos como totalidades completas. Podemos imaginar un conjunto que está caracterizado en términos de una *propiedad* concreta. Por ejemplo, el conjunto de las cosas *rojas* está caracterizado en términos de la propiedad de *rojez*: una cosa pertenece a dicho conjunto si y sólo si tiene rojez. Esto nos permite invertir el planteamiento y hablar de una propiedad en términos de un simple objeto, a saber, el conjunto de todas las cosas que tienen dicha propiedad. Con este punto de vista, la «rojez» es el conjunto de todas las cosas rojas. (Podemos concebir también que algunos otros conjuntos están «ahí», con sus elementos caracterizados por alguna propiedad no tan simple.)

* Un *conjunto* significa simplemente una colección de cosas —objetos físicos o conceptos matemáticos— que puede considerarse como un todo. En matemáticas, los elementos (es decir, miembros) de un conjunto son muy a menudo conjuntos ellos mismos, ya que pueden reunirse conjuntos para formar otros conjuntos. En consecuencia, se pueden considerar conjuntos de conjuntos, o conjuntos de conjuntos de conjuntos, etc.

Esta idea de definir los conceptos en términos de conjuntos era fundamental para el procedimiento, introducido en 1884 por el influente lógico germano Gottlob Frege, según el cual los *números* pueden definirse en términos de conjuntos. Por ejemplo, ¿qué queremos decir mediante el número 3? Sabemos qué es la propiedad de «trinidad» pero ¿qué es el propio 3? Ahora bien, «trinidad» es una propiedad de *colecciones* de objetos, es decir, es una propiedad de *conjuntos*: un conjunto tiene esta propiedad particular de «trinidad» si y sólo si el conjunto tiene exactamente tres miembros. Por ejemplo, el conjunto de los ganadores de medallas en una competición olímpica concreta tiene esta propiedad de «trinidad». También la tienen el conjunto de las ruedas en un triciclo, o el conjunto de las hojas de un trébol normal, o el conjunto de las soluciones de la ecuación $x^3 - 6x^2 + 11x - 6 = 0$. ¿Cuál es, entonces, la definición de Frege del número 3? Según Frege, 3 debe ser un conjunto *de* conjuntos: el conjunto de *todos* los conjuntos con la propiedad de «trinidad»¹. De este modo, un conjunto tiene tres miembros si y sólo si pertenece al conjunto 3 de Frege.

Esto puede parecer una definición circular pero, en realidad, no lo es. En general, podemos definir los *números* como totalidades de conjuntos equivalentes, donde «equivalentes» significa en este caso «que tienen elementos que pueden ser emparejados uno-a-uno» (es decir, en términos ordinarios esto equivaldría a «que tienen el mismo número de miembros»). El número 3 es entonces aquel conjunto particular del que uno de sus miembros es un conjunto que contiene, por ejemplo, simplemente una manzana, una naranja y una pera. Nótese que ésta es una definición de «3» bastante diferente de la de «**3**» de Church dada en la p. 104. Pueden darse también otras definiciones que son bastante más populares en estos días.

Ahora bien, ¿qué hay de la paradoja de Russell? Ésta se refiere a un conjunto R definido de la manera siguiente:

R es el conjunto de todos los conjuntos que no son miembros de sí mismos.

Por lo tanto, R es ciertamente una colección de conjuntos; y el criterio para que un conjunto *X* pertenezca a esta colección es que el mismo conjunto *X* no se encuentre entre sus *propios* miembros.

¿Es absurdo suponer que un conjunto pueda ser realmente un miembro de sí mismo? Ciertamente no. Consideremos, por ejemplo, el conjunto *I* de los conjuntos *infinitos* (conjuntos con infinitos miembros). Ciertamente hay una infinidad de conjuntos infinitos *diferentes*,

de modo que I es él mismo infinito. Por consiguiente ¡ I pertenece a sí mismo! Entonces ¿cómo es que la idea de Russell nos da una paradoja? Preguntemos: ¿es el propio conjunto R de Russell un miembro de sí mismo, o no lo es? Si *no* es un miembro de sí mismo entonces debería pertenecer a R puesto que R consta precisamente de aquellos conjuntos que no son miembros de sí mismos. Por consiguiente, R pertenece a R , después de todo: una contradicción. Por otro lado, si R es un miembro de sí mismo, entonces, puesto que «él mismo» es realmente R , pertenece al conjunto cuyos miembros se caracterizan por *no* ser miembros de sí mismos, es decir, no es miembro de sí mismo después de todo: ¡de nuevo una contradicción!*

Esta consideración no era una frivolidad. Russell estaba simplemente utilizando, de una forma bastante extrema, el mismo tipo de razonamiento matemático general de teoría de conjuntos que los matemáticos estaban empezando a emplear en sus demostraciones. Evidentemente las cosas se les habían ido de las manos y resultaba conveniente ser mucho más preciso sobre los tipos de razonamiento que deberían permitirse y los que no. Era obviamente necesario que los razonamientos permitidos estén libres de contradicción y que permitan que, a partir de enunciados que se sabe previamente que son verdaderos, sólo se deriven enunciados verdaderos. El propio Russell, junto con su colega Alfred North Whitehead, llegó a desarrollar un sistema matemático de axiomas y reglas de inferencia altamente formalizado, cuyo propósito era el que fuera posible traducir en su esquema todos los tipos de razonamientos matemáticos correctos. Las reglas estaban cuidadosamente seleccionadas para impedir los tipos paradójicos de razonamiento que conducían a la propia paradoja de Russell. El esquema concreto que presentaron Russell y Whitehead constituyó una obra monumental. Sin embargo, era algo incómodo y los tipos de razonamiento matemático que realmente incorporaba resultaban ser bastante limitados. El gran matemático David Hilbert, a quien encontramos antes en el Capítulo 2, se embarcó en la tarea de establecer un esquema mucho más manejable y comprensible. Se incluirían *todos* los tipos de razonamiento matemáticamente correctos

* Existe una divertida manera de expresar la paradoja de Russell en términos ordinarios. Imaginemos una biblioteca en la que hay dos catálogos, en uno de los cuales se incluyen precisamente todos los libros de la biblioteca que hacen referencia a sí mismos en algún lugar, y en el otro se incluyen precisamente todos los libros que no hacen mención de sí mismos. ¿En cuál de estos catálogos debe ser incluido el segundo catálogo?

para cualquier área matemática particular. Además, Hilbert pretendía que fuera posible *demostrar* que el esquema estaba libre de contradicción. Sólo entonces las matemáticas estarían situadas, de una vez por todas, sobre unos fundamentos inatacables.

Sin embargo, las esperanzas de Hilbert y sus seguidores quedaron defraudadas cuando, en 1931, el brillante lógico matemático austriaco de 25 años, Kurt Gödel, presentó un sorprendente teorema que destrozaba efectivamente el programa de Hilbert. Lo que Gödel demostró era que *cualquiera* de estos sistemas matemáticos precisos («formales») de axiomas y reglas de inferencia, siempre que sea lo bastante amplio como para contener descripciones de proposiciones aritméticas simples (como el «último teorema de Fermat» considerado en el Capítulo 2) y siempre que esté libre de contradicción, debe contener algunos enunciados que no son demostrables ni indemostrables con los medios permitidos dentro del sistema. La verdad de tales enunciados es así «indecidible» mediante los procedimientos aceptados. De hecho, Gödel pudo demostrar que el mismo enunciado de la consistencia del propio sistema axiomático debe ser una de estas proposiciones «indecidibles». Será importante que comprendamos la naturaleza de esta «indecidibilidad». Veremos por qué el argumento de Gödel rompe el núcleo mismo del programa de Hilbert. Veremos también cómo el argumento de Gödel nos da la posibilidad, mediante intuición directa, de ir más allá de las limitaciones de cualquier sistema matemático formalizado particular bajo consideración. Esta comprensión será crucial para gran parte de la discusión posterior.

Sistemas matemáticos formales

Será preciso que seamos un poco más explícitos sobre lo que entendemos por un «sistema matemático formal de axiomas y reglas de inferencia». Debemos suponer que existe algún alfabeto de símbolos en cuyos términos se expresan nuestros enunciados matemáticos. Ciertamente estos símbolos deben ser adecuados para permitir una notación para los números naturales, de modo que la «aritmética» pueda incorporarse dentro de nuestro sistema. Podemos, si así lo queremos, utilizar simplemente la notación arábiga usual 0, 1, 2, 3,..., 9, 10, 11, 12,..., para los números, aunque esto hace que la especificación de las reglas sea un poco más complicada de lo necesario. Resultará una simplificación mucho mayor si utilizamos, por ejemplo, 0, 01, 011, 0111, 01111,..., para denotar la serie de los números naturales (o, a

modo de compromiso, podríamos utilizar la notación binaria). No obstante, puesto que ello provocaría confusión en la discusión que sigue, me atenderé a la notación arábica usual en mis descripciones, cualquiera que sea la notación que el sistema pudiera usar *realmente*. Podríamos necesitar un símbolo «espacio» para separar las distintas «palabras» o «números» de nuestro sistema pero, ya que esto también puede resultar confuso, utilizaremos sólo una coma (,) para este propósito cuando sea necesario. También necesitaremos letras que denoten números naturales arbitrarios («variables») (o quizá enteros, o racionales, etc. —aunque nos limitaremos aquí a los números naturales), digamos $t, u, v, w, x, y, z, t', t'', t''', \dots$. Pueden ser necesarias letras con prima $t', t'', \dots$, ya que no queremos poner ningún límite definido al número de variables que pueden figurar en una expresión. Consideramos la *prima* (') como otro símbolo más dentro del sistema formal, de modo que el número real de *símbolos* permanece finito. Necesitaremos símbolos para las operaciones aritméticas básicas $=, +, \times$, etc., quizá para varios tipos de paréntesis $(,), [,]$, y para los símbolos *lógicos* tales como $\&(\text{«y»}), \Rightarrow (\text{«implica»}), \vee (\text{«o»}), \Leftrightarrow (\text{«si y sólo si»}), \sim (\text{«no» o «no es el caso que...»})$. Además, nos harán falta los «cuantificadores» lógicos: el *cuantificador existencial* $\exists$ («existe... tal que») y el *cuantificador universal* $\forall$ («para todo... tenemos»). Entonces podemos formar enunciados tales como el «último teorema de Fermat».

$$\sim \exists w, x, y, z [(x+1)^{w+3} + (y+1)^{w+3} = (z+1)^{w+3}]$$

(ver Capítulo 2, p. 91). (Podría haber escrito «0111» para «3», y utilizado quizá una notación para «elevar a una potencia» que se ajustara mejor al formalismo; pero, como ya he dicho, me estoy ateniendo a los símbolos tradicionales para no introducir confusiones innecesarias.) El enunciado anterior se lee (terminando en el primer paréntesis cuadrado):

«No es el caso que existan números naturales w, x, y, z tales que...»

Podemos también reescribir el último teorema de Fermat utilizando $\forall$:

$$\forall x, y, z, w [(x+1)^{w+3} + (y+1)^{w+3} \neq (z+1)^{w+3}],$$

que se lee (terminando después del símbolo «no» tras el primer paréntesis):

«Para todos los números naturales w, x, y, z no es el caso que...»
que es lógicamente equivalente a lo anterior.

Necesitamos letras para denotar proposiciones completas, y para ello utilizaré letras mayúsculas: P, Q, R, S,... Una proposición tal podría ser de hecho la anterior afirmación de Fermat:

$$F = \sim \exists x, y, z, w [(x+1)^{w+3} + (y+1)^{w+3} = (z+1)^{w+3}].$$

Una proposición podría también *depender* de una o más variables; por ejemplo, podríamos estar interesados en la afirmación de Fermat para alguna *potencia particular* * $w+3$:

$$G(w) = \sim \exists x, y, z [(x+1)^{w+3} + (y+1)^{w+3} = (z+1)^{w+3}].$$

de modo que, $G(0)$ afirma que «ningún cubo puede ser suma de cubos positivos», $G(1)$ afirma lo mismo para las potencias cuartas, y así sucesivamente. (Nótese que ahora falta la « w » tras « $\exists$ ».) La afirmación de Fermat es ahora que $G(w)$ es válida para *todo* w :

$$F = \forall w [G(w)]$$

$G()$ es un ejemplo de lo que se conoce como una *función proposicional* es decir, una proposición que depende de una o más variables.

Los *axiomas* del sistema constituirán una lista finita de proposiciones generales cuya verdad, dados los significados de los símbolos, se supone autoevidente. Por ejemplo, para proposiciones o funciones proposicionales arbitrarias $P, Q, R()$, tendremos entre nuestros axiomas:

$$\begin{aligned} (P \& Q) &\Rightarrow P, \\ \sim(\sim P) &\Leftrightarrow P, \\ \sim \exists x [R(x)] &\Leftrightarrow \forall x [\sim R(x)], \end{aligned}$$

cuya «verdad autoevidente» se comprueba inmediatamente a partir de sus *significados*. (El primero afirma simplemente que: «si P y Q son ambas verdaderas, entonces P es verdadera»; la segunda afirma la equivalencia entre «no es el caso que P es falsa» y « P es verdadera»; la tercera está ejemplificada por la equivalencia lógica de las dos for-

* Aunque todavía no se sabe si es verdadera la proposición general F de Fermat, se sabe que son verdaderas las proposiciones individuales $G(0), G(1), G(2), G(3), \dots$ hasta aproximadamente $G(125000)$. Es decir, se sabe que ningún cubo puede ser la suma de cubos positivos, ninguna potencia cuarta puede ser suma de potencias cuartas, etc., y así sucesivamente hasta el enunciado para potencias 125000-ésimas.

mas de enunciar el último teorema de Fermat dadas más arriba.) Podríamos incluir también axiomas aritméticos básicos, tales como

$$\begin{aligned} &\forall x, y [x + y = y + x] \\ &\forall x, y, z [(x + y) \times z = (x \times z) + (y \times z)], \end{aligned}$$

aunque podría ser preferible construir estas operaciones matemáticas a partir de algo más primitivo y deducir estos enunciados como teoremas. Las *reglas de inferencia* serán cosas (autoevidentes) como:

«a partir de P y $P \Rightarrow Q$ podemos deducir Q »

«a partir de $\forall x [R(x)]$ podemos deducir cualquier proposición que se obtenga sustituyendo x por un número natural concreto en $R(x)$ ».

Éstas son instrucciones que nos dicen cómo podemos derivar nuevas proposiciones a partir de proposiciones ya establecidas.

Ahora, partiendo de los axiomas y aplicando una y otra vez las reglas, podemos construir una larga lista de proposiciones. En cualquier etapa, podemos hacer entrar de nuevo en juego cualquiera de los axiomas, y podemos siempre seguir haciendo uso de las proposiciones que ya hemos añadido a nuestra creciente lista. Las proposiciones de una cualquiera de estas listas correctamente concatenadas se conocen como *teoremas* (aunque muchas de ellas sean bastante triviales o sin interés como enunciados matemáticos). Si tenemos una proposición concreta P que queremos *demostrar*, entonces trataremos de encontrar dicha lista, correctamente concatenada con arreglo a estas reglas, y que termina con nuestra concreta proposición P . Tal lista nos proporcionaría una *demostración* de P dentro del sistema; y, en consecuencia, P será un teorema.

La idea del programa de Hilbert consistía en encontrar, para cualquier área bien definida de las matemáticas, una lista de axiomas y reglas de inferencia suficientemente amplia que incorporara *todas* las formas de razonamiento matemático correcto apropiadas para dicha área. Fijemos la *aritmética* como nuestra área en cuestión de las matemáticas (donde se incluyen los cuantificadores $\exists$ y $\forall$, de modo que puedan formarse enunciados como el del último teorema de Fermat). No habría ninguna ventaja para nosotros en considerar aquí cualquier área matemática más general que ésta. La aritmética es *ya* bastante general para aplicar el método de Gödel. Según el programa de Hilbert, si se acepta que semejante amplio sistema de axiomas y reglas de inferencia nos ha sido ya dado para la aritmética, entonces dispondremos de un criterio definido para la «corrección» de la demostración matemática de cualquier proposición aritmética. Existía la espe-

ranza de que tal sistema de axiomas y reglas fuera *completo*, en el sentido de que nos permitiera en principio decidir la verdad o falsedad de *cualquier* enunciado matemático que pueda formularse dentro del sistema.

Hilbert esperaba que para cualquier cadena de símbolos que represente una proposición matemática, digamos P , deberíamos ser capaces de demostrar o bien P o bien $\sim P$, según sea P verdadera o falsa. Aquí debemos suponer que la cadena es *sintácticamente correcta* en su construcción, en donde «sintácticamente correcta» significa en esencia «gramaticalmente» correcta —es decir, que satisface todas las reglas gramaticales del formalismo, como que los paréntesis estén correctamente emparejados, etc.— de modo que P tiene un significado verdadero o falso bien definido. Si pudiera hacerse realidad la esperanza de Hilbert, ¡esto nos dispensaría incluso de preocuparnos de lo que las proposiciones *significan*! P sería simplemente una cadena de símbolos sintácticamente correcta. A la cadena de símbolos P se le asignaría el valor de verdad **verdadero** si P es un teorema (es decir, si P es demostrable dentro del sistema), y se le asignaría el valor de verdad **falso** si, por el contrario, $\sim P$ es un teorema. Para que esto tenga sentido se requiere *consistencia* además de que no falta nada. Es decir, no deben existir cadenas de símbolos P para los que tanto P como $\sim P$ sean teoremas. ¡De lo contrario, P podría ser **verdadera** y **falsa** al mismo tiempo!

El punto de vista de que se puede prescindir de los significados de los enunciados matemáticos, considerándolos como nada más que cadenas de símbolos en algún sistema matemático formal, es el punto de vista matemático del *formalismo*. A ciertas personas les gusta esta idea, según la cual las matemáticas se convierten en una especie de «juego sin significado». Sin embargo, no es una idea que me seduzca. Es, en realidad, el «significado» —y no la ciega computación algorítmica— lo que constituye la substancia de las matemáticas. ¡Afortunadamente, Gödel asestó un golpe devastador al formalismo! Veamos cómo lo hizo.

El teorema de Gödel

Una parte del argumento de Gödel era muy detallada y complicada; no es necesario, sin embargo, que examinemos las complejidades de esta parte. Por otra parte, la idea central era sencilla, bella y profunda; podremos apreciar esta idea. La parte complicada (que tam-

bién contiene mucho ingenio) consistía en mostrar en detalle cómo se pueden codificar realmente las reglas de inferencia individuales del sistema formal, y también el uso de sus diversos axiomas, en *operaciones aritméticas*. (¡El hacer comprender que esto era algo fructífero era, no obstante, un aspecto de la parte profunda!) Para llevar a cabo esta codificación necesitamos encontrar alguna forma conveniente de etiquetar las proposiciones mediante números naturales. Una forma consistiría simplemente en utilizar algún tipo de orden «alfabético» para las cadenas de símbolos del sistema formal con una longitud específica, dentro de una ordenación global según la longitud de la cadena. (Así, podrían ordenarse alfabéticamente las cadenas de longitud uno, seguidas de las cadenas de longitud dos, alfabéticamente ordenadas, seguidas de las cadenas de longitud tres, etc.) Esto se llama orden *lexicográfico* *. En realidad, Gödel utilizó originalmente un sistema de numeración más complicado, pero las diferencias no son importantes para nosotros. Nos interesaremos especialmente en las *funciones proposicionales* que dependen de una *sola* variable, como es el caso de la $G(w)$ anterior. Sea la n -ésima de estas funciones proposicionales (en el orden elegido para las cadenas de símbolos) aplicada a w

$$P_n(w).$$

Podemos permitirnos, si queremos, el que nuestra numeración sea un poco «chapucera», de forma que algunas de estas expresiones no sean sistáticamente correctas. (Esto hace la codificación aritmética mucho más fácil que si tratáramos de omitir todas las expresiones sintácticamente incorrectas.) Si $P_n(w)$ es sintácticamente correcta será algún enunciado aritmético concreto perfectamente bien definido que concierne a los dos números naturales n y w . *Cuál* sea exactamente este enunciado aritmético dependerá de los detalles del sistema de numeración específico que hayamos elegido. Eso pertenece a la parte complicada del argumento y no nos interesa aquí. Las cadenas de proposiciones que constituyen una *demonstración* de algún teorema en el

* Podemos pensar en el orden lexicográfico como la ordenación normal para los números naturales escritos en «base $k+1$ » utilizando, para los $k+1$ dígitos, los diversos símbolos del sistema formal, junto con un nuevo «cero» que no se utiliza nunca. (Esta última complicación surge debido a que los números que comienzan por cero son los mismos que aquellos en los que se omite el cero.) Un sencillo orden lexicográfico de cadenas con nueve símbolos es el dado por los números naturales que pueden escribirse en notación decimal ordinaria sin el cero: 1, 2, 3, 4,..., 8, 9, 11, 12,..., 21, 22,..., 99, 111, 112,...

sistema pueden ser etiquetadas también mediante números naturales utilizando el esquema de ordenación escogido. Denotemos mediante

$$\Pi_n$$

la n -ésima demostración. (De nuevo, podemos utilizar una «numeración chapucera» en la que para algunos valores de n la expresión « Π_n » no es sintácticamente correcta y por lo tanto no demuestra ningún teorema.)

Consideremos ahora la siguiente función proposicional, que depende del número natural w :

$$\sim \exists x [\Pi_x \text{ demuestra } P_w(w)].$$

El enunciado dentro del paréntesis cuadrado se da parcialmente en palabras, pero es un enunciado perfecta y exactamente bien definido. Afirma que la x -ésima demostración es realmente una demostración de la proposición que constituye $P_w()$ aplicada al propio valor w . El cuantificador existencial negado, fuera del paréntesis, sirve para eliminar una de las variables («no existe un x tal que...»), de modo que nos queda una función proposicional aritmética que sólo depende de una variable w . La expresión global afirma que *no* existe demostración de $P_w(w)$. Supondré que está estructurada de una manera sintácticamente correcta (incluso si $P_w(w)$ no lo está —en cuyo caso el enunciado sería *verdadero*, puesto que no puede haber demostración de una expresión sintácticamente incorrecta). De hecho, debido a las traducciones a la aritmética que se supone que hemos llevado a cabo, el enunciado de arriba es en realidad un enunciado *aritmético* relativo al número natural w (siendo la parte encerrada entre paréntesis cuadrados un enunciado aritmético bien definido relativo a *dos* números naturales x y w). No se supone que sea obvio que los enunciados puedan ser realmente codificados en la aritmética, aunque pueden serlo. Mostrar que tales enunciados pueden ser realmente codificados de esta forma es el principal «trabajo duro» incluido en la parte complicada del argumento de Gödel. Al igual que antes, *cuál* sea exactamente el enunciado aritmético dependerá en gran medida de la estructura detallada de los axiomas y reglas de nuestro sistema formal. Puesto que todo eso pertenece a la parte complicada, no nos interesaremos ahora por los detalles.

Hemos enumerado todas las funciones proposicionales que dependen de una sola variable, de modo que la que acabamos de escribir debe tener asignado un número. Escribamos este número como k .

Nuestra función proposicional es la k -ésima de la lista. Por consiguiente

$$\sim \exists x [\Pi_x \text{ demuestra } P_w(w)] = P_k(w).$$

Examinemos ahora esta función para el valor- w particular: $w = k$. Tenemos

$$\sim \exists x [\Pi_x \text{ demuestra } P_k(k)] = P_k(k).$$

La proposición específica $P_k(k)$ es un enunciado aritmético perfectamente bien definido (sintácticamente correcto). ¿Tiene una demostración dentro de nuestro sistema formal? ¿Tiene demostración su negación $\sim P_k(k)$? La respuesta a ambas preguntas debe ser «no». Podemos verlo examinando el *significado* subyacente en el procedimiento de Gödel. Aunque $P_k(k)$ es sólo una proposición aritmética, la hemos construido de modo que afirma lo que se ha escrito en el lado izquierdo: «no existe demostración, dentro del sistema, de la proposición $P_k(k)$ ». Si hemos sido cuidadosos al establecer nuestros axiomas y reglas de inferencia, y suponiendo que hayamos hecho bien nuestra numeración, entonces no puede haber ninguna demostración de esta $P_k(k)$ dentro del sistema. En efecto, si hubiera tal demostración, el significado del enunciado que $P_k(k)$ realmente afirma, a saber, que *no* existe demostración, sería falso, de modo que $P_k(k)$ tendría que ser falsa como proposición aritmética. ¡Nuestro sistema formal no debería estar tan mal construido como para permitir que se demuestran proposiciones falsas! Por consiguiente, debe ser el caso que, de hecho, *no* hay demostración de $P_k(k)$. Pero esto es precisamente lo que $P_k(k)$ está tratando de decirnos. Por lo tanto, lo que afirma $P_k(k)$ debe ser un enunciado *verdadero*, de modo que $P_k(k)$ debe ser verdadera como proposición aritmética. ¡Hemos encontrado una proposición *verdadera* que *no tiene demostración dentro del sistema*!

¿Qué sucede con su negación $\sim P_k(k)$? Se concluye que haríamos mejor en no encontrar tampoco una demostración de esta otra. Acabamos de establecer que $\sim P_k(k)$ debe ser falsa (puesto que $P_k(k)$ es verdadera), ¡y se supone que no podemos demostrar proposiciones falsas dentro del sistema! Así, ni $P_k(k)$ ni $\sim P_k(k)$ son demostrables dentro de nuestro sistema formal. Esto establece el teorema de Gödel.

La intuición matemática

Nótese que aquí ha sucedido algo muy notable. Con frecuencia la gente considera el teorema de Gödel como algo negativo, algo que muestra las necesarias limitaciones del razonamiento matemático for-

malizado. Por muy abiertos que creamos haber sido, siempre habrá algunas proposiciones que escapen a la red. Pero ¿debería preocuparnos la proposición particular $P_k(k)$? ¡En el curso del argumento anterior hemos establecido realmente que $P_k(k)$ es un enunciado *verdadero*! De algún modo nos las hemos arreglado para *ver* que $P_k(k)$ es verdadero pese al hecho de que esto no es formalmente demostrable dentro del sistema. Los formalistas matemáticos *estrictos* deberían estar realmente preocupados, ya que mediante su propio razonamiento hemos establecido que la noción formalista de «verdad» debe ser necesariamente incompleta. *Cualquiera* que sea el sistema formal (consistente) que se utilice para la aritmética, existen enunciados que podemos ver que son verdaderos pero a los que no hemos asignado el valor de verdad **verdadero** mediante el procedimiento propuesto por los formalistas tal como se describió antes. Un formalista estricto quizá trataría de evitar esto no hablando para nada del concepto de verdad sino refiriéndose simplemente a la *demostrabilidad* dentro de algún sistema formal dado. Sin embargo, esto parece muy restrictivo. Ni siquiera se puede estructurar el argumento de Gödel en la forma dada más arriba, utilizando este punto de vista, puesto que partes esenciales de dicho argumento hacen uso de razonamientos sobre lo que es verdadero y lo que no es verdadero ². Algunos formalistas adoptan un punto de vista más «pragmático» al afirmar que no les preocupan enunciados tales como $P_k(k)$ debido a que son extremadamente complejos y sin interés como proposiciones de la aritmética. Estas personas alegarán:

Sí, existe el extraño enunciado, tal como $P_k(k)$, para el que mi noción de demostrabilidad o **verdad** no coincide con su instintiva noción de verdad, pero estos enunciados no intervienen nunca en las matemáticas serias (al menos no en las que yo estoy interesado) ya que tales enunciados son absurdamente complicados y no naturales en tanto que matemáticas.

Ciertamente se da el caso de que las proposiciones como $P_k(k)$ serán, si se las escribe al completo, extremadamente incómodas y extrañas en tanto que enunciados matemáticos acerca de números. Sin embargo, en años recientes, se han expuesto algunos enunciados razonablemente sencillos, de un carácter matemático muy aceptable, que son realmente equivalentes a proposiciones de tipo-Gödel ³. Éstas son indemostrables a partir de los axiomas normales de la aritmética.

tica, pero no obstante se deducen de una propiedad «obviamente verdadera» que tiene el propio sistema axiomático.

La falta de interés que profesan los formalistas por la «verdad matemática» me parece un punto de vista muy extraño para adoptar en una filosofía de las matemáticas. Más aún: no es realmente tan pragmático. Cuando los matemáticos llevan a cabo sus razonamientos, no desean tener que estar comprobando continuamente si sus argumentos pueden ser formulados o no en términos de los axiomas y reglas de inferencia de algún sistema formal complicado. Solamente necesitan estar seguros de que sus argumentos son vías válidas para llegar a la verdad. El argumento de Gödel es otro procedimiento válido, así que me parece que $P_k(k)$ es una verdad matemática tan buena como cualquiera que pueda obtenerse de forma más convencional utilizando los axiomas y reglas de inferencia que puedan establecerse de antemano.

Un procedimiento que él mismo sugiere es el siguiente. Aceptemos que $P_k(k)$, que por el momento designaré simplemente por G_0 , es una proposición perfectamente válida; por consiguiente, podemos añadirla a nuestro sistema como un axioma adicional. Por supuesto, nuestro nuevo sistema corregido tendrá su *propia* proposición de Gödel, llamémosla G_1 , que de nuevo se ve que es un enunciado perfectamente válido acerca de números. En consecuencia, añadimos también G_1 a nuestro sistema. Esto nos da un nuevo sistema corregido que tendrá su propia proposición de Gödel G_2 (de nuevo perfectamente válida), que podemos añadir a continuación, obteniendo la siguiente proposición de Gödel G_3 , que también añadimos, y así sucesivamente, repitiendo indefinidamente el proceso. ¿Qué sucede con el sistema resultante cuando nos permitimos utilizar la lista *completa* $G_0, G_1, G_2, G_3, \dots$ como axiomas adicionales? ¿Pudiera ser que *ésta* fuera completa? Puesto que ahora tenemos un sistema de axiomas ilimitado (infinito) quizá no es evidente que sea aplicable el procedimiento de Gödel. Sin embargo, este continuo añadido de proposiciones de Gödel es un esquema perfectamente sistemático y puede ser reexpresado como un sistema lógico ordinario finito de axiomas y reglas de inferencia. Este sistema tendrá su propia proposición de Gödel, llamémosla G_ω , que puede ser añadida de nuevo y formar entonces la proposición de Gödel $G_{\omega+1}$ del sistema resultante. Repitiendo, como antes, obtenemos una lista $G_\omega, G_{\omega+1}, G_{\omega+2}, G_{\omega+3}, \dots$ de proposiciones, todas ellas enunciados perfectamente válidos acerca de números naturales, y que pueden ser todas añadidas a nuestro sistema formal. Esto es de nuevo perfectamente sistemático y nos lleva a un

nuevo sistema que engloba a todo el lote; pero éste tendrá de nuevo su proposición de Gödel, llamémosla $G_{\omega+\omega}$ que podemos reescribir como G_{ω^2} , y todo el procedimiento puede empezar otra vez de modo que tendremos una nueva lista de axiomas G_{ω^2} , G_{ω^2+1} , G_{ω^2+2} , etc., infinita pero sistemática, que conduce aún a un nuevo sistema —y una nueva proposición de Gödel G_{ω^3} . Repitiendo todo el procedimiento obtenemos G_{ω^4} y luego G_{ω^5} y así sucesivamente. Ahora bien, *este* procedimiento es completamente sistemático y tiene su propia proposición de Gödel G_{ω^2} .

¿Termina esto alguna vez? En cierto sentido, no; pero nos conduce a algunas consideraciones matemáticas difíciles en las que no podemos entrar en detalle aquí. El procedimiento anterior fue discutido por Alan Turing en un artículo de 1939 ⁴. ¡De hecho, y de forma muy notable, *cualquier* proposición verdadera (pero sólo cuantificada universalmente) de la aritmética puede ser obtenida mediante un proceso repetido de «gödelización» de este tipo! Ver Feferman (1988). Sin embargo, esto supone en cierto grado una petición de principio sobre cómo *decidimos* realmente si una proposición es verdadera o falsa. El punto crítico, en cada paso, consiste en ver cómo codificar el añadido de una familia infinita de proposiciones de Gödel que proporciona un simple axioma adicional (o un número finito de axiomas). Esto requiere que nuestra familia infinita pueda ser sistematizada de alguna forma algorítmica. Para estar seguros de que tal sistematización hace *correctamente* lo que se supone que hace, necesitamos utilizar *intuiciones* procedentes del exterior del sistema —igual que lo hicimos en primer lugar para ver que $P_k(k)$ era una proposición verdadera. Son estas intuiciones las que no pueden ser sistematizadas —y, en realidad, ¡deben estar fuera de *cualquier* acción algorítmica!

La intuición a partir de la que concluimos que la proposición de Gödel $P_k(k)$ es realmente un enunciado verdadero de la aritmética es un ejemplo de un tipo general de procedimiento que los lógicos conocen como *principio de reflexión*: así, «reflejando» el *significado* del sistema de axiomas y reglas de inferencia, y convenciéndonos de que éstos proporcionan realmente vías válidas para alcanzar las verdades matemáticas, podemos ser capaces de codificar esta intuición en nuevos enunciados matemáticos verdaderos que no eran deducibles de aquellos mismos axiomas y reglas. La derivación de la verdad de $P_k(k)$, como se resumió arriba, dependía de un principio semejante. Otro principio de reflexión, importante para el argumento original de Gödel (aunque no dado arriba), descansa en la deducción de nuevas verdades matemáticas a partir del hecho de que un sistema axiomá-

tico, que ya creemos que es válido para obtener verdades matemáticas, es realmente *consistente*. Los principios de reflexión implican con frecuencia razonamientos sobre conjuntos infinitos, y siempre hay que ser cuidadosos al utilizarlos de modo que no estén demasiado cerca del tipo de argumento que pudiera conducirnos a una paradoja tipo-Russell. Los principios de reflexión proporcionan la propia antítesis del razonamiento formalista. Si se es cuidadoso, nos capacitan para salir fuera de los rígidos confinamientos de cualquier sistema formal y obtener nuevas intuiciones matemáticas que no parecían disponibles antes. Podría haber muchos resultados perfectamente aceptables en nuestra literatura matemática cuyas demostraciones requieran intuiciones que quedan lejos de los axiomas y reglas originales de los sistemas formales estándar para la aritmética. Todo esto muestra que los procedimientos mentales mediante los que los matemáticos llegan a sus juicios de verdad no están simplemente enraizados en los procedimientos de algún sistema formal específico. *Vemos* la validez de la proposición de Gödel $P_k(k)$ aunque no podamos derivarla de los axiomas. El tipo de «visión» que está implicada en un principio de reflexión requiere un acto de intuición esencial matemática que no es el resultado de las operaciones puramente algorítmicas que pudieran ser codificadas en algún sistema matemático formal. Volveremos a tratar esta materia en el Capítulo 10.

El lector puede notar una cierta similitud entre el argumento que establece la verdad, aunque «indemostrabilidad», de $P_k(k)$ y el argumento de la paradoja de Russell. También hay una similitud con el argumento de Turing que establece la no-existencia de una máquina de Turing que resuelva el problema de la detención. Estas similitudes no son accidentales. Existe un fuerte hilo conductor histórico entre los tres. Turing encontró su argumento después de estudiar el trabajo de Gödel. El mismo Gödel conocía bien la paradoja de Russell y fue capaz de transformar el razonamiento paradójico de este tipo, que lleva demasiado lejos el uso de la lógica, en un argumento matemático válido. (Todos estos argumentos tienen su origen en el «corte diagonal» de Cantor, descrito en el capítulo anterior, p. 119).

¿Por qué deberíamos aceptar los argumentos de Gödel y Turing cuando hemos tenido que rechazar el razonamiento que conduce a la paradoja de Russell? Los primeros son mucho más nítidos y son irreprochables como argumentos matemáticos, mientras que la paradoja de Russell descansa en razonamientos más vagos que involucran conjuntos «inmensos». Pero debe admitirse que las diferencias no son realmente tan claras como a uno le gustaría que fuesen. El intento de

clarificar estas diferencias era un motivo poderoso tras la idea global del formalismo. El argumento de Gödel muestra que el punto de vista del formalismo estricto es insostenible, pero no nos lleva a un punto de vista alternativo completamente fiable. Para mí, el resultado final sigue sin estar resuelto. El método que se adopta * de hecho en las matemáticas contemporáneas para evitar el tipo de razonamiento con conjuntos «inmensos» que conduce a la paradoja de Russell no es enteramente satisfactorio. Además, aún suele establecerse en términos típicamente formalistas —o, alternativamente, en términos que no nos dan una confianza plena en que no puedan aparecer contradicciones.

Sea como fuere, me parece que una consecuencia evidente del argumento de Gödel es que el concepto de verdad matemática no puede ser encapsulado en ningún esquema formalista. Esto es quizá evidente aun sin el teorema de Gödel. En efecto, ¿cómo vamos a decidir qué axiomas o reglas de inferencia adoptar en un caso cualquiera cuando tratamos de establecer un sistema formal? Nuestra guía en la decisión de las reglas a adoptar debe ser siempre nuestra comprensión intuitiva de lo que es «autoevidentemente verdadero», dados los significados de los símbolos del sistema. ¿Cómo vamos a decidir qué sistemas formales son razonables para ser adoptados —es decir, que están de acuerdo con nuestras ideas intuitivas sobre «autoevidencia» y «significado»— y cuáles no? Ciertamente, la noción de autoconsistencia no es adecuada para ello. Podemos tener muchos sistemas autoconsistentes que no son «razonables» en este sentido, en los que los axiomas y reglas de inferencia tienen significados que rechazaríamos como falsos, o quizá no tienen significado en absoluto. «Autoevidencia» y «significado» son conceptos que seguirían siendo necesarios aun sin el teorema de Gödel.

Sin embargo, sin el teorema de Gödel hubiera sido posible imaginar que las nociones intuitivas de «autoevidencia» y «significado» podrían haber sido empleadas sólo una vez y para siempre, simplemente para establecer el sistema formal en primer lugar, y prescindir

* Se hace una distinción entre «conjuntos» y «clases», en la que se permite que los conjuntos se agrupen para formar otros conjuntos o tal vez clases, pero *no* se permite que las clases se agrupen para formar colecciones mayores de cualquier tipo, siendo consideradas como «demasiado grandes» para esto. Sin embargo, no existe ninguna regla para decidir cuándo se permite que una colección pueda ser considerada como un conjunto o cuando deba considerarse necesariamente que sólo es una clase, ¡aparte de la regla circular que establece que los conjuntos son aquellas colecciones que pueden agruparse para formar otras colecciones!

de ellas en lo sucesivo como parte de un argumento matemático claro para determinar la verdad. Entonces, según el punto de vista formalista, estas «vagas» nociones intuitivas hubieran tenido algún papel que jugar como parte del pensamiento matemático *preliminar*, como una guía hacia el descubrimiento del argumento formal apropiado; pero no desempeñarían ningún papel en la demostración real de la verdad matemática. El teorema de Gödel muestra que este punto de vista no es sostenible en una filosofía de los fundamentos de las matemáticas. La noción de verdad matemática va más allá del concepto global de formalismo. Hay algo absoluto e «infuso» en la verdad matemática. De esto trata el platonismo matemático, como se discutió al final del último capítulo. Cualquier sistema formal concreto tiene una cualidad provisional y «de factura humana». Tales sistemas pueden desempeñar papeles muy valiosos en las discusiones matemáticas, pero sólo pueden proporcionar una guía parcial (o aproximada) a la verdad. La verdad matemática real va más allá de las simples construcciones humanas.

¿Platonismo o intuicionismo?

He señalado dos escuelas contrarias de filosofía matemática, que se inclinan fuertemente hacia el platonismo antes que hacia el punto de vista formalista. En realidad he sido bastante simplista en mis distinciones y pueden hacerse muchas matizaciones. Por ejemplo, se puede argumentar bajo el encabezamiento de platonismo bien tengan los objetos del pensamiento matemático algún tipo de «existencia» real o bien si es sólo el concepto de «verdad» matemática el que es absoluto. No he querido plantear aquí estas distinciones. A mi modo de ver, el carácter absoluto de la verdad matemática y la existencia platónica de los conceptos matemáticos son esencialmente la misma cosa. La «existencia» que debe atribuirse al conjunto de Mandelbrot, por ejemplo, es una característica de su naturaleza «absoluta». El que un punto del plano de Argand pertenezca o no al conjunto de Mandelbrot es una cuestión absoluta, independiente de qué matemático, o qué computador, lo esté examinando. Es la «independencia-del-matemático» del conjunto de Mandelbrot lo que le confiere su existencia platónica. Además, sus detalles más finos quedan fuera de lo que nos es accesible mediante el uso de computadores. Estos dispositivos sólo pueden darnos aproximaciones a una estructura que tiene en sí misma una existencia más profunda e «independiente-del-computador». Reconozco, sin embargo, que puede haber muchos otros puntos de vista

que se puedan mantener razonablemente a propósito de esta cuestión. No tenemos aquí que preocuparnos mucho por estas diferencias.

Hay también diferencias de punto de vista sobre hasta qué extremo estamos dispuestos a llevar nuestro platonismo —si realmente uno afirma ser un platónico. El propio Gödel era un gran platónico. Los tipos de enunciados matemáticos que he estado considerando hasta ahora son más bien «tibios» tal como van las cosas⁵. Pueden surgir enunciados mucho más controvertidos, particularmente en la teoría de conjuntos. Cuando se consideran todas las ramificaciones de la teoría de conjuntos se tropieza con conjuntos tan desmesuradamente enormes, y contruidos de manera tan vaga, que incluso un decidido platónico como yo mismo puede honestamente empezar a dudar que su existencia, o inexistencia, sea realmente algo «absoluto»⁶. Puede llegar un momento en que los conjuntos tengan una definición tan intrincada y conceptualmente dudosa que la cuestión de la verdad o falsedad de enunciados matemáticos relativos a ellos pueda empezar a adquirir algo de la cualidad de «cuestión de opinión» en lugar de la de «infusa». El que uno esté preparado para llevar el platonismo hasta el final, junto con Gödel, y exigir que la verdad o falsedad de los enunciados matemáticos relativos a tan enormes conjuntos sea siempre algo «absoluto» o platónico, o bien se detenga en algún punto anterior y exija una verdad o falsedad absoluta sólo cuando los conjuntos son razonablemente constructivos y no tan desmesuradamente enormes, no es un asunto que tenga aquí gran relevancia para nuestra discusión. Los conjuntos (finitos o infinitos) que tendrán importancia para nosotros son ridículamente minúsculos comparados con aquellos a los que me acabo de referir. Por ello las diferencias entre estas diversas visiones platónicas no nos afectan grandemente.

Existen, no obstante, otros puntos de vista matemáticos, como el que se conoce como *intuicionismo* (y que otros llaman *finitismo*), que van al otro extremo en donde se rechaza aceptar la existencia consumada de cualquier conjunto infinito *. El intuicionismo fue iniciado en 1924 por el matemático holandés L.E.J. Brouwer como una respuesta alternativa —diferente del formalismo— a las paradojas (como la de Russell) que pueden aparecer cuando se utilizan demasiado libremente los conjuntos infinitos en el razonamiento matemático. Las raíces de este punto de vista pueden rastrearse hasta Aristóteles, que había sido pupilo de Platón pero había rechazado los puntos de vista

* El intuicionismo fue llamado así debido a que se suponía que reflejaba el pensamiento humano.

de Platón acerca de la existencia absoluta de las entidades matemáticas y sobre la aceptabilidad de conjuntos infinitos. Según el intuicionismo, los conjuntos (infinitos o no) no deben pensarse como teniendo «existencia» por sí mismos sino que deben pensarse simplemente en términos de las reglas por las que se pueda determinar su pertenencia a ellos.

Una característica distintiva del intuicionismo de Brouwer es el rechazo de la «ley del tercio excluso». Esta ley afirma que la negación de la negación de un enunciado es equivalente a la afirmación de dicho enunciado. (En símbolos: $\sim(\sim P) \Leftrightarrow P$, una relación que encontramos antes.) * ¡Quizá Aristóteles no se hubiera sentido feliz con la negación de algo tan lógicamente «obvio» como esto! En términos ordinarios de «sentido común», la ley del tercio excluso puede considerarse como una verdad autoevidente: si es falso que algo no es verdadero, entonces este algo ¡es ciertamente verdadero! (Esta ley es la base del método matemático de «*reductio ad absurdum*», cf. p. 92.) Pero los intuicionistas se sienten capaces de negar esta ley. Esto se debe básicamente a que están adoptando una actitud diferente hacia el concepto de *existencia*, exigiendo que se presente una construcción (mental) definida antes de aceptar que un objeto matemático existe realmente. Por ello, para un intuicionista «existencia» significa «existencia constructiva». En un argumento matemático que procede por *reductio ad absurdum* uno desarrolla alguna hipótesis con la intención de mostrar que sus consecuencias conducen a una contradicción, contradicción que proporciona la deseada demostración de que la hipótesis en cuestión es falsa. La hipótesis podría tomar la forma de un enunciado acerca de que una entidad matemática con ciertas propiedades requeridas no existe. Cuando esto conduce a una contradicción, uno infiere, en *matemáticas ordinarias*, que la entidad requerida no existe realmente. Pero tal argumento, por sí mismo, no proporciona medios para *construir* efectivamente tal entidad. Para un intuicionista, este tipo de existencia no es existencia en absoluto; y es en ese sentido en el que rechazan aceptar la ley del tercio excluso y el método de *reductio ad absurdum*. En realidad, Brouwer estaba profundamente

* La ley $\sim(\sim P) \Leftrightarrow P$ se conoce más frecuentemente como «ley de la doble negación», reservándose el nombre de «ley del tercio excluso» para la ley $P \vee \sim P$ (o P es verdadera o P es falsa). Ambas son equivalentes en la lógica clásica ordinaria, aunque no se puede decir lo mismo en la lógica intuicionista que no es veritativo-funcional ni admite la interdefinición de las conectivas. Obviamente, la lógica intuicionista niega ambas leyes. [Nota del traductor.]

insatisfecho con tal «existencia» no-constructiva ⁷. Sin una construcción real, aseguraba, dicho concepto de existencia no tiene significado. ¡En la lógica brouweriana, de la falsedad de la no-existencia de un objeto no se puede deducir que exista realmente este objeto!

En mi opinión, aunque hay algo encomiable en buscar la constructividad en la existencia matemática, el punto de vista del intuicionismo de Brouwer es demasiado extremo. Brouwer expuso sus ideas por primera vez en 1924, mas de diez años antes del trabajo de Church y Turing. Ahora que el concepto de constructividad —en términos de la idea de computabilidad de Turing— puede estudiarse dentro del marco *convencional* de la filosofía matemática, no hay necesidad de llegar a los extremos a los que Brouwer quería llevarnos. Podemos discutir la constructividad como un tema separado de la cuestión de la existencia matemática. Si seguimos con el intuicionismo, debemos negarnos el uso de tipos de argumentos muy poderosos dentro de las matemáticas, y el tema se vuelve de algún modo agobiante e impotente.

No deseo extenderme sobre las diversas dificultades o absurdos aparentes a que nos lleva el punto de vista intuicionista, aunque quizá sea útil que mencione sólo algunos de los problemas. Un ejemplo citado a menudo por Brouwer se refiere a la expansión decimal de π :

$$3,141592653589793\dots$$

¿Existe una serie de diez sietes consecutivos en algún lugar de esta expansión, es decir

$$\pi = 3,141592653589793\dots 777777777\dots$$

o no existe tal serie? En términos matemáticos ordinarios, todo lo que podemos decir por ahora es que o existe o no existe —¡y no sabemos cuál de las dos posibilidades ocurre! Este parecería ser un enunciado bastante inofensivo. Sin embargo, los intuicionistas negarán que sea válido decir «o existe una serie de diez sietes consecutivos en algún lugar de la expansión decimal de π , o no existe» —¡al menos hasta que se haya establecido (de alguna manera constructiva aceptable para los intuicionistas) que realmente existe tal serie, o bien se haya establecido que no existe ninguna! Un cálculo directo bastaría para mostrar que una serie de diez sietes consecutivos existe en algún lugar de la expansión decimal de π , pero se necesitaría alguna especie de teorema matemático para establecer que no hay tal serie. Ningún computador ha llegado aún lo suficientemente lejos en el cálculo de π para determinar que existe una serie semejante. Lo que uno espe-

ría sobre bases probabilistas es que tal serie exista, pero aun si un computador pudiera calcular consistentemente dígitos a un ritmo de, digamos, uno por segundo ¡llevaría probablemente un tiempo del orden de entre cien y mil años encontrar la serie! Me parece mucho más probable que, antes que por computación directa, la existencia de dicha serie sea establecida matemáticamente (probablemente como un corolario de algún resultado mucho más poderoso e interesante) —¡aunque quizá no de una forma aceptable para los intuicionistas!

Este problema concreto no tiene interés matemático real. Sólo se da como un ejemplo que es fácil de explicar. En la forma de intuicionismo extremo de Brouwer, él afirmaría que, en el momento presente, la afirmación «existe una serie de diez siete consecutivos en algún lugar de la expansión decimal de π no es verdadera ni falsa. Si, en alguna fecha posterior, se estableciera el resultado correcto en uno u otro sentido, mediante computación o mediante demostración matemática (intuicionista), entonces la afirmación se convertiría en «verdadera» o «falsa», según fuera el caso. Un ejemplo similar sería el «último teorema de Fermat». Según el intuicionismo extremo de Brouwer, éste tampoco es verdadero ni falso en el tiempo presente, pero podría llegar a ser uno u otro en alguna fecha posterior. Para mí, tal subjetividad y dependencia temporal de la verdad matemática es inadmisibles. Es, en verdad, una cuestión muy subjetiva la de si, o cuando, un resultado matemático puede aceptarse como oficialmente «demostrado». La verdad matemática no debe descansar en semejantes criterios sociodependientes. Además, tener un concepto de la verdad matemática que cambia con el tiempo es, cuando menos, muy incómodo e insatisfactorio para una matemática que esperamos pueda ser empleada fiablemente en una descripción del mundo físico. No todos los intuicionistas adoptarían una posición tan estricta como la de Brouwer. De todas formas, el punto de vista intuicionista es decididamente incómodo, incluso para los que simpatizan con los objetivos del constructivismo. Pocos matemáticos actuales se alinean incondicionalmente con el intuicionismo, aunque sólo sea porque es muy restrictivo respecto a los tipos de razonamiento matemático que nos permite utilizar.

He descrito brevemente las tres corrientes principales de la filosofía matemática actual: formalismo, platonismo e intuicionismo. No he ocultado mis fuertes simpatías por el punto de vista platónico de que la verdad matemática es absoluta, externa y eterna, y no se basa en criterios hechos por el hombre; y que los objetos matemáticos tienen una existencia intemporal por sí mismos, independiente de la sociedad

humana o de objetos físicos particulares. He tratado de presentar mis argumentos a favor de este punto de vista en esta sección, en la sección anterior y al final del Capítulo 3. Espero que el lector estará preparado para seguir conmigo en este camino. Será importante para mucho de lo que encontraremos más adelante.

Teoremas tipo-Gödel a partir del resultado de Turing

En mi presentación del teorema de Gödel he omitido muchos detalles, y también he dejado de lado lo que históricamente fue la parte quizá más importante de su argumento: la que se refiere a la «indecibilidad» de la consistencia de los axiomas. Mi propósito aquí *no* ha sido el de subrayar este «problema de la demostrabilidad de la consistencia de los axiomas», tan importante para Hilbert y sus contemporáneos, sino mostrar que, utilizando nuestras intuiciones de los significados de las operaciones en cuestión, se *ve* claramente que una proposición de Gödel específica —ni demostrable ni indemostrable utilizando los axiomas y reglas del sistema formal considerado— ¡es una proposición *verdadera*!

He mencionado que Turing desarrolló su propio argumento posterior que establecía la insolubilidad del problema de la detención después de haber estudiado el trabajo de Gödel. Los dos argumentos tienen mucho en común y, de hecho, los aspectos claves del resultado de Gödel pueden derivarse directamente utilizando el método de Turing. Veamos cómo funciona esto, y de ahí obtendremos una intuición algo diferente sobre lo que hay detrás del teorema de Gödel.

Una propiedad esencial de un sistema matemático formal es que debería ser una cuestión computable el decidir si una cadena dada de símbolos constituye o no una demostración, dentro del sistema, de un enunciado matemático dado. Después de todo, la idea general en la formalización de la noción de demostración matemática es que no habrá que hacer juicios posteriores sobre lo que es un razonamiento válido y lo que no lo es. Debe ser posible verificar de una forma completamente mecánica y previamente determinada si una supuesta demostración es o no una demostración; es decir, debe haber un *algoritmo* para verificar demostraciones. Por otra parte, no exigimos que deba ser necesariamente una cuestión algorítmica el *encontrar* demostraciones (o refutaciones) de enunciados matemáticos propuestos.

Resulta, de hecho, que siempre *existe* un algoritmo para encontrar

una demostración dentro de cualquier sistema matemático formal en el que dicha demostración exista. En efecto, supongamos que nuestro sistema está formulado en términos de algún lenguaje simbólico, que a su vez es expresable en términos de algún «alfabeto» finito de símbolos. Como antes, ordenemos *lexicográficamente* nuestras cadenas de símbolos, lo que significa, recordemos, hacer una ordenación alfabética para cada longitud de cadena dada, ordenando primero alfabéticamente todas las cadenas de longitud uno, a continuación las de longitud dos, luego las de longitud tres y así sucesivamente (p. 145). De esta forma tenemos todas las demostraciones correctamente construidas ordenadas numéricamente según un esquema lexicográfico. Teniendo nuestra lista de demostraciones, tenemos también una lista de *todos* los teoremas del sistema formal, pues los teoremas son precisamente las proposiciones que aparecen en la última línea de las demostraciones correctamente construidas. El listado es perfectamente computable: en efecto, podemos considerar la lista lexicográfica de *todas* las cadenas de símbolos del sistema, tengan o no sentido como demostraciones, y luego comprobar la primera cadena con nuestro algoritmo de verificar demostraciones para ver si es una demostración y descartarla si no lo es; luego comprobamos la segunda de la misma manera y la descartamos si no es una demostración, y luego la tercera, luego la cuarta y así sucesivamente. De este modo, si existe una demostración acabaremos por encontrarla en algún lugar de la lista.

Por consiguiente, si Hilbert hubiera tenido éxito en encontrar su sistema matemático —un sistema de axiomas y reglas de inferencia lo bastante fuerte para que podamos decidir, mediante demostración formal, la verdad o falsedad de cualquier proposición matemática correctamente formulada dentro del sistema— entonces *existiría* un método algorítmico general para decidir la verdad de cualquiera de estas proposiciones. ¿Por qué es esto así? Lo es porque si, mediante el procedimiento esbozado arriba, podemos llegar a encontrar la proposición que estamos buscando como la línea final en la demostración, entonces hemos *demostrado* esa proposición. Si, por el contrario, llegamos a encontrar la *negación* de nuestra proposición como la línea final, entonces la hemos *refutado*. Si el esquema de Hilbert fuera completo siempre ocurriría una u otra de estas eventualidades (y, si fuera consistente, nunca ocurrirían las dos juntas). Así, nuestro procedimiento mecánico terminará siempre en algún paso y tendríamos un algoritmo universal para decidir la verdad o falsedad de todas las proposiciones del sistema. Esto contradiría el resultado de Turing, como se presentó en el Capítulo 2, de que no existe algoritmo general para decidir pro-

posiciones matemáticas. Por consiguiente hemos demostrado, en efecto, el teorema de Gödel de que *ningún* esquema del tipo propuesto por Hilbert puede ser completo en el sentido que hemos estado discutiendo.

En realidad, el teorema de Gödel es más concreto que esto puesto que el tipo de sistema formal en el que Gödel estaba interesado debía ser apropiado sólo para proposiciones de la aritmética, y no para proposiciones de las matemáticas en general. ¿Podemos disponer que todas las operaciones necesarias de las máquinas de Turing puedan realizarse utilizando sólo la aritmética? Para decirlo de otra manera, ¿podemos expresar todas las funciones *computables* de números naturales (es decir, las funciones recursivas, o algorítmicas —los resultados de la acción de una máquina de Turing) en términos de aritmética ordinaria? Realmente es casi cierto que podemos, pero no del todo. Necesitamos añadir una operación extra a las reglas estándar de la aritmética y la lógica (incluyendo $\exists$ y $\forall$). Esta operación selecciona simplemente

«el menor número natural x tal que $K(x)$ es verdadero»,

donde $K()$ es cualquier función proposicional dada calculable aritméticamente —para la que se supone que *existe* tal número, es decir, que $\exists x[K(x)]$ es verdadero. (Si no existiera tal número, entonces nuestra operación seguiría «actuando indefinidamente» * tratando de localizar el no-existente x requerido.) En cualquier caso, el argumento precedente establece, sobre la base del resultado de Turing, que el programa de Hilbert de reducir ramas enteras de las matemáticas a cálculos dentro de algún sistema formal es realmente insostenible.

Tal como están las cosas, este método no prueba tan fácilmente que tengamos una proposición de Gödel (como $P_k(k)$) que es *verdadera*, aunque no demostrable dentro del sistema. Sin embargo, si recordamos el argumento dado en el Capítulo 2 sobre «cómo superar a un algoritmo» (cf. p. 98), veremos que podemos hacer algo muy parecido. En dicho argumento pudimos probar que, dado un algoritmo para decidir si la acción de una máquina de Turing se detiene, podemos plantear una acción de una máquina de Turing que *nosotros* vemos que no se detiene, aunque el algoritmo no puede hacerlo.

* Es realmente esencial que se permita que puedan darse estas desafortunadas posibilidades, de modo que podamos describir potencialmente *cualquier* operación algorítmica. Recuérdese que para describir máquinas de Turing en general, debemos permitir que haya máquinas de Turing que no se detengan nunca.

(Recuérdese que insistíamos en que el algoritmo debe informarnos correctamente de cuándo se *detendrá* la acción de una máquina de Turing, si bien a veces puede fallar al no decirnos que la acción de una máquina de Turing no se detendrá —pues él mismo seguirá funcionando indefinidamente.) De este modo, como sucedía en la situación anterior con el teorema de Gödel, tenemos una proposición que podemos ver, mediante una intuición, que debe ser realmente verdadera (no-parada de la acción de la máquina de Turing) aunque la acción algorítmica dada no es capaz de decírnoslo.

Conjuntos recursivamente enumerables

Hay una manera de describir los ingredientes básicos de los resultados de Turing y Gödel de un modo gráfico, en términos del lenguaje de la *teoría de conjuntos*. Esto nos permite alejarnos de las descripciones arbitrarias en términos de simbolismos específicos o sistemas formales, de modo que queden de manifiesto los resultados esenciales. Consideraremos sólo conjuntos (finitos o infinitos) de *números naturales* $0, 1, 2, 3, 4, \dots$, de manera que examinamos colecciones de éstos, tales como $\{4, 5, 8\}$, $\{0, 57, 100003\}$, $\{6\}$, $\{1, 2, 3, 4, \dots, 999\}$, $\{1, 2, 3, 4, \dots\}$, $\{0, 2, 4, 6, 8, \dots\}$, o incluso el conjunto total $\mathbf{N} = \{0, 1, 2, 3, 4, \dots\}$ o el conjunto vacío $\emptyset = \{ \}$. Nos interesaremos sólo en cuestiones de *computabilidad*, a saber: ¿«Qué tipos de conjuntos de números naturales pueden ser generados mediante algoritmos y cuáles no?»

Para abordar este asunto podemos pensar, si así lo deseamos, que cada simple número natural n denota una cadena específica de símbolos en un sistema formal particular. Ésta sería la « n -ésima» cadena de símbolos, llamémosla Q_n , de acuerdo con una cierta ordenación lexicográfica de las proposiciones del sistema (expresadas de forma «sintácticamente correcta»). Entonces cada número natural representa una proposición. El conjunto de *todas* las proposiciones del sistema formal estará representado por el conjunto total $\mathbf{N}$ y, por ejemplo, los teoremas del sistema formal serán imaginados constituyendo un conjunto más pequeño de números naturales, digamos el conjunto P . Sin embargo, los detalles de cualquier sistema particular de numeración de las proposiciones no son importantes. Todo lo que necesitaremos, para establecer una correspondencia entre números naturales y proposiciones, será un algoritmo conocido para obtener cada proposición Q_n (escrita en la notación simbólica adecuada) a partir de su

correspondiente número natural n , y otro algoritmo conocido para obtener n a partir de Q_n . Dados estos dos algoritmos conocidos estamos en libertad de *identificar* el conjunto de los números naturales $\mathbf{N}$ con el conjunto de proposiciones de un sistema formal específico.

Escojamos un sistema formal que sea consistente y suficientemente amplio para incluir todas las acciones de todas las máquinas de Turing —y, más aún, «razonable»— en el sentido de que sus axiomas y reglas de inferencia son cosas que pueden asumirse como «autoevidentemente verdaderas». Ahora, *algunas* de las proposiciones $Q_0, Q_1, Q_2, Q_3, \dots$ del sistema formal tendrán *demostraciones* dentro del sistema. Estas proposiciones «demostrables» tendrán números que constituirán algún conjunto en $\mathbf{N}$, de hecho el conjunto P de «teoremas» antes considerado. En efecto, ya hemos visto que existe un *algoritmo* para generar, una detrás de otra, todas las proposiciones con demostración en algún sistema formal dado. (Como se esbozó antes, la « n -ésima demostración» Π_n se obtiene algorítmicamente a partir de n . Todo lo que tenemos que hacer es mirar la última línea de la n -ésima demostración para encontrar la n -ésima proposición demostrable dentro del sistema, es decir, el n -ésimo «teorema».) De este modo tenemos un algoritmo para generar los elementos de P uno detrás de otro (quizá con repeticiones —esto no supone ninguna diferencia).

Un conjunto, tal como P , que puede generarse de esta manera mediante algún algoritmo se llama *recursivamente enumerable*. Nótese que el conjunto de las proposiciones que son *indemostrables* dentro del sistema —es decir, las proposiciones cuya negación es demostrable— es también recursivamente enumerable, pues basta con enumerar simplemente las proposiciones demostrables y tomar sus negaciones a medida que avanzamos. Hay muchos otros subconjuntos de $\mathbf{N}$ que son recursivamente enumerables, y no necesitamos hacer referencia a nuestro sistema formal para definirlos. Ejemplos sencillos de conjuntos recursivamente enumerables son el conjunto de los números pares

$$\{0, 2, 4, 6, 8, \dots\},$$

el conjunto de los cuadrados

$$\{0, 1, 4, 9, 16, \dots\},$$

y el conjunto de los primos

$$\{2, 3, 5, 7, 11, \dots\}.$$

Evidentemente podemos generar cada uno de estos conjuntos por medio de un algoritmo. En cada uno de estos tres ejemplos sucederá *también* que el complemento del conjunto —es decir, el conjunto de los números naturales que *no* están en el conjunto— es recursivamente enumerable. Los conjuntos complementarios en estos tres casos son, respectivamente:

$$\{1,3,5,7,9,\dots\};$$

$$\{2,3,5,6,7,8,10,\dots\};$$

y

$$\{0,1,4,6,8,9,10,12,\dots\}.$$

Sería un asunto sencillo proporcionar también un algoritmo para estos conjuntos complementarios. De hecho, para cualquier número natural n dado, podemos decidir algorítmicamente si es o no un número par, o si es o no un cuadrado, o si es o no primo. Esto nos proporciona un algoritmo para generar *ambos*, el conjunto y el conjunto complementario, pues podemos recorrer los números naturales uno a uno y decidir en cada caso si pertenece al conjunto original o al conjunto complementario. Un conjunto que tiene la propiedad de que tanto él como su complementario son recursivamente enumerables se denomina conjunto *recursivo*. Evidentemente el complemento de un conjunto recursivo es también un conjunto recursivo.

Ahora bien, ¿existen conjuntos que sean recursivamente enumerables pero *no* recursivos? Hagamos una pausa, por un momento, para señalar lo que esto supone. Puesto que los elementos de tal conjunto pueden ser generados mediante un algoritmo, tendremos un medio de decidir si un elemento sospechoso de pertenecer al conjunto —y que, supongamos por el momento, sí pertenece al conjunto— *está*, en efecto, en el conjunto. Todo lo que tenemos que hacer es permitir que nuestro algoritmo actúe sobre los elementos del conjunto hasta que finalmente llegue al elemento particular que estamos examinando. Pero supongamos ahora que nuestro elemento sospechoso *no* está realmente en el conjunto. En este caso nuestro algoritmo no nos dará resultado, pues seguirá actuando indefinidamente sin llegar nunca a una decisión. Para eso necesitamos un algoritmo que genere el conjunto *complementario*. Si *este* algoritmo encuentra a nuestro sospechoso, entonces estamos seguros de que el elemento no está en el conjunto. Con ambos algoritmos estaríamos en buenas condiciones. Podríamos alternar ambos algoritmos y atrapar al sospechoso de una

forma u otra. Esta situación feliz es la que ocurre con un conjunto *recursivo*. Pero ahora estamos considerando un conjunto que es sólo recursivamente enumerable pero *no* recursivo: ¡nuestro presunto algoritmo para generar el conjunto complementario no existe! Entonces se nos presenta la curiosa situación de que, para un elemento *perteneciente* al conjunto, podemos decidir algorítmicamente que efectivamente *está* en el conjunto, ¡pero no podemos, mediante un algoritmo, decidir esta cuestión con garantías para elementos que *no* están en el conjunto!

¿Se presenta alguna vez tan curiosa situación? ¿Existen realmente, conjuntos recursivamente enumerables que no son recursivos? En concreto, ¿qué hay sobre el conjunto P ? ¿Es un conjunto recursivo? Sabemos que es recursivamente enumerable, así que lo que tenemos que decidir es si el conjunto complementario es también recursivamente enumerable. ¡De hecho no lo es! ¿Cómo podemos decir esto? Bien, recuérdese que se ha supuesto que las acciones de las máquinas de Turing figuran entre las operaciones permitidas dentro de nuestro sistema formal. Denotemos por T_n la n -ésima máquina de Turing. Entonces el enunciado

« $T_n(n)$ se para»

es una proposición —escribámosla $S(n)$ — que podemos expresar en nuestro sistema formal, para cada número natural n . La proposición $S(n)$ será verdadera para algunos valores de n , y falsa para otros. El conjunto de *todas* las $S(n)$, cuando n recorre los números naturales $0, 1, 2, 3, \dots$ estará representado como algún subconjunto S de $\mathbf{N}$. Recuérdese ahora el resultado fundamental de Turing (Capítulo 2, p. 92) de que no hay algoritmo que afirme « $T_n(n)$ no se para» precisamente en estos casos en los que efectivamente $T_n(n)$ no se para. Esto prueba que el conjunto de las $S(n)$ *falsas* *no* es recursivamente enumerable.

Observamos que la parte de S que está en P consta precisamente de aquellas $S(n)$ que son *verdaderas*. ¿Por qué es esto así? Ciertamente, si una $S(n)$ particular es demostrable, entonces debe ser verdadera (¡debido a que hemos escogido un sistema formal «razonable»!), así que la parte de S que está en P debe constar solamente de proposiciones $S(n)$ *verdaderas*. Más aún, ninguna proposición $S(n)$ verdadera puede estar fuera de P , pues si $T_n(n)$ se para entonces podemos dar una demostración dentro del sistema de que efectivamente lo hace *.

* La demostración constaría, de hecho, de una serie de pasos que reflejarán la

Supongamos ahora que el complemento de P fuera recursivamente enumerable. Entonces tendríamos algún algoritmo para generar los elementos de este conjunto complementario. Podemos poner en marcha este algoritmo y anotar todas las proposiciones $S(n)$ que encontremos. Todas éstas son las $S(n)$ falsas, así que nuestro método nos da efectivamente una enumeración recursiva del conjunto de $S(n)$ falsas. Pero señalábamos arriba que las $S(n)$ falsas *no* son recursivamente enumerables. Esta contradicción establece que, después de todo, el complemento de P no puede ser enumerado recursivamente; por consiguiente, el conjunto P *no es recursivo*, que es lo que queríamos establecer.

Estas propiedades prueban que nuestro sistema formal no puede ser completo: debe haber proposiciones que no son demostrables ni refutables dentro del sistema. En efecto, si no hubiera tales proposiciones «indecidibles», el complemento del conjunto P tendría que ser el conjunto de proposiciones *refutables* (todo lo que no es demostrable tendría que ser refutable). Pero hemos visto que las proposiciones refutables constituyen un conjunto recursivamente enumerable, de modo que P sería *recursivo*. Sin embargo, P *no* es recursivo —una contradicción que establece la incompletitud requerida. Ésta es la principal estocada del teorema de Gödel.

¿Qué sucede ahora con el subconjunto T de $\mathbf{N}$ que representa las proposiciones *verdaderas* de nuestro sistema formal? ¿Es T recursivo? ¿Es T recursivamente enumerable? ¿Es el complemento de T recursivamente enumerable? La respuesta a todas estas preguntas es «No». Una manera de verlo consiste en notar que las proposiciones falsas de la forma

$$\langle T_n(n) \text{ se para} \rangle$$

no pueden ser generadas mediante un algoritmo, como hemos señalado más arriba. En consecuencia, las proposiciones falsas en conjunto no pueden ser generadas mediante un algoritmo puesto que semejante algoritmo debería enumerar, en particular, todas las anteriores proposiciones $\langle T_n(n) \text{ se para} \rangle$ falsas. Análogamente, el conjunto de todas las proposiciones *verdaderas* no puede ser generado mediante un algoritmo (puesto que semejante algoritmo podría ser modificado de forma trivial para producir todas las proposiciones falsas sin más que hacerle tomar la *negación* de cada proposición que genera). Pues-

acción de la máquina funcionando hasta que se pare. La demostración quedará completada en cuanto se pare la máquina.

to que las proposiciones verdaderas no son entonces recursivamente enumerables (y tampoco lo son las falsas) aquellas constituyen una colección enormemente más complicada y profunda que la de las proposiciones demostrables dentro del sistema. Esto ilustra, una vez más, aspectos del teorema de Gödel: que el concepto de *verdad* matemática es sólo parcialmente accesible a través del argumento formal.

Existen, no obstante, algunos casos sencillos de proposiciones aritméticas verdaderas que forman conjuntos recursivamente enumerables. Por ejemplo, las proposiciones verdaderas de la forma

$$\exists w, x, \dots, z [f(w, x, \dots, z) = 0],$$

donde $f()$ es una función construida a partir de las operaciones aritméticas ordinarias de adición, sustracción, multiplicación, división y elevación a una potencia, constituyen un conjunto recursivamente enumerable (que denominaré A), como no es difícil de ver. Un ejemplo de una proposición de esta forma —aunque no sabemos si es cierta— es la negación del «último teorema de Fermat», para lo que podemos tomar $f()$ dada por

$$f(w, x, y, z) = (x + 1)^{w+3} + (y + 1)^{w+3} - (z - 1)^{w+3}$$

Sin embargo, el conjunto A resulta ser no recursivo (un hecho que *no* es tan fácil de ver —aunque es una consecuencia del auténtico argumento original de Gödel). ¡Por lo tanto, no disponemos de ningún algoritmo mediante el que, en principio, podamos decidir la verdad o falsedad del «último teorema de Fermat»!



Fig. 4.1. Representación muy esquemática de un conjunto recursivo.

En la Figura 4.1 he tratado de representar esquemáticamente un conjunto recursivo como una región con un contorno sencillo, de modo que uno puede imaginar que se puede decir directamente si un punto dado pertenece o no al conjunto. Podemos pensar que cada punto de la figura representa un número natural. El conjunto complementario está entonces representado también como una región de apariencia sencilla. En la Figura 4.2 he tratado de representar un conjunto recursivamente enumerable pero *no*-recursivo como un conjunto con un contorno complicado, en donde se supone que el conjunto a un lado del contorno —el lado recursivamente enumerable— parece más sencillo que el que está al otro lado. Las figuras son muy esquemáticas, y no se pretende que sean «geométricamente precisas» en ningún sentido. ¡En particular, no tiene especial significación el hecho de que estas figuras hayan sido representadas en un plano bidimensional!

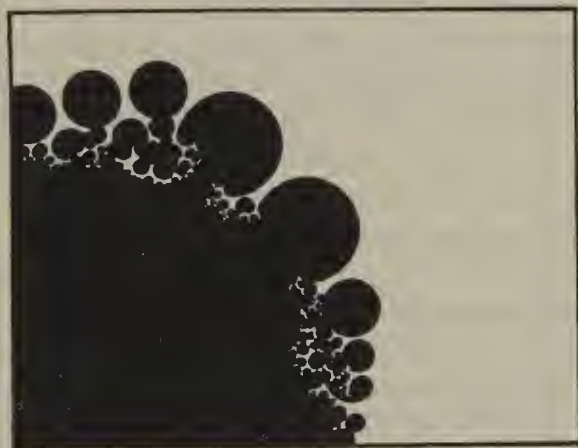


Fig. 4.2. Representación muy esquemática de un conjunto recursivamente enumerable (región negra) que no es recursivo. La idea consiste en que la región blanca está definida sólo como «lo que queda» cuando se elimina la región negra generada computablemente; y no es asunto computable el verificar que un punto está realmente en la región blanca.

En la Figura 4.3 he indicado esquemáticamente cómo las regiones P , T y A yacen dentro del conjunto N .

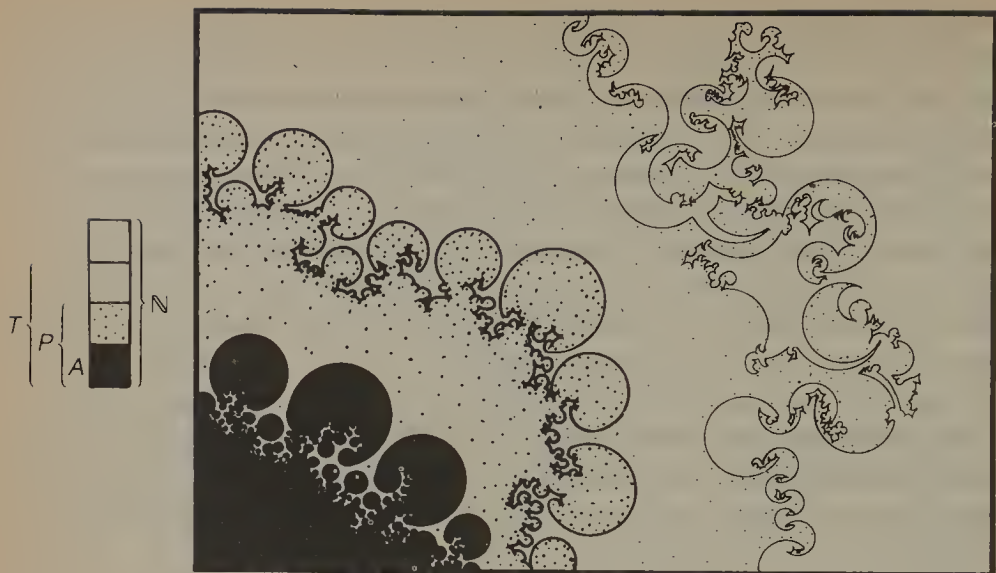


Fig. 4.3. Representación muy esquemática de varios conjuntos de proposiciones. El conjunto P de las proposiciones que son demostrables dentro del sistema es, como el A , recursivamente enumerable pero no recursivo; el conjunto T de las proposiciones verdaderas no es siquiera recursivamente enumerable

¿Es recursivo el conjunto de Mandelbrot?

Los conjuntos no recursivos deben tener la propiedad de ser esencialmente complicados. Su complejidad debe desafiar, en cierto sentido, todos los intentos de sistematización o, de lo contrario, esa misma sistematización llevaría a algún método algorítmico apropiado. En el caso de un conjunto no-recursivo no hay una vía algorítmica general para decidir si un elemento (o «punto») pertenece o no al conjunto. Ahora bien, al comienzo del Capítulo 3 fuimos testigos de un cierto conjunto de apariencia extraordinariamente complicada: el conjunto de Mandelbrot. Aunque las reglas que proporcionan su definición son sorprendentemente simples, el propio conjunto exhibe una variedad sin fin de estructuras altamente elaboradas. ¿Podría ser éste un ejemplo de conjunto no-recursivo que se muestra verdaderamente ante nuestros ojos mortales?

El lector no tardará en notar, no obstante, que este paradigma de complicación ha sido conjurado, para que nuestros ojos puedan verlo, mediante la magia de la moderna tecnología de computadores de alta velocidad. ¿No son los computadores electrónicos las encarnaciones

mismas de la acción algorítmica? Efectivamente esto es cierto, pero debemos considerar de qué forma produce realmente estas imágenes el computador. Para verificar si un punto del plano de Argand —un número complejo c — pertenece al conjunto de Mandelbrot (en color negro) o al conjunto complementario (en color blanco), el computador empezará con 0, luego aplica la iteración

$$z \rightarrow z^2 + c$$

a $z = 0$ para obtener c , y luego a $z = c$ para obtener $c^2 + c$ y luego a $c^2 + c$ para obtener $c^4 + 2c^3 + c^2 + c$, y así sucesivamente. Si esta secuencia $0, c, c^2 + c, c^4 + 2c^3 + c^2 + c, \dots$ permanece acotada, entonces el punto representado por c se colorea en negro; en caso contrario se colorea en blanco. ¿Cómo determina la máquina si la secuencia permanece o no permanece acotada? ¡En principio, la pregunta implica saber qué sucede tras un *infinito* número de términos de la secuencia! Esto, en sí mismo, no es una materia computable. Afortunadamente, hay formas de determinar, tras sólo un número finito de términos, cuándo la secuencia se ha hecho no acotada. (De hecho, en cuanto alcanza el círculo de radio $1 + \sqrt{2}$ centrado en el origen podemos estar seguros de que la secuencia es no acotada.)

Así, en cierto sentido, el *complemento* del conjunto de Mandelbrot (es decir, la región *blanca*) es recursivamente enumerable. Si el número complejo c está en la región blanca entonces hay un algoritmo para verificar este hecho. ¿Qué sucede con el propio conjunto de Mandelbrot —la región negra? ¿Existe un algoritmo que nos asegure que un punto sospechoso de estar en la región negra, está efectivamente en la región negra? Parece ser que, por el momento, no se conoce la respuesta a esta pregunta. He consultado a varios colegas y expertos, y ninguno parece saber de tal algoritmo; ni tampoco han tropezado con ninguna demostración de que no exista tal algoritmo. Al menos, no parece haber ningún algoritmo *conocido* para la región negra. ¡Quizá el complemento del conjunto de Mandelbrot es realmente un ejemplo de conjunto recursivamente enumerable que no es recursivo!

Antes de explorar más a fondo esta sugerencia será necesario abordar algunos asuntos que no he mencionado. Estos temas tendrán importancia para nosotros en nuestras discusiones finales sobre la computabilidad en física. Realmente he sido algo inexacto en la discusión precedente. He aplicado términos como «recursivamente enumerable» y «recursivo» a conjuntos de puntos en el plano de Argand, es decir, a conjuntos de números complejos. Estos términos deberían utilizarse

estrictamente sólo para los números naturales o para otros conjuntos *numerables*. Hemos visto en el Capítulo 3 (p. 119) que los números reales no son numerables y, por tanto, tampoco lo pueden ser los números complejos —puesto que los números reales pueden considerarse como tipos particulares de números complejos, a saber, números complejos con parte imaginaria nula (cf. p. 124). De hecho, hay exactamente «tantos» números complejos como números reales, a saber, «*C*» de ellos. (Para establecer una relación uno-a-uno entre los números complejos y los números reales podemos, hablando en términos aproximados, tomar las expansiones decimales de las partes real e imaginaria de cada número complejo e intercalarlas para dar los dígitos impares y pares del correspondiente número real: v.g. el número complejo $3,6781... + i512,975...$ correspondería al número real $50132.6977851...$)

Un modo de evitar este problema sería referirnos sólo a números complejos *computables*, pues vimos en el Capítulo 3 que los números reales computables —y, por tanto, también los números complejos computables— son numerables. Sin embargo, existe una seria dificultad en esto: ¿no existe ningún algoritmo general para decidir si dos números computables, dados en términos de sus respectivos algoritmos, son iguales entre sí o no lo son! (Podemos formar algorítmicamente su diferencia pero no podemos decidir algorítmicamente si esta diferencia es o no cero. Imaginemos dos algoritmos que generan los dígitos $0,99999...$ y $1,00000...$ respectivamente; nunca podremos saber si los 9s, o los 0s continuarán indefinidamente, y por tanto los números son iguales, o si finalmente aparecerá algún otro dígito, y los números son desiguales.) Por consiguiente, nunca podríamos saber si estos números son iguales. Una consecuencia de esto es que incluso con un conjunto tan sencillo como el *disco unidad* en el plano de Argand (el conjunto de puntos cuya distancia al origen no es mayor que la unidad, es decir, la región negra en la Fig. 4.4.) no habría algoritmo para decidir con certeza si un número complejo yace en el disco. El problema no se presenta con los puntos en el interior (o con los puntos fuera del disco) sino sólo con puntos que yacen sobre el mismo borde del disco. Supongamos que se nos da simplemente un algoritmo que genera los dígitos de las partes real e imaginaria de algún número complejo. Si sospechamos que este número complejo yace realmente sobre el círculo unidad no podemos verificarlo necesariamente. No existe algoritmo para decidir si el número computable

$$x^2 + y^2$$

es realmente igual a 1 o no, siendo éste el criterio para decidir si el número complejo $x + iy$ yace o no sobre el círculo unidad.

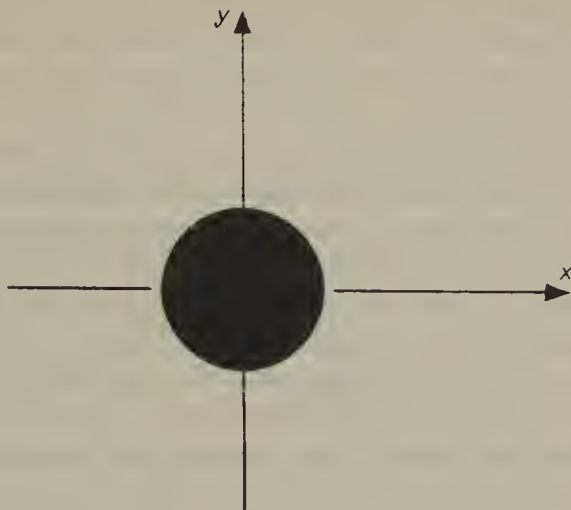


Fig. 4.4. El disco unidad contaría ciertamente entre los conjuntos «recursivos», pero esto requiere un punto de vista adecuado.

Evidentemente no es esto lo que buscamos. ¡El disco unidad *debería* contar ciertamente entre los recursivos! ¡No hay muchos conjuntos más simples que el disco unidad! Una manera de soslayar este problema podría ser la de *no considerar* el contorno. Para puntos que están estrictamente en el interior, o estrictamente en el exterior, existen algoritmos para verificarlo. (No hay más que generar los dígitos de $x^2 + y^2$ uno tras otro, y encontrar finalmente un dígito diferente de 9 tras el punto decimal en 0,99999... o diferente de 0 en 1,00000...) En este sentido el disco unidad es recursivo. Pero este punto de vista es bastante incómodo para las matemáticas puesto que con frecuencia necesitamos expresar argumentos en términos de lo que sucede *en* el contorno. Es posible, por el contrario, que semejante punto de vista fuera apropiado para la física. Tendremos que considerar de nuevo este tipo de resultado más adelante.

Podríamos adoptar otro punto de vista estrechamente relacionado que no hace referencia en absoluto a números complejos computables. En lugar de tratar de numerar los números complejos en el interior o el exterior del conjunto en cuestión, podemos simplemente buscar un algoritmo que decida, *dado* el número complejo, si yace en

el conjunto o si yace en el complemento del conjunto. Por «dado» quiero decir que, para cada número complejo que estamos verificando, se nos presentan uno tras otro —quizá por arte de magia— los sucesivos dígitos de las partes real e imaginaria hasta donde queramos. No exijo que exista un algoritmo, conocido o desconocido, para *presentar* estos dígitos. Un conjunto de números complejos se consideraría «recursivamente enumerable» siempre que exista un solo algoritmo tal que *cualquiera* que se le presentase de esta forma una sucesión de dígitos semejante respondiera finalmente «sí», tras un número finito de pasos, si y sólo si el número complejo yace realmente en el conjunto. Al igual que sucedía con el primer punto de vista sugerido arriba, resulta que este nuevo punto de vista «omite» los contornos. Entonces, tanto el interior del disco unidad como el exterior del disco unidad contarán como enumerablemente recursivos en este sentido, mientras que el propio contorno no lo haría.

No está completamente claro para mí que cualquiera de estos puntos de vista sea realmente el que necesitamos ⁹. Cuando se aplica al conjunto de Mandelbrot, la filosofía de «omitir el contorno» puede dejar fuera mucha de la complejidad del conjunto. Este conjunto consta en parte de «gotas» —regiones con interiores— y en parte de «zarcillos». La complicación más extrema parece estar en los zarcillos, que pueden retorcerse violentamente. Sin embargo, los zarcillos no están en el interior del conjunto y, por tanto, serían «omitidos» si adoptáramos una cualquiera de estas dos filosofías. Pese a todo, todavía no está claro que el conjunto de Mandelbrot sea «recursivo» aun cuando sólo se consideren las gotas. La cuestión parece descansar en cierta conjetura indemostrada relativa al conjunto de Mandelbrot: ¿es éste lo que se denomina «localmente conexo»? No me propongo explicar aquí el significado o la relevancia de este término. Quiero indicar simplemente que éstos son temas difíciles y plantean cuestiones relativas al conjunto de Mandelbrot que todavía no están resueltas, algunas de las cuales están en la vanguardia de la investigación matemática en curso.

Podríamos adoptar también otros puntos de vista para evitar el problema que plantea el hecho de que los números complejos no son numerables. En lugar de considerar *todos* los números complejos computables podemos considerar un subconjunto apropiado de tales números que tengan la propiedad de que *es* una cuestión computable el decidir si dos de ellos son o no iguales. Un subconjunto sencillo sería el de los números complejos «*racionales*» en el que las partes real e imaginaria de los números son ambas números racionales. No creo

que esto eliminara muchos de los zarcillos del conjunto de Mandelbrot aun cuando este punto de vista es muy restrictivo. Algo más satisfactorio podría ser el considerar los números *algebraicos* —aquellos números complejos que son soluciones de ecuaciones algebraicas con coeficientes enteros. Por ejemplo, todas las soluciones de

$$129z^7 - 33z^5 + 725z^4 + 16z^3 - 2z - 3 = 0$$

son números algebraicos. Los números algebraicos son numerables y computables, y es ciertamente una cuestión computable el decidir si dos de ellos son o no iguales. (Resulta que muchos de ellos yacen sobre el contorno del círculo unidad y en los zarcillos del conjunto de Mandelbrot.) Podemos, si así lo deseamos, plantear en sus términos la cuestión de si el conjunto de Mandelbrot es o no recursivo.

Pudiera ser que los números algebraicos sean apropiados en el caso de los dos conjuntos recién considerados pero, en general, no resuelven todas nuestras dificultades. En efecto, consideremos el conjunto (la región negra de la Fig. 4.5) definido por la relación

$$y \geq e^x$$

para $x + iy (= z)$ en el plano de Argand. Tanto el interior del conjunto como el interior del complemento del conjunto son ambos recursivamente enumerables con arreglo a cualquiera de los puntos de vista expresados más arriba, pero (como se deduce de un famoso teorema demostrado en 1882 por F. Lindemann) el contorno, $y = e^x$,

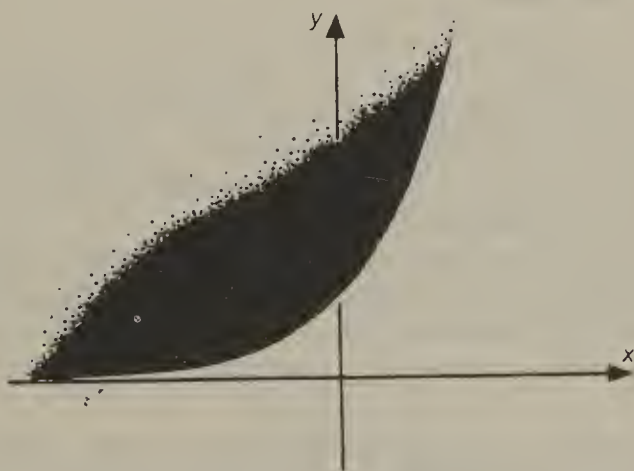


Fig. 4.5. El conjunto definido por la relación exponencial $y \geq e^x$, debería contar también entre los «recursivos».

contiene sólo *un* punto algebraico, a saber, el punto $z = i$. ¡Los números algebraicos no nos sirven de ayuda en este caso para explorar la naturaleza algorítmica del contorno! No sería difícil encontrar otra subclase de números computables que bastaran en este caso particular, pero nos quedamos con la sensación de que todavía no ha sido alcanzado el punto de vista correcto.

Algunos ejemplos de matemáticas no-recursive

Existen muchas áreas de las matemáticas en las que aparecen problemas que son no recursivos. Así, se nos puede presentar una clase de problemas para los que la respuesta en cada caso es o «sí» o «no», pero para los que no existe ningún algoritmo general que decida cuál de los dos es el caso. Algunas de estas clases de problemas tienen una apariencia notablemente sencilla.

Consideremos, en primer lugar, el problema de encontrar soluciones enteras de sistemas de ecuaciones algebraicas con coeficientes enteros. Tales ecuaciones se llaman ecuaciones *diofánticas* (en honor del matemático griego Diofanto, que vivió en el siglo III a. de C. y que estudió este tipo de ecuaciones). Un conjunto de tales ecuaciones podría ser

$$z^3 - y - 1 = 0, \quad yz^2 - 2x - 2 = 0, \quad y^2 - 2xz + z + 1 = 0$$

y el problema consiste en decidir si tienen solución para valores *enteros* de x , y , z . De hecho, en este caso particular la tienen, siendo una solución la dada por

$$x = 13, \quad y = 7, \quad z = 2.$$

Sin embargo no existe algoritmo para decidir esta cuestión para un conjunto arbitrario * de ecuaciones diofánticas: ¡la aritmética diofántica, pese a la naturaleza elemental de sus ingredientes, es parte de las matemáticas no-algorítmicas!

(Un ejemplo algo menos elemental lo constituye la *equivalencia topológica de variedades*. Lo menciono sólo brevemente, debido a que tiene alguna posible relevancia para los temas discutidos en el Capítulo 8. Para entender lo que es una «variedad» consideremos primero

* Esto da una respuesta negativa al décimo problema de Hilbert, mencionado en p. 61. (Ver, por ejemplo, Devlin 1988.) Aquí, el número de variables no está restringido. Sin embargo, se sabe que no son necesarias más de nueve para que sea válida esta propiedad no-algorítmica.

un lazo de cuerda, que es una variedad de *una* sola dimensión, y consideremos luego una superficie cerrada, una variedad de *dos* dimensiones. A continuación tratemos de imaginar un tipo de «superficie» que pueda tener *tres* o un número mayor de dimensiones. «Equivalencia topológica» de dos variedades significa que una de ellas puede ser deformada hasta hacerla coincidir con la otra mediante un movimiento continuo —sin rasgar ni pegar. Por ejemplo, una superficie esférica y la superficie de un cubo son topológicamente equivalentes, mientras que ambas son topológicamente no-equivalentes a la superficie de un anillo o de una taza de té —siendo estas dos últimas topológicamente equivalentes entre sí. Ahora bien, para variedades bidimensionales existe un algoritmo para decidir si dos de ellas son o no topológicamente equivalentes —lo que corresponde, en realidad, a contar el número de «asas» que tiene cada superficie. En el caso de tres dimensiones no se conoce la respuesta a esta cuestión en el momento de escribir este libro, pero para cuatro o más dimensiones *no* existe algoritmo para decidir la equivalencia. El caso tetradimensional es posiblemente de cierta importancia para la física puesto que, según la teoría de la relatividad general de Einstein, espacio y tiempo constituyen conjuntamente una 4-variedad; ver Capítulo 5, p. 268. Geroch y Hartle, 1987, han sugerido que esta propiedad no-algorítmica podría tener relevancia para la «gravitación cuántica»; cf. también Capítulo 8.)

Consideremos un problema de tipo diferente, llamado el *problema de las palabras* ¹⁰. Supongamos que tenemos algún alfabeto de símbolos y consideremos varias cadenas de estos símbolos, que llamaremos *palabras*. Las palabras mismas no necesitan tener ningún significado pero se dará cierta lista (finita) de «igualdades» entre ellas que podemos utilizar para derivar otras nuevas «igualdades». Esto se hace sustituyendo palabras de la lista inicial en otras palabras (normalmente más largas) que las contienen como porciones. Cada una de estas porciones puede ser reemplazada por otra porción que, según la lista, se considera igual a ella. El problema entonces es decidir, para un par de palabras dado, si son o no «iguales» con arreglo a estas reglas.

Como ejemplo podríamos tener para nuestra lista inicial

LAS = AS
 ASO = A
 NASO = RON
 SAN = LIRON
 GAS = DEL

A partir de estas podemos derivar, por ejemplo

$$\text{BOA} = \text{BOLA}$$

mediante el uso de sustituciones sucesivas de la segunda, la primera, y de nuevo la segunda de las relaciones de la lista inicial:

$$\text{BOA} = \text{BOASO} = \text{BOLASO} = \text{BOLA}$$

El problema ahora es: dado un par de palabras, ¿podemos obtener una a partir de la otra utilizando simplemente estas sustituciones? ¿Podemos, por ejemplo, ir de GASOLINA a DEAN o, pongamos por caso, de GASTAR a DELATAR? En el primer caso la respuesta resulta ser «sí», mientras que en el segundo caso es «no». Cuando la respuesta es «sí» la manera normal de probarlo sería simplemente mostrando una cadena de igualdades en la que cada palabra se obtiene de la precedente mediante el uso de una relación permitida. Así (indicando en **negrita** las letras que van a ser cambiadas, y en *itálica* las letras que acabamos de cambiar):

$$\begin{aligned} \text{GASOLINA} &= \text{GALINA} = \text{GALINASO} = \text{GALIRON} \\ &= \text{GASAN} = \text{DELAN} = \text{DELASON} = \text{DEASON} = \text{DEAN} \end{aligned}$$

¿Cómo podemos decir que es imposible ir de GASTAR a DELATAR por medio de las reglas permitidas? Para esto necesitamos pensar un poco más, pero no es difícil de ver que hay varias formas de hacerlo. La más simple parece ser la siguiente: en cada «igualdad» de nuestra lista inicial, el número de As más el número de Rs más el número de Ds es el mismo en cada lado. Por tanto, el número total de As, Rs y Ds no puede cambiar a lo largo de cualquier sucesión de sustituciones permitidas. Sin embargo, dicho número es 3 para GASTAR mientras que es 4 para DELATAR. Por consiguiente, no hay manera de ir de GASTAR a DELATAR mediante sustituciones permitidas.

Nótese que cuando las dos palabras son «iguales» podemos mostrarlo exhibiendo simplemente una cadena formal de símbolos permitidos, utilizando las reglas que nos han sido dadas; mientras que en el caso en que son «desiguales» tenemos que recurrir a argumentos *acerca* de las reglas que nos han sido dadas. Hay un algoritmo evidente que podemos utilizar para establecer la «igualdad» entre palabras, siempre que las palabras *sean* efectivamente «iguales». Todo lo que tenemos que hacer es un listado lexicográfico de todas las secuencias posibles de palabras, y luego tachar de la lista todas las cadenas en las que hay un par de palabras consecutivas en donde la segunda

no se sigue de la primera mediante ninguna regla permitida. Las secuencias restantes proporcionan todas las «igualdades» buscadas entre palabras. Sin embargo, no existe, en general, un algoritmo tan obvio para decidir cuando dos palabras *no* son «iguales», y tenemos que recurrir a la «inteligencia» para establecer el hecho. (Realmente necesité algún tiempo antes de reparar en el «truco» anterior para establecer que GASTAR y DELATAR no son «iguales». Con otro ejemplo podría ser necesario un «truco» bastante diferente. La inteligencia, dicho sea de paso, es también útil —aunque no necesaria— para establecer la *existencia* de una «igualdad»).

De hecho, para la lista *particular* de las cinco «igualdades» que constituyen la lista en el caso anterior no es excesivamente difícil dar un algoritmo para verificar si dos palabras son «desiguales» cuando efectivamente son «desiguales». Sin embargo, ¡para *encontrar* el algoritmo que funciona en este caso necesitamos ejercitar un poco la inteligencia! En realidad, resulta que no hay un único algoritmo que se pueda utilizar universalmente para *todas* las elecciones posibles de la lista inicial. En este sentido no hay solución algorítmica para el problema de las palabras. ¡El problema general de las palabras pertenece a las matemáticas no-recursivas!

Existen incluso ciertas selecciones *particulares* de la lista inicial para las que no existe ningún algoritmo que decida cuándo dos palabras son desiguales. Una de estas viene dada por

$$\begin{aligned}AH &= HA \\OH &= HO \\AT &= TA \\OT &= TO \\TAI &= IT \\HOI &= IH \\THAT &= ITHT\end{aligned}$$

(Esta lista está adaptada de una dada en 1955 por G. S. Tseitin y Dana Scott; ver Gardner 1985, p. 144.) Por lo tanto, este problema particular de las palabras constituye *por sí mismo* un ejemplo de matemáticas no-recursivas, en el sentido de que utilizando esta lista inicial particular no podemos decidir algorítmicamente si dos palabras dadas son o no «iguales».

El problema general de las palabras surgió a partir de consideraciones de lógica matemática formalizada («sistemas formales» etc., como consideramos anteriormente). La lista inicial juega el papel de un sistema de axiomas y la regla de sustitución de palabras, el papel

de las reglas formales de inferencia. La demostración de la no-recursividad del problema de las palabras surge de tales consideraciones.

Como ejemplo final de un problema matemático que es no-recursivo consideremos la cuestión del recubrimiento del plano euclídeo con formas poligonales, en donde se nos dan un número finito de formas diferentes y se pregunta si es posible recubrir completamente el plano, sin huecos ni solapamientos, utilizando solamente estas formas y no otras. Una disposición de formas semejantes se denomina una *teselación* del plano. Estamos familiarizados con el hecho de que estas teselaciones son posibles si sólo se utilizan cuadrados, o sólo triángulos equiláteros, o sólo hexágonos regulares (como se ilustra en la Fig. 10.2, Capítulo 10), pero no si sólo se utilizan pentágonos regulares. Habrá muchas otras formas que teselarán el plano, tales como cada uno de los dos pentágonos *irregulares* ilustrados en la Figura 4.6. Con un *par* de formas las teselaciones pueden ser mucho más elaboradas. En la Figura 4.7 se dan dos ejemplos sencillos. Todos los ejemplos dados hasta ahora tienen la propiedad de ser *periódicos*, lo que significa que son exactamente repetitivos en dos direcciones independientes. En términos matemáticos decimos que existe un *paralelogramo período* —un paralelogramo tal que, si lo marcamos de alguna manera y luego lo repetimos una y otra vez en las dos direcciones paralelas a sus lados, reproducirá el patrón de la teselación dada. En la Figura 4.8 se muestra un ejemplo en el que una teselación periódica con una tesela en forma de espina es representada a la izquierda y relacionada con un paralelogramo período cuya teselación periódica se muestra a la derecha.

Ahora bien, existen muchas teselaciones del plano que *no* son periódicas. La Figura 4.9 muestra tres teselaciones «espirales» no-periódicas con la misma tesela en forma de espina que la de la Figura 4.8. Esta forma particular de tesela se conoce como «versátil» (¡por razones obvias!) *, y fue ideada por B. Grünbaum y G. C. Shepard (1981, 1987), aparentemente basada en una forma anterior debida a H. Voderberg. Nótese que el versátil teselará *tanto* periódica *como* no-periódicamente. Esta propiedad es compartida por muchas otras formas de tesela y conjuntos de formas de tesela. ¿Existen teselas simples o conjuntos de teselas que teselen el plano *sólo* no-periódicamente? La respuesta a esta pregunta es «sí». En la Figura 4.10 he representado un conjunto de seis teselas construido por el matemático

* Juego de palabras entre «tile» = tesela y «versatile» = versátil (Nota del traductor.)

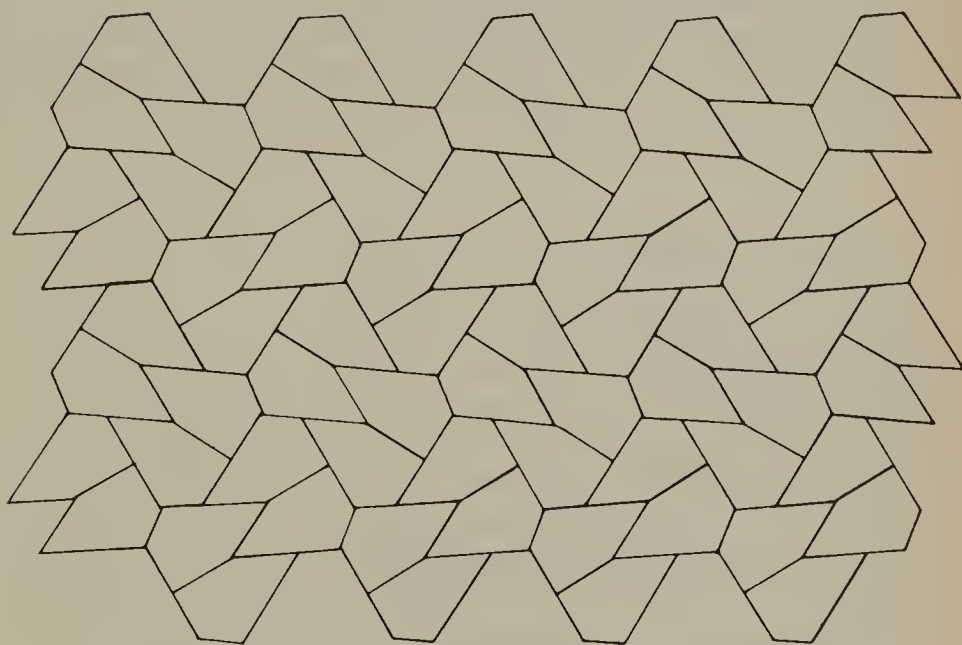
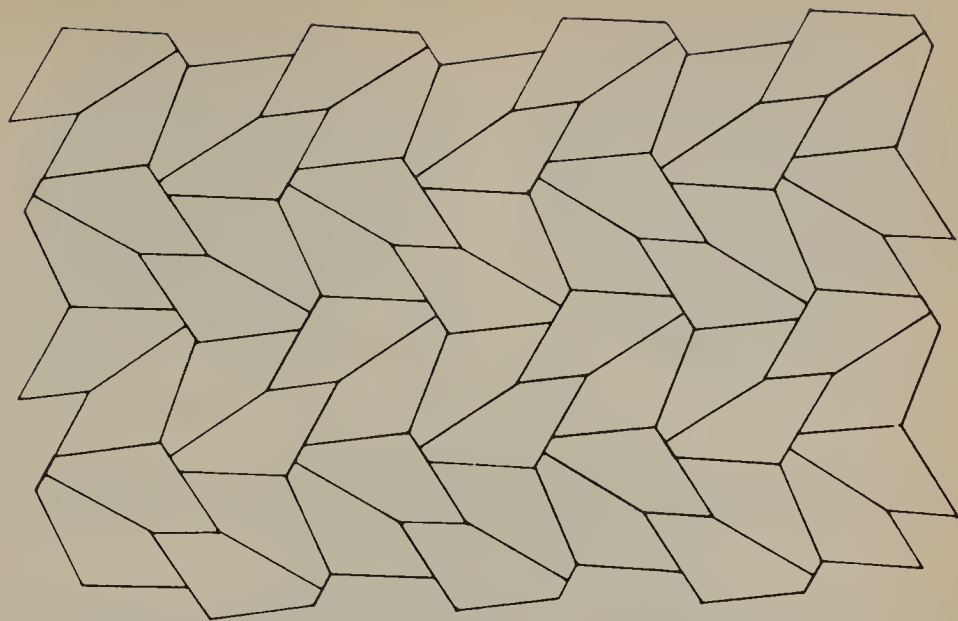


Fig. 4.6. Dos ejemplos de teselaciones periódicas del plano, utilizando una sola forma cada una de ellas (descubierto por Majorie Rice en 1976).

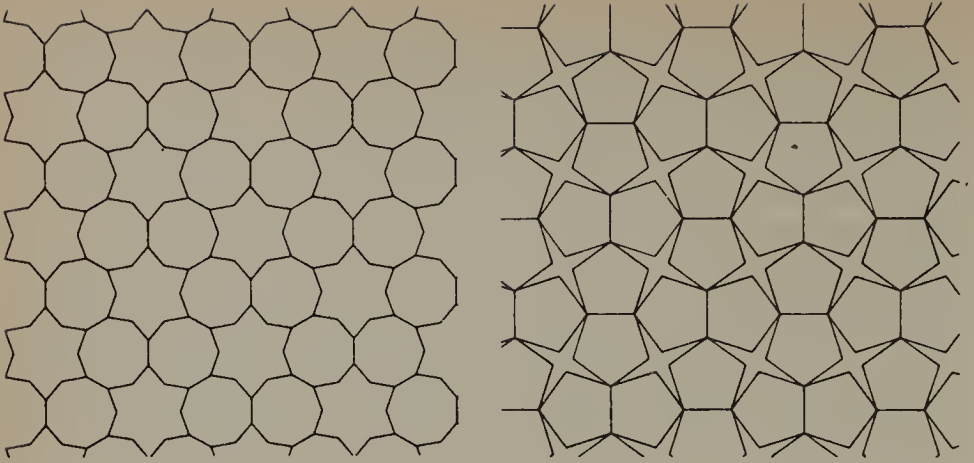


Fig. 4.7. Dos ejemplos de teselaciones periódicas del plano, utilizando dos formas cada una.

americano Raphael Robinson (1971) que teselan el plano entero pero sólo de manera no-periódica.

Merece la pena entrar un poco en la historia de cómo surgió este conjunto no-periódico de teselas (cf. Grunbaum y Shepard 1987). En 1961 el lógico chino-americano Hao Wang abordó la cuestión de si hay un *procedimiento de decisión* para el problema de la teselación,

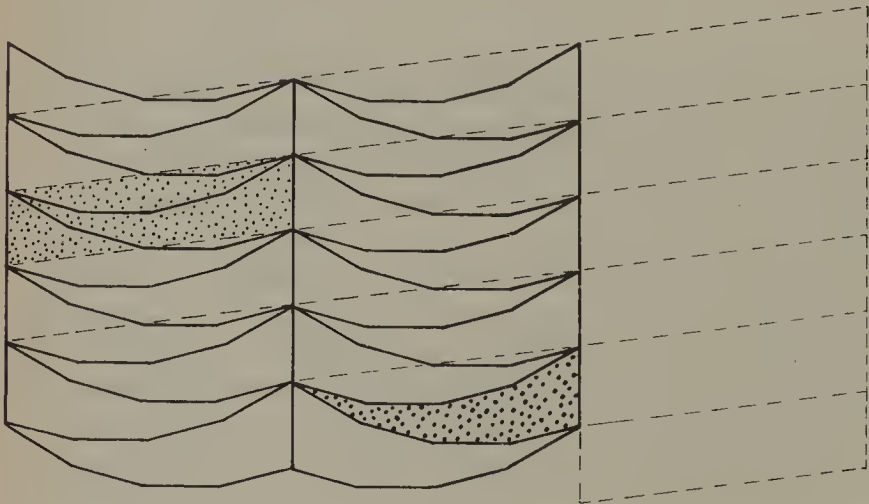


Fig. 4.8. Una teselación periódica, ilustrada con relación a su paralelogramo período.

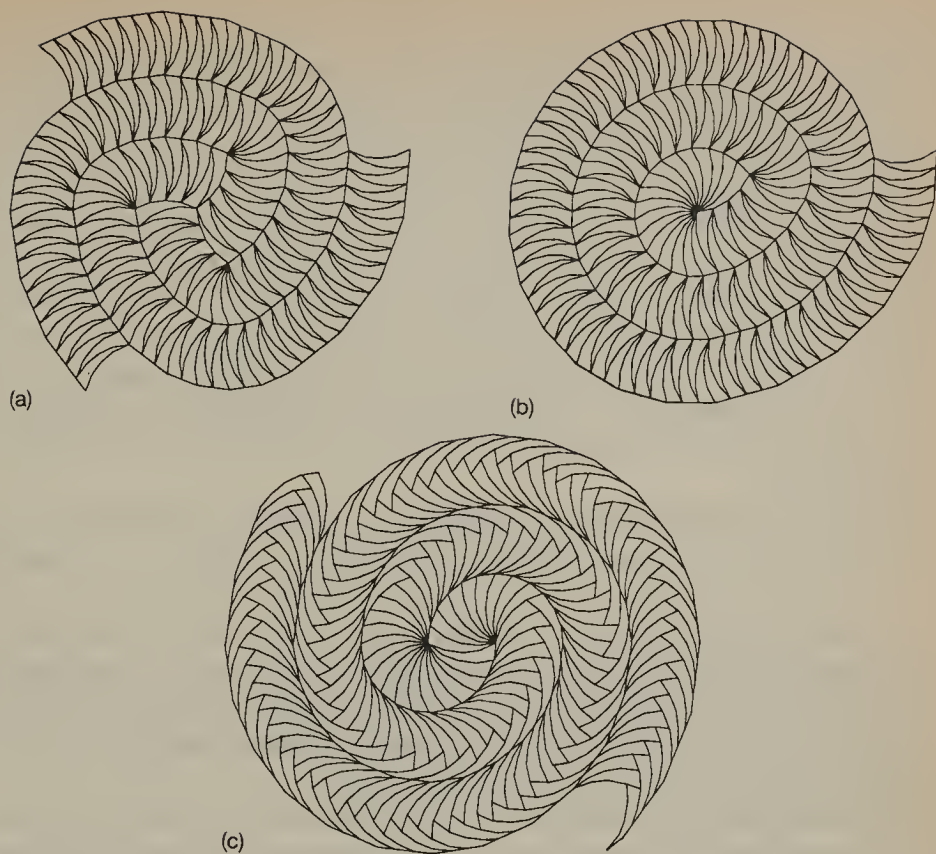


Fig. 4.9. Tres teselaciones «espirales» no periódicas, utilizando la misma forma «versátil» que se utilizó en la Fig. 4.8.

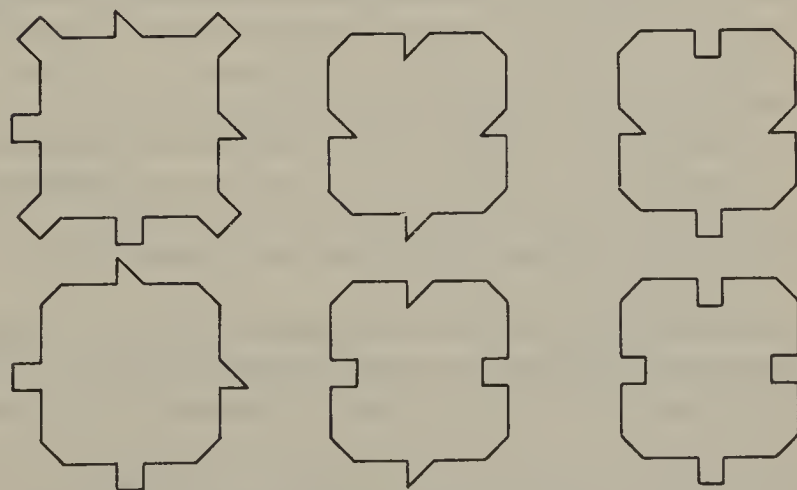


Fig. 4.10. Las seis teselas de Raphael Robinson que teselan el plano sólo de forma no periódica.

es decir, si existe un *algoritmo* para decidir si un conjunto finito dado de formas poligonales diferentes teselará ¡el plano entero! * Wang fue capaz de demostrar que efectivamente existiría dicho procedimiento de decisión *si* se pudiera probar que todo conjunto finito de teselas diferentes que tesele el plano de alguna forma, también teselará el plano periódicamente. Yo pienso que en esa época existía la sensación generalizada de que era poco probable que pudiera existir un conjunto que violase esta condición —es decir, un conjunto «aperiódico» de teselas. Sin embargo, en 1966, Robert Berger fue capaz de probar, siguiendo algunas de las líneas que Hao Wang había sugerido, que *no* existe procedimiento de decisión para el problema de la teselación: ¡el problema de la teselación forma también parte de las matemáticas no-recursivas! ¹¹

Por consiguiente, se sigue del resultado previo de Hao Wang que debe existir un conjunto aperiódico de teselas, y Berger fue capaz de mostrar el primer conjunto aperiódico de teselas. No obstante, debido a lo complicado de esta línea argumental, su conjunto incluía un número desorbitadamente alto de teselas diferentes —originalmente 20426. Con algún ingenio adicional, Berger pudo reducir su número a 104. Más tarde, en 1971, Raphael Robinson pudo rebajar el número hasta las seis mostradas en la Figura 4.10.

En la Figura 4.11 se muestra otro conjunto aperiódico de seis teselas. Yo mismo lo presenté en 1973 siguiendo una línea de pensamiento bastante diferente. (Volveré a este tema en el Capítulo 10 en cuya Figura 10.3 se representa un mosaico teselado con estas formas.) Después de que el conjunto aperiódico de seis teselas de Robinson hubiera llamado mi atención, empecé a pensar sobre la manera de reducir el número, y mediante varias operaciones de recortado y repegado pude reducirlo a dos. En la Figura 4.12 se muestran dos esquemas alternativos. Las estructuras necesariamente no-periódicas que exhiben las teselaciones completas tienen muchas propiedades notables, incluyendo una estructura cuasi-periódica con simetría quintuple aparentemente imposible en cristalografía. Volveré más tarde a estos temas.

Es tal vez notable que un área de las matemáticas aparentemente «trivial» como ésta —a saber, el recubrimiento del plano con formas congruentes—, que parece casi un «juego de niños», forme efectiva-

* En realidad, Hao Wang consideró un problema ligeramente diferente, con teselas cuadradas, sin rotación, y empalmando bordes coloreados; pero aquí no nos importan las diferencias.

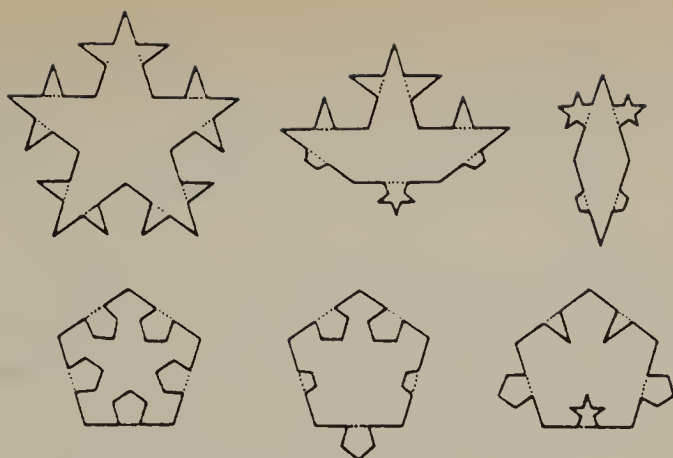


Fig. 4.11. Otro conjunto de seis teselas que tesela el plano sólo de forma no-periódica.

mente parte de las matemáticas no-recursivas. De hecho, hay muchos problemas difíciles y no resueltos en este área. No se sabe, por ejemplo, si existe un conjunto aperiódico que conste de *una sola* tesela.

El problema de la teselación, tal como lo tratan Wang, Berger y Robinson, utiliza teselas basadas en cuadrados. Yo estoy admitiendo aquí polígonos de forma general, y se necesita alguna manera adecuadamente computable de mostrar las teselas individuales. Una manera de hacerlo sería dar sus vértices como puntos en el plano de Argand, y estos puntos pueden ser dados de un modo perfectamente adecuado como números algebraicos.

¿Es el conjunto de Mandelbrot semejante a la matemática no-recursiva?

Volvamos ahora a nuestra discusión anterior del conjunto de Mandelbrot. Voy a suponer, con propósito de ilustración, que el conjunto de Mandelbrot es no-recursivo en cierto sentido apropiado. Puesto que su complemento es recursivamente enumerable esto querría decir que el propio conjunto no sería recursivamente enumerable. Pienso que es probable que la forma del conjunto de Mandelbrot tenga algunas lecciones que enseñarnos sobre la naturaleza de los conjuntos no-recursivos y las matemáticas no-recursivas.

Volvamos a la Figura 3.2 que ya encontramos en el Capítulo 3.

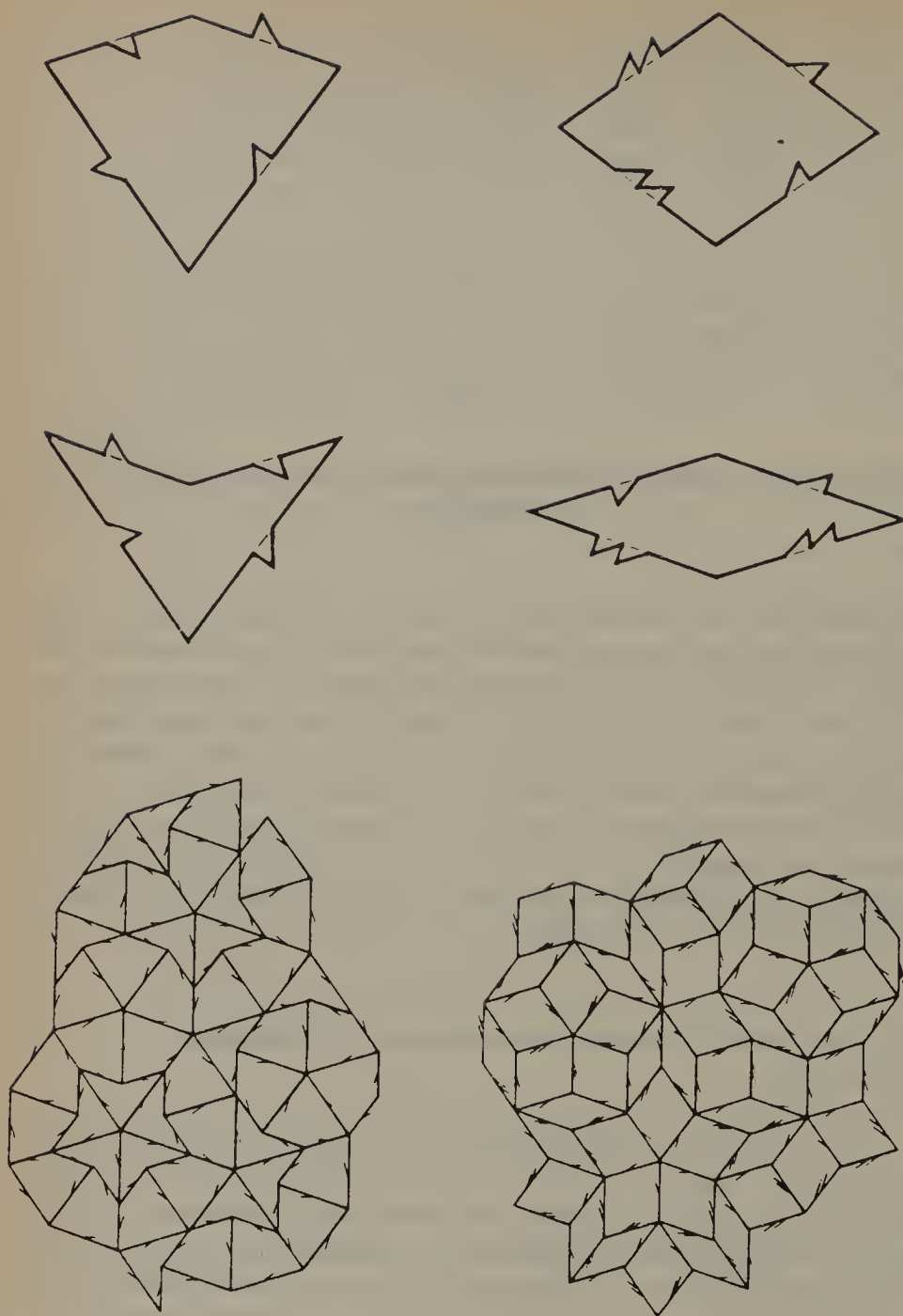


Fig. 4.12. Dos pares, cada uno de los cuales teselará sólo de forma no-periódica («teselas de Penrose»), y las regiones del plano teseladas con cada par.

Nótese que la mayor parte del conjunto parece estar ocupada por una extensa región en forma de corazón, que he llamado A, en la Figura 4.13. La forma se conoce como un *cardioide* y su región interior puede definirse matemáticamente como el conjunto de puntos c del plano de Argand que surgen de la expresión

$$c = z - z^2$$

donde z es un número complejo cuya distancia al origen es menor que $1/2$. Este conjunto es, en efecto, recursivamente enumerable en el sentido sugerido antes: existe un algoritmo tal que, cuando se aplica a un punto en el interior de la región, verificará que el punto está efectivamente en dicha región interior. El algoritmo real se obtiene fácilmente a partir de la fórmula anterior.

Consideremos ahora la región en forma de disco inmediatamente a la izquierda del cardioide principal (región B en la Fig. 4.13). Su interior es el conjunto de puntos

$$c = z - 1$$

donde z dista del origen menos de $1/4$. Esta región es realmente el interior de un disco, es decir, el conjunto de puntos dentro de un círculo exacto. De nuevo esta región es recursivamente enumerable en el sentido anterior. ¿Qué sucede con las otras «verrugas» del car-

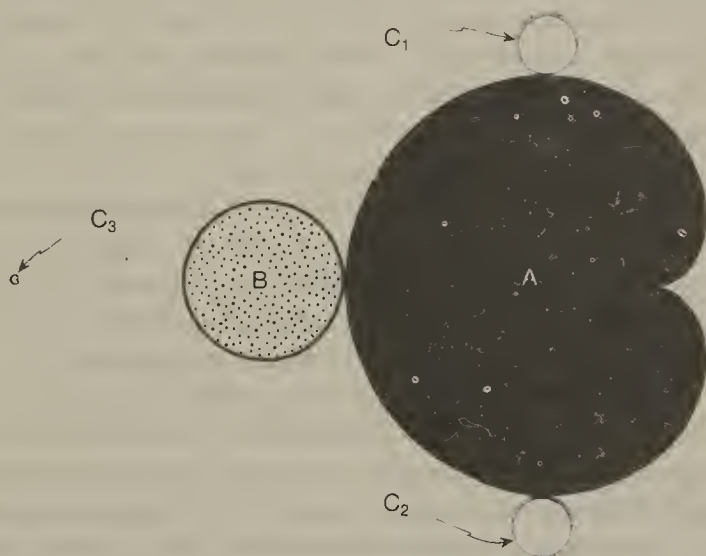


Fig. 4.13. Las partes principales del interior del conjunto de Mandelbrot pueden ser definidas mediante sencillas ecuaciones algorítmicas.

dioide? Consideremos las dos verrugas que siguen en tamaño. Estas son las gotas más o menos circulares que aparecen aproximadamente encima y debajo del cardioide en la Figura 3.2 y están marcadas como C_1 , C_2 en la Figura 4.13. Vienen dadas en términos del conjunto

$$c^3 + 2c^2 + (1 - z)c + (1 - z)^2 = 0$$

en donde ahora z recorre la región que está a una distancia $1/8$ del origen. En realidad, esta ecuación nos da no solamente estas dos gotas (a la vez) sino también la forma «bebé» de tipo cardioide que aparece a la izquierda en la Figura 3.2 —la región principal de la Figura 3.1— y es la región marcada C_3 en la Figura 4.13. De nuevo estas regiones (juntas o separadas) constituyen conjuntos recursivamente enumerables (en el sentido sugerido antes) en virtud de la existencia de la fórmula anterior.

A pesar de la sugerencia que he estado haciendo de que el conjunto de Mandelbrot puede ser no-recursive, hemos sido capaces de vaciar ya las áreas mayores del conjunto con algún algoritmo perfectamente bien definido y no-demasiado-complicado. Parece que este proceso continuará. Todas las regiones más visibles del conjunto —y ciertamente un porcentaje abrumador de su área (si no toda ella)— puede tratarse algorítmicamente. Si, como estoy suponiendo, el conjunto completo es realmente no-recursive, entonces las regiones que no pueden alcanzarse mediante nuestros algoritmos deben ser muy delicadas y difíciles de encontrar. Más aún, cuando hayamos localizado una de estas regiones existe la probabilidad de que podamos ver entonces cómo mejorar nuestros algoritmos para que pueda llegarse también a aquellas regiones concretas. Pero entonces habría *otras* regiones (si mi suposición de la no-recursiveidad es correcta) ocultas aún más profundamente en oscuridades de sutileza y complejidad, que ni siquiera nuestro algoritmo mejorado puede alcanzar. Una vez más, mediante esfuerzos prodigiosos de intuición, ingenio y aplicación podríamos ser capaces de localizar una de estas regiones; pero todavía habría otras que se nos escaparían, y así sucesivamente.

Pienso que esto no es diferente de la manera frecuente de proceder de las matemáticas en áreas en las que los problemas son difíciles y presumiblemente no-recursive. Los problemas más comunes con los que probablemente podemos tropezar en algún área específica pueden con frecuencia ser tratados con sencillos procedimientos algorítmicos —procedimientos que pueden haber sido conocidos durante siglos. Pero algunos escaparán de la red y se necesitan procedimientos más sofisticados para tratarlos. Por supuesto, los que aún escapan

intrigarán particularmente a los matemáticos y les aguijonearán para desarrollar métodos aún más poderosos. Éstos tendrían que basarse en intuiciones cada vez más profundas sobre la naturaleza de las matemáticas implicadas. Quizá hay algo de esto en nuestra comprensión del mundo físico.

En los problemas de las palabras y de la teselación considerados antes podemos empezar a captar algo de este tipo de cosa (aunque no son áreas en las que la maquinaria matemática haya sido aún desarrollada en una gran medida). Pudimos utilizar un argumento muy sencillo en un caso particular para probar que una determinada palabra no puede ser obtenida a partir de otra mediante las reglas permitidas. No es difícil imaginar que entrarán en juego líneas argumentales mucho más sofisticadas para tratar los casos más complejos. Existiría entonces la posibilidad de que estas nuevas líneas de razonamiento puedan desarrollarse en un procedimiento algorítmico. Sabemos que ningún procedimiento puede ser suficiente para todos los ejemplos del problema de las palabras, aunque los ejemplos que escapan tendrían que ser contruidos con mucho cuidado y sutileza. En realidad, desde el momento en que *sepamos* como se construyen estos ejemplos —desde el momento que sepamos con certeza que un caso particular ha eludido nuestro algoritmo— entonces podremos mejorar nuestro algoritmo para incluir también este caso. Solamente pueden escapar pares de palabras que no son «iguales», de modo que en cuanto sepamos que han escapado, sabemos que no son «iguales», y este hecho puede ser añadido a nuestro algoritmo. ¡Nuestra intuición mejorada conduciría a un algoritmo mejorado!

Teoría de la complejidad

Los argumentos que he dado más arriba, y en los capítulos precedentes, respecto a la naturaleza, existencia y limitaciones de los algoritmos han sido en gran medida a un nivel «de principios». No he discutido en absoluto la cuestión de si los algoritmos que aparecen pueden ponerse en práctica de algún modo. Incluso para problemas en donde es evidente que existen algoritmos y cómo pueden construirse, puede ser necesario mucho ingenio y trabajo arduo para desarrollar tales algoritmos en forma utilizable. A veces un poco de intuición e ingenio conduce a una reducción considerable en la complejidad de un algoritmo y, a veces, a mejoras absolutamente enormes en su velocidad. Estas cuestiones son a menudo muy detalladas y

técnicas, y en los últimos años se ha realizado un gran trabajo en muchos contextos diferentes a propósito de la construcción, comprensión y mejora de los algoritmos —un área de trabajo en rápida expansión y desarrollo. No sería pertinente que intentase entrar en una discusión detallada de tales cuestiones. Sin embargo, existen diversas generalidades que son conocidas, o conjeturadas, acerca de ciertas limitaciones *absolutas* al incremento en la velocidad de un algoritmo. Resulta que, incluso entre problemas matemáticos que *son* de naturaleza algorítmica, existen algunas clases de problemas que son intrínsecamente mucho más difíciles de resolver algorítmicamente que otros. Los difíciles sólo pueden ser resueltos mediante algoritmos muy lentos (o, quizá, mediante algoritmos que requieren una cantidad desorbitadamente grande de espacio de almacenamiento, etc.). La teoría que trata cuestiones de este tipo se llama *teoría de la complejidad*.

La teoría de la complejidad se interesa no tanto en la dificultad de resolver algorítmicamente problemas *individuales*, sino en familias infinitas de problemas tales que habría un algoritmo general para encontrar respuestas a todos los problemas de una familia. Los diferentes problemas de la familia tendrían diferentes «tamaños», donde el tamaño de un problema viene medido por algún número natural n . (Dentro de un momento tendré más que decir sobre el modo en que este número n caracteriza realmente el tamaño del problema.) La longitud de tiempo —o, más correctamente, el número de pasos elementales— que necesitará el algoritmo para cada problema particular de la clase vendrá dada por algún número natural N que depende de n . Para ser un poco más precisos digamos que, considerados *todos* los problemas de un tamaño particular n , el mayor número de pasos que necesita el algoritmo es N . Ahora bien, a medida que n se hace más y más grande, es probable que el número N también se haga más y más grande. De hecho, es probable que N crezca mucho más rápidamente que n . Por ejemplo, N podría ser aproximadamente proporcional a n^2 , o a n^3 , o quizá a 2^n (que, para n grande, es mucho mayor que cada uno de los n , n^2 , n^3 , n^4 y n^5 —mayor, de hecho, que n^r para cualquier número r dado), o incluso N podría ser aproximadamente proporcional a, pongamos por caso, 2^{2^n} (que todavía es mucho mayor).

El número de «pasos» podría depender, por supuesto, del tipo de máquina computadora en que fuera ejecutado el algoritmo. Si la máquina computadora es una máquina de Turing del tipo descrito en el Capítulo 2, en el que sólo hay una cinta —lo que es bastante poco eficaz— entonces el número N podría crecer más rápidamente (es decir, la máquina podría funcionar más lentamente) que si se permi-

tieran dos o más cintas. Para evitar equívocos de este tipo, se hace una amplia clasificación de las posibles maneras de las que crece N como función de n , de modo que, independientemente del tipo de máquina de Turing que se utilice, la medida de la tasa de crecimiento de N caiga siempre en la misma categoría. Una categoría semejante, denominada **P** (que significa «tiempo polinómico»), incluye todas las tasas que son, como mucho, múltiplos fijos* de uno de los n , n^2 , n^3 , n^4 , n^5 ... Es decir, para cualquier problema de categoría **P** (donde por «problema» entiendo realmente una familia de problemas con un algoritmo general para resolverlos), tenemos

$$N \leq K \times n^r$$

siendo los números K y r *constantes* (independientes de n). Esto significa que N no es mayor que algún múltiplo de n elevado a una potencia fija.

Un tipo simple de problema que ciertamente pertenece a **P** es el de multiplicar dos números. Para explicar esto describiré primero la forma en que el número n caracteriza el tamaño del par concreto de números que van a ser multiplicados. Podemos imaginar que cada número se escribe en la notación binaria y que $n/2$ es simplemente el número de dígitos binarios de cada número, dando un *total* de n dígitos binarios —es decir, n *bits*. (Si uno de los números es mayor que el otro podemos sencillamente hacer empezar el más corto con una sucesión de ceros para adaptarlo a la longitud del más grande.) Por ejemplo, si $n = 14$, podríamos estar considerando

$$1011010 \times 0011011$$

(que es 1011010×11011 , pero con ceros añadidos a la cifra más corta). El modo más directo de llevar a cabo esta multiplicación es escribirla:

* Un «polinomio» se refiere realmente a una expresión más general, tal como $7n^4 - 3n^3 + 6n + 15$, pero esto no nos da más generalidad. Para cualquiera de estas expresiones, todos los términos que incluyen potencias menores de n pierden importancia cuando n se hace grande (de modo que en este ejemplo concreto podemos ignorar todos los términos excepto $7n^4$).

$$\begin{array}{r}
 1011010 \\
 \times 0011011 \\
 \hline
 1011010 \\
 1011010 \\
 0000000 \\
 1011010 \\
 1011010 \\
 0000000 \\
 0000000 \\
 \hline
 010010111110
 \end{array}$$

recordando que, en el sistema binario, $0 \times 0 = 0$, $0 \times 1 = 0$, $1 \times 0 = 0$, $1 \times 1 = 1$, $0 + 0 = 0$, $0 + 1 = 1$, $1 + 0 = 1$, $1 + 1 = 10$. El número de multiplicaciones binarias individuales es $(n/2) \times (n/2) = n^2/4$, y puede haber hasta $(n^2/4) - (n/2)$ sumas binarias individuales (incluyendo las que se llevan de una columna a la siguiente). Esto hace $(n^2/2) - (n/2)$ operaciones aritméticas individuales en total —y añadiremos algunas más extra para los pasos lógicos que implica el llevar una cifra en la suma. El número total de pasos es esencialmente $N = n^2/2$ (ignorando los términos de menor orden) que ciertamente es polinómico ¹².

En general, para una clase de problemas tomamos la medida n del «tamaño» del problema como el *número total de dígitos binarios* (o *bits*) necesarios para especificar los datos libres del problema de ese tamaño particular. Esto significa que, para n dado, habrá hasta 2^n diferentes casos del problema para ese tamaño dado (puesto que cada dígito puede ser una de las dos posibilidades, 0 ó 1, y hay n dígitos en total), y estos tienen que ser cubiertos uniformemente por el algoritmo en no más de N pasos.

Existen muchos ejemplos de (clases de) problemas que *no* están en **P**. Por ejemplo, para realizar la operación del cálculo de 2^{2^r} a partir del número r necesitaremos alrededor de 2^n pasos solamente para *escribir la respuesta*, y ya no digamos para realizar el cálculo, siendo n el número de dígitos binarios en la representación binaria r . La operación de calcular 2^{2^r} necesita alrededor de 2^{2^r} pasos ¡solamente para escribirla, etc.! ¡Éstos son mucho mayores que polinomios y, por consiguiente, no pertenecen ciertamente a **P**!

Más interesantes son los problemas en los que pueden escribirse las respuestas, e incluso verificarse su corrección, en un tiempo polinómico. Hay una categoría importante de (algorítmicamente resolubles clases de) problemas que se caracterizan por esta propiedad. Se

les llama (clases de) problemas **NP**. Más exactamente, si un problema individual de una clase de problemas en **NP** tiene una solución, entonces el algoritmo dará esa solución y debe ser posible verificar en un tiempo polinómico que la solución propuesta es realmente una solución. En los casos en que el problema no tiene solución el algoritmo lo dirá, pero no se exige que verifique —en tiempo polinómico o cualquier otro— que realmente no hay solución.

Los problemas **NP** aparecen en muchos contextos, tanto dentro de las propias matemáticas como en el mundo práctico. Sólo daré un sencillo ejemplo matemático: el problema de encontrar lo que se denomina un «circuito hamiltoniano» en un grafo (un nombre más bien intimidatorio para una idea extremadamente simple). Por un «grafo» se entiende una colección finita de puntos, o «vértices», de los cuales un cierto número de pares están conectados mediante líneas —llamadas las «aristas» del grafo. (No estamos interesados aquí en propiedades geométricas o de «distancia», sino sólo en qué vértices están conectados entre sí. Por consiguiente, no importa realmente que los vértices estén representados en un plano —suponiendo que no nos preocupa que las «aristas» se crucen— o en un espacio tridimensional.) Un circuito hamiltoniano es simplemente un camino cerrado (o lazo) que consta solamente de aristas del grafo y que pasa exactamente sólo una vez por cada vértice. Un ejemplo de un grafo, con un circuito hamiltoniano dibujado en él, se representa en la Fig. 4.14. El problema del circuito hamiltoniano consiste en decidir, para cualquier

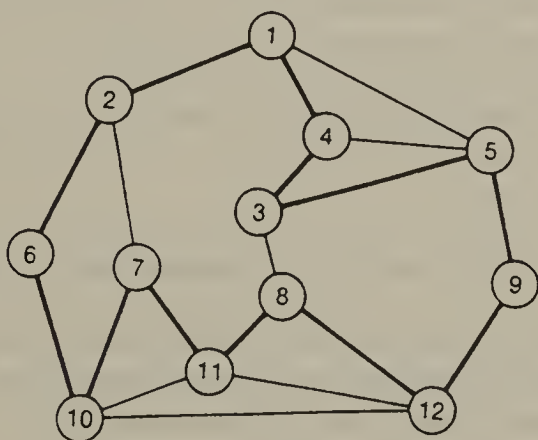


Fig. 4.14. Un grafo con un circuito hamiltoniano indicado (líneas más oscuras). Existe solamente otro circuito hamiltoniano más que el lector puede tratar de localizar.

grafo dado, si existe o no un circuito hamiltoniano y representarlo explícitamente en el caso de que exista. Hay varias maneras de presentar un grafo en términos de dígitos binarios. No importa mucho cuál sea el método utilizado. Un procedimiento sería numerar los vértices 1, 2, 3, 4, 5,... y hacer luego una lista de los pares en un orden fijo adecuado:

(1,2),(1,3),(2,3),(1,4),(2,4),(3,4),(1,5),(2,5),(3,5),(4,5),(1,6),...

A continuación hacemos una correspondiente lista de «0s» y «1s» en la que ponemos un «1» allí donde el par corresponde a una arista y un «0» si no lo hace. Así, la secuencia binaria

10010110110...

significará que el vértice 1 está unido al vértice 2, al vértice 4 y al vértice 5,... el vértice 3 está unido al vértice 4 y al vértice 5,... el vértice 4 está unido al vértice 5,... etc. (como en la Fig. 4.14). El circuito hamiltoniano podría venir dado, si quisiéramos, simplemente como una subcolección de estas aristas que estaría descrita mediante una secuencia binaria con muchos más ceros que antes. El procedimiento de verificación es algo que puede conseguirse mucho más rápidamente que el descubrimiento inicial del circuito hamiltoniano. Todo lo que necesitamos hacer es verificar que el circuito propuesto es realmente un circuito, que sus aristas pertenecen realmente a las del grafo original, y que cada vértice del grafo se utiliza exactamente dos veces, cada una de ellas en extremos de dos aristas diferentes. Este procedimiento de verificación es algo que puede conseguirse fácilmente en un tiempo polinómico.

En realidad este problema no sólo es **NP** sino que se conoce como **NP-completo**. Esto significa que cualquier otro problema **NP** puede ser reducido a éste en un tiempo polinómico —de modo que si alguien fuera lo bastante hábil para encontrar un algoritmo que resuelva el problema del circuito hamiltoniano en tiempo *polinómico*, es decir, probar que el problema del circuito hamiltoniano está realmente en **P**, entonces se concluiría de ello que *¡todos los problemas NP están realmente en P!* Tal circunstancia tendría extraordinarias implicaciones. De un modo general, los problemas que están en **P** se consideran «*tratables*» (es decir, «resolubles en un intervalo de tiempo aceptable»), para n razonablemente grandes, en un rápido computador moderno, mientras que los problemas en **NP** que no están en **P** se consideran «*intratables*» (es decir que, aunque resolubles en principio, son «irresolubles en la práctica») para n razonablemente grande —in-

dependientemente de los incrementos, de cualquier tipo previsible, en velocidad de operación del computador que puedan imaginarse. (El tiempo real que se necesitaría para un problema **NP** no en **P**, para n grande, se haría rápidamente mayor que la edad del universo, ¡lo que no sirve de mucho en un problema práctico!) Cualquier ingenioso algoritmo para resolver el problema del circuito hamiltoniano en un tiempo polinómico ¡podría transformarse en un algoritmo para resolver *cualquier* otro problema **NP** en un tiempo polinómico!

Otro problema que es **NP-completo** es el «problema del viajante», que es muy similar al problema del circuito hamiltoniano excepto que las diversas aristas tienen números asociados a ellas y se busca el circuito hamiltoniano para el que la suma de los números (la «distancia» recorrida por el viajante) es *mínima*. Una vez más, una solución en un tiempo polinómico del problema del viajante conduciría a una solución en tiempo polinómico para *todos* los demás problemas **NP**. (¡Si se encontrara tal solución ello constituiría una noticia de portada! En efecto, existen en particular sistemas secretos de codificación, introducidos en los últimos años, que dependen de un problema de factorización de números enteros altos, siendo éste otro problema **NP**. Si este problema pudiera resolverse en tiempo polinómico, entonces tales códigos podrían ser probablemente descifrados utilizando los potentes ordenadores modernos; pero si no es así, los códigos parecen estar a salvo. Ver Gardner 1989.)

Es opinión común de los expertos la de que realmente es *imposible* resolver, con cualquier dispositivo similar a una máquina de Turing, un problema **NP-completo** en tiempo polinómico, y que, por consiguiente, **P** y **NP** *no* son el mismo. Es muy probable que esta creencia sea correcta, pero todavía nadie ha sido capaz de demostrarlo. Éste sigue siendo el más importante problema no resuelto de la teoría de la complejidad.

Complejidad y computabilidad en los objetos físicos

La teoría de la complejidad es importante, para nuestras consideraciones en este libro, debido a que plantea otro tema algo diferente de la cuestión de si las cosas son o no algorítmicas: a saber, si las cosas que se sabe que son algorítmicas, son o no realmente algorítmicas de un modo *útil*. En los últimos capítulos tendré menos que decir acerca de cuestiones de la teoría de la complejidad que de computabilidad. De hecho, me inclino a pensar (aunque, sin duda, sobre

bases bastante insuficientes) que, a diferencia de la misma cuestión básica de la computabilidad, los resultados de la teoría de la complejidad no son cruciales en relación con los fenómenos mentales. Más aún, tengo la sensación de que las cuestiones de la factibilidad de los algoritmos apenas se rozan en la teoría de la complejidad tal como hoy existe.

Sin embargo, bien pudiera estar equivocado sobre el papel de la complejidad. Como señalaré más adelante (en el Capítulo 9, p. 498), la teoría de la complejidad para *objetos físicos reales* podría diferir quizá en aspectos significativos de lo que acabamos de discutir. Para que esta posible diferencia se haga manifiesta será necesario aprovechar algunas de las propiedades mágicas de la mecánica cuántica —una misteriosa pero poderosamente precisa teoría del comportamiento de átomos y moléculas, y de muchos otros fenómenos, algunos de los cuales son importantes a una escala mucho mayor. Aprenderemos algo de esta teoría en el Capítulo 6. Según un reciente conjunto de ideas introducidas por David Deutsch (1985), es posible *en principio* construir un «computador cuántico» para el que existen (clases de) problemas que no están en **P**, pero que podrían ser resueltos por dicho dispositivo en tiempo polinómico. No es en absoluto evidente, por ahora, cómo podría construirse un dispositivo físico real que se comporte (fiablemente) como un computador cuántico —y, además, la clase particular de problemas considerada hasta ahora es decididamente artificial— pero parece existir la posibilidad *teórica* de que un dispositivo físico cuántico pueda mejorar una máquina de Turing.

¿Sería posible que un cerebro humano, que para nuestra discusión estoy considerando como un «dispositivo físico» aunque de una sorprendente sutileza y delicado diseño, tanto como complicación, estuviera sacando provecho de la magia de la teoría cuántica? ¿Comprendemos el modo en que podrían utilizarse beneficiosamente los efectos cuánticos para la resolución de problemas o la formación de juicios? ¿Es concebible que tengamos que ir aún «más allá» de la teoría cuántica actual para hacer uso de estas posibles ventajas? ¿Es ciertamente probable que los dispositivos físicos reales puedan mejorar la teoría de la complejidad para máquinas de Turing? ¿Y qué pasa con la teoría de la *computabilidad* para dispositivos físicos reales?

Para abordar estas cuestiones debemos apartarnos de los temas puramente matemáticos y preguntar, en los próximos capítulos, ¡cómo se comporta realmente el mundo físico!

5. El mundo clásico

El status de la teoría física

¿Qué necesitamos conocer acerca del funcionamiento de la Naturaleza para poder apreciar de qué modo puede formar parte de ella la consciencia? ¿Importa realmente cuáles sean las leyes que gobiernan los elementos constituyentes de los cuerpos y los cerebros? Si nuestras percepciones conscientes consisten simplemente en la activación de algoritmos, como muchos defensores de la IA quieren hacernos creer, entonces no sería de mucha relevancia cuáles sean realmente estas leyes. Cualquier dispositivo que *fuera* capaz de ejecutar un algoritmo sería tan bueno como cualquier otro. Quizá, por el contrario, haya más que simples algoritmos en nuestras sensaciones de consciencia. Tal vez sea realmente relevante el modo detallado en que estamos constituidos, como lo sean las leyes físicas que realmente gobiernan la substancia de la que estamos compuestos. Quizá necesitamos comprender qué cualidad profunda subyace en la naturaleza misma de la materia y determina el modo en que esta materia debe comportarse. La física no ha alcanzado aún este punto. Todavía quedan muchos misterios que desentrañar y muchas intuiciones directas que obtener. A pesar de ello, muchos físicos y fisiólogos dirían que *ya* sabemos bastante sobre las leyes físicas relevantes para el funcionamiento de un objeto de tamaño-normal como es un cerebro humano. Aunque es indudable que un cerebro humano es excepcionalmente complicado como sistema físico, y que aún no conocemos mucho de su estructura detallada y actuación relevante, pocos estarían dispuestos a afirmar que es precisamente en los principios *físicos* que

subyacen a su comportamiento donde existe una significativa falta de comprensión.

Defenderé más adelante el planteamiento poco convencional de que, contrariamente a esta opinión, todavía *no* comprendemos suficientemente bien la física como para que el funcionamiento de nuestros cerebros pueda describirse adecuadamente en sus términos, ni siquiera en principio. Para hacer este planteamiento, será necesario que presente en primer lugar una visión general del status de la teoría física actual. Este capítulo concierne a la llamada «física clásica», que incluye tanto la mecánica de Newton como la relatividad de Einstein. Aquí, por «clásico» entendemos esencialmente las teorías dominantes antes de la llegada, alrededor de 1925 (a través del trabajo inspirado de físicos como Planck, Einstein, Bohr, Heisenberg, Schrödinger, de Broglie, Born, Jordan, Pauli y Dirac), de la *teoría cuántica*, una teoría con incertidumbre, indeterminismo y misterio, que describe el comportamiento de moléculas, átomos y partículas subatómicas. La teoría clásica es, por el contrario, *determinista* en el sentido de que el futuro está siempre completamente fijado por el pasado. Aun así, la física clásica tiene mucho de misterioso, pese al hecho de que el conocimiento que se ha conseguido a lo largo de siglos nos ha llevado a una imagen de gran precisión fenoménica. Tendremos que examinar también la teoría cuántica (en el Capítulo 6) pues creo que, contrariamente a lo que parece ser la opinión mayoritaria entre los fisiólogos, es probable que los fenómenos cuánticos sean importantes en el modo de operar del cerebro —pero eso es un tema para los siguientes capítulos.

Lo que la ciencia ha conseguido hasta ahora ha sido espectacular. Sólo tenemos que mirar a nuestro alrededor para ser testigos de lo que el extraordinario poder de nuestra comprensión de la Naturaleza nos ha ayudado a obtener. La tecnología del mundo moderno se ha derivado, en buena medida, de una gran riqueza de experiencia empírica. Sin embargo, es la *teoría* física la que subyace a nuestra tecnología de un modo mucho más fundamental, y es la teoría física la que nos interesará aquí. Las teorías de que disponemos ahora tienen una precisión bastante notable. Pero no es sólo su precisión la que les da su fuerza. Es también el hecho de que se han mostrado extraordinariamente susceptibles de un tratamiento matemático preciso y detallado. Son estos hechos en conjunto los que nos han dado una ciencia de una potencia verdaderamente impresionante.

Buena parte de esta teoría física no es especialmente reciente. Si puede singularizarse un suceso sobre todos los demás, éste es la pu-

blicación, en 1687, de los *Principia* de Isaac Newton. Esta obra extraordinaria demostró cómo, a partir de unos pocos principios físicos, se puede comprender, y a menudo predecir con sorprendente precisión, gran parte del comportamiento real de los objetos físicos. (Una buena parte de los *Principia* tenía que ver también con notables desarrollos en las técnicas matemáticas, aunque más tarde Euler y otros proporcionaron métodos más prácticos.) El propio trabajo de Newton, como él fue el primero en admitir, debía mucho a los logros de pensadores anteriores, entre los que eran preeminentes los nombres de Galileo Galilei, René Descartes y Johannes Kepler. Pero había importantes conceptos subyacentes que procedían de pensadores todavía más antiguos, como las ideas geométricas de Platón, Eudoxo, Euclides, Arquímedes y Apolonio. Más adelante tendré más que decir sobre ello.

Las desviaciones del esquema básico de la dinámica de Newton vendrían más tarde. La primera fue la teoría electromagnética de James Clerk Maxwell, desarrollada a mediados del siglo XIX. Esta englobaba no sólo el comportamiento clásico de los campos eléctricos y magnéticos sino también el de la luz ¹. Esta importante teoría será objeto de nuestra atención más adelante en este mismo capítulo. La teoría de Maxwell es de considerable importancia para la tecnología actual, y no hay duda de que los fenómenos electromagnéticos son relevantes para el funcionamiento de nuestros cerebros. Lo que es menos evidente, sin embargo, es que también pueden ser importantes para nuestros procesos mentales las dos grandes teorías de la relatividad asociadas al nombre de Albert Einstein. La teoría *especial* de la relatividad, que se desarrolló a partir de un estudio de las ecuaciones de Maxwell, fue desarrollada por Henri Poincaré, Hendrick Antoon Lorentz y Einstein (y más tarde le fue dada una elegante descripción geométrica por Hermann Minkowski) para explicar el enigmático comportamiento de los cuerpos cuando se mueven a velocidades próximas a la de la luz. La famosa ecuación de Einstein « $E = mc^2$ » era parte de esta teoría. Pero el impacto de la teoría en la tecnología ha sido muy pequeño (excepto allí donde incide sobre la física nuclear) y su relevancia para el funcionamiento de nuestros cerebros parecería ser, en el mejor de los casos, marginal. No obstante, la relatividad especial nos dice algo profundo sobre la realidad física, en relación con la naturaleza del *tiempo*. Veremos en los próximos capítulos que esto conduce a algunos profundos enigmas concernientes a la teoría cuántica que podrían tener importancia en relación con nuestra percepción del «flujo del tiempo». Además, necesitaremos comprender

la teoría especial para poder estimar debidamente la teoría de la relatividad *general* de Einstein —la teoría que utiliza el espacio-tiempo curvo para describir la gravedad. El impacto de *esta* teoría sobre la tecnología casi no ha existido hasta ahora *, y ¡parecería fantástico en extremo sugerir que pueda tener alguna relevancia para el funcionamiento de nuestros cerebros! Pero, curiosamente, es en realidad la teoría *general* la que tendrá la mayor relevancia para nuestras discusiones posteriores, más particularmente en los Capítulos 7 y 8, en los que tendremos que aventurarnos en los más remotos confines del espacio y el tiempo para hacernos una idea de los cambios que considero que son necesarios antes de que pueda salir a la luz una imagen debidamente coherente de la teoría cuántica (¡pero dejémoslo para más adelante!)

Éstas son las áreas generales de la física *clásica*. ¿Qué pasa con la física cuántica? A diferencia de la teoría de la relatividad, la teoría cuántica *está* empezando a tener un impacto realmente significativo sobre la tecnología. Esto se debe en parte a los conocimientos que ha proporcionado en ciertas áreas tecnológicamente importantes tales como la química y la metalurgia. Algunos dirán, de hecho, que estas áreas han quedado incluidas dentro de la física en virtud de las nuevas intuiciones precisas que hemos tenido gracias a la teoría cuántica. Además de esto, la teoría cuántica nos ha presentado muchos fenómenos *nuevos*, el más familiar de los cuales es, supongo, el láser. ¿No podría ser que algunos aspectos esenciales de la teoría cuántica estuvieran desempeñando también un papel crucial en la física que subyace a nuestros procesos mentales?

¿Qué hay de los conocimientos físicos de origen más reciente? Algunos lectores pueden haberse encontrado con ideas expresadas con entusiasmo, que incluyen nombres como «quarks» (cf. p. 201), «GUT» (Teorías de Gran Unificación), el «escenario inflacionario» (ver nota final 12 en p. 575), «supersimetría», «teoría de (super)cuerdas», etc. ¿Cómo se comparan estos nuevos esquemas con los que acabo de mencionar? ¿Necesitaremos saber también algo sobre ellos? Para colocar las cosas en una perspectiva más apropiada, formularé

* Casi, pero no del todo; la precisión exigida para el comportamiento de las sondas espaciales requiere realmente que sus órbitas se calculen tomando en consideración los efectos de la relatividad general; y existen dispositivos capaces de localizar una posición en la Tierra con tal exactitud (de hecho, con un error de apenas unos decímetros) que deben utilizar realmente los efectos de la curvatura del espacio-tiempo de la relatividad general.

tres amplias categorías de teorías físicas básicas. Denomino a éstas como sigue:

1. SOBERBIAS.
2. ÚTILES.
3. TENTATIVAS.

En la categoría de SOBERBIAS deben ir todas las que he estado discutiendo en los párrafos que preceden a éste. Para calificarlas de SOBERBIAS no estimo necesario que la teoría se aplique sin refutación a los fenómenos del mundo, sino que sólo exijo que el alcance y exactitud con que se apliquen sea *excepcional*, en un sentido apropiado. ¡Tal como estoy utilizando el término «soberbia», resulta extraordinariamente notable el simple hecho de que existan teorías en esta categoría! No conozco ninguna teoría básica de ninguna otra ciencia que pudiera entrar debidamente en esta categoría. Quizá la teoría de la selección natural, que propusieron Darwin y Wallace, sea la que está más próxima aunque todavía a una buena distancia.

La más antigua de las teorías SOBERBIAS es la geometría euclídea de la que algo aprendemos en la escuela. Los antiguos pudieron no considerarla como una teoría física en absoluto, pero eso es lo que realmente es: una sublime y soberbiamente precisa teoría del espacio físico (y de la geometría de los cuerpos rígidos). ¿Por qué considero la geometría euclídea una teoría física en lugar de una rama de las matemáticas? Irónicamente, una de las razones más evidentes para adoptar esta opinión es que ahora sabemos que la geometría euclídea *¡no es completamente exacta* como descripción del espacio físico en que habitamos! La teoría de la relatividad general de Einstein nos dice ahora que el espacio(-tiempo) es realmente «curvo» (es decir, *no* exactamente euclídeo) en presencia de un campo gravitatorio. Pero este hecho no descalifica a la caracterización de la geometría euclídea como SOBERBIA. Para dimensiones en la escala del metro las desviaciones respecto a la planitud euclídea son ínfimas, y los errores al tratar la geometría como euclídea son ¡menores que el diámetro de un átomo de hidrógeno!

Es razonable decir que también debería ser calificada de SOBERBIA la teoría de la *estática* (que trata de cuerpos que no están en movimiento), tal como la desarrollaron Arquímedes, Pappus y Stevin en forma de una hermosa ciencia. Esta teoría está ahora incluida en la mecánica newtoniana. Las ideas profundas de la *dinámica* (cuerpos en movimiento) introducidas por Galileo alrededor del 1600, y desarrolladas por Newton hasta constituir una magnífica y amplia teoría, deben entrar indudablemente en la categoría de SOBERBIAS. Cuan-

do se aplica a los movimientos de planetas y satélites la precisión observada de la teoría es excepcional: mejor que una parte en diez millones. El mismo esquema newtoniano se aplica aquí en la Tierra —o entre las estrellas y galaxias— con una precisión comparable. Análogamente, la teoría de Maxwell es válida con gran exactitud sobre un extraordinario dominio, que en el extremo inferior se extiende hasta la escala minúscula de átomos y partículas subatómicas, y en el superior, hasta la escala de las galaxias, ¡aproximadamente un millón de millones de millones de millones de millones de veces mayor! (En el extremo más pequeño de esta escala las ecuaciones de Maxwell deben combinarse adecuadamente con las reglas de la mecánica cuántica.) Ciertamente debe ser calificada también de SOBERBIA.

La relatividad especial de Einstein (anticipada por Poincaré y elegantemente reformulada por Minkowski) da una descripción maravillosamente precisa de los fenómenos en los que la velocidad de los objetos llega a ser próxima a la de la luz —velocidades a las que las descripciones de Newton comienzan a fallar. La teoría de la relatividad general de Einstein, de suprema belleza y originalidad, generaliza la teoría dinámica de (la gravitación de) Newton y mejora su exactitud, a la vez que hereda toda la notable precisión de esta teoría respecto al movimiento de planetas y satélites. Además, explica con detalle diversos hechos observacionales que son incompatibles con el esquema newtoniano. Uno de estos hechos (el «pulsar binario», cf. p. 273) muestra que la teoría de Einstein tiene una exactitud de alrededor de una parte en 10^{14} . Ambas teorías de la relatividad —la segunda de las cuales incluye a la primera— deben clasificarse realmente como SOBERBIAS (casi tanto en razón de su elegancia matemática como por su exactitud).

El rango de fenómenos que se explican de acuerdo con la extraordinariamente bella y revolucionaria teoría de la mecánica cuántica, y la exactitud con que concuerda con los experimentos, nos dice claramente que también la teoría cuántica debe ser calificada de SOBERBIA. No se conocen discrepancias observacionales con dicha teoría —aunque su fuerza reside, más allá de esto, en el número de fenómenos antes inexplicables que la teoría explica ahora. Las leyes de la química, la estabilidad de los átomos, la agudeza de las líneas espectrales (cf. p. 290) y sus muy específicas estructuras observadas, el curioso fenómeno de la superconductividad (resistencia eléctrica nula) y el comportamiento de los láseres son sólo algunos de éstos.

Estoy estableciendo altos niveles para la categoría de SOBER-

BIAS, pero nos hemos acostumbrado a esto en la física. Ahora bien, ¿qué hay sobre las teorías más recientes? En mi opinión sólo hay una de ellas que pueda posiblemente calificarse como SOBERBIA y no es particularmente reciente: una teoría llamada *electrodinámica cuántica* (o QED), que surgió del trabajo de Jordan, Heisenberg y Pauli, fue formulada por Dirac en 1926-1934, y hecha manejable por Bethe, Feynman, Schwinger y Tomonaga en 1947-1948. Esta teoría apareció como una combinación de los principios de la mecánica cuántica con los de la ecuación fundamental, debida a Dirac, que gobierna el movimiento y spin del electrón. La teoría en conjunto no tiene la irresistible elegancia o consistencia de las teorías SOBERBIAS anteriores, pero se califica así en virtud de su precisión verdaderamente excepcional. Un resultado particularmente digno de mención es el valor del momento magnético del electrón. (Los electrones se comportan como minúsculos imanes de carga eléctrica en rotación. El término «momento magnético» se refiere a la fuerza de este minúsculo imán.) El valor calculado para este momento magnético a partir de la QED es 1.001 15965246 (en las unidades apropiadas, con un margen de error de alrededor de 20 en las dos últimas cifras), mientras que el valor experimental más reciente es 1.001 159652193 (con un posible error de alrededor de 10 en las dos últimas cifras). Como ha señalado Feynman, esta precisión equivale a determinar la distancia entre Nueva York y Los Angeles ¡con un error menor que el espesor de un cabello humano! No tendremos aquí necesidad de conocer esta teoría pero, por completitud, mencionaré brevemente algunas de sus características fundamentales hacia el final del próximo capítulo *.

Existen algunas teorías actuales que yo colocaría en la categoría de ÚTILES. Dos de éstas no serán necesarias aquí pero son dignas de mención. La primera es la del modelo de *quarks* de Gell-Mann-Zweig para las partículas subatómicas llamadas *hadrones* (los protones, neutrones, mesones, etc. que constituyen los núcleos atómicos —o, más correctamente, las partículas «fuertemente interactivas») y la (posterior) teoría detallada de sus interacciones, conocida como *cromodinámica cuántica*, o QCD. La idea consiste en que todos los hadrones están formados por constituyentes conocidos como «quarks» que interaccionan entre sí mediante una cierta generalización de la teoría de Maxwell (llamada teoría de «Yang-Mills»). En segundo lugar, existe una teoría (debida a Glashow, Salam, Ward y Weinberg, y que utiliza una vez más la teoría de Yang-Mills) que combina las

* Ver el libro *QED* de Feynman (1985) para una exposición popular de esta teoría.

fuerzas electromagnéticas con las interacciones «débiles» que son las responsables de la desintegración radiactiva. Esta teoría incorpora una descripción de los llamados *leptones* (electrones, muones, neutrinos; también de las partículas W y Z —las partículas «débilmente interactivas»). Hay un buen soporte experimental para ambas teorías. Sin embargo, ellas son, por varias razones, algo más desordenadas de lo que uno quisiera (como sucedía con la QED, pero más en este caso) y su exactitud observada y poder predictivo quedan, por el momento, a mucha distancia del nivel «excepcional» que se exige para su inclusión en la categoría de SOBERBIAS. Estas dos teorías juntas (la segunda incluyendo la QED) se conocen a veces como el *modelo estándar*.

Finalmente, existe una teoría de otro tipo que creo que también pertenece, cuando menos, a la categoría de ÚTILES. Ésta es la llamada teoría del *big bang* origen del Universo *. Esta teoría jugará un papel importante en las discusiones de los Capítulos 7 y 8.

No creo que ninguna otra teoría pueda entrar en la categoría de ÚTILES. Existen muchas ideas populares actualmente (o recientemente). Algunas de ellas son: las teorías de «Kaluza-Klein», «supersimetría» (o «supergravedad»), y las ahora muy de moda teorías de «cuerdas» (o «supercuerdas»), además de las teorías «GUT» (y ciertas ideas derivadas de ellas, como el «escenario inflacionario», cf. nota 12 en p. 575). Todas ellas entran firmemente, en mi opinión, en la categoría de TENTATIVAS. (Ver Barrow 1988, Close 1983, Davies y Brown 1988, Squires 1985.) La diferencia importante entre las categorías de ÚTILES y TENTATIVAS es la falta de cualquier soporte experimental significativo para las teorías en esta última categoría ³. Esto no quiere decir que quizá alguna de ellas no pudiera ascender espectacularmente a la categoría de las ÚTILES o incluso a la de SOBERBIAS. Algunas de estas teorías contienen realmente ideas originales muy prometedoras, pero por ahora siguen siendo ideas sin soporte experimental. La categoría de las TENTATIVAS es una categoría muy amplia. Las ideas implícitas en algunas de ellas podrían contener las semillas de un nuevo avance substancial en el conocimiento, mientras que algunas otras me dan la impresión de ser artificiosas o estar decididamente descaminadas. (Estuve tentado de se-

* Me refiero aquí a lo que se conoce como el «modelo estándar» del big bang. Existen muchas variantes de la teoría del big bang, de las que las más populares actualmente proporcionan lo que se conoce como «escenario inflacionario» —en mi opinión, claramente en la categoría de TENTATIVA.

parar una cuarta categoría de la respetable categoría de TENTATIVAS, y llamarla, por ejemplo, DESCAMINADAS; ¡pero luego lo pensé mejor ya que no quiero perder a la mitad de mis amigos!)

No debería sorprendernos que las principales teorías SOBERBIAS sean antiguas. A lo largo de la historia debe haber habido muchas más teorías que entrarían en la categoría de TENTATIVAS, pero la mayoría de ellas han sido olvidadas. Análogamente, debe haber habido otras muchas en la categoría de ÚTILES que se han desvanecido desde entonces; pero había también algunas que han sido incorporadas en teorías que más tarde llegaron a ser SOBERBIAS por sí mismas. Consideremos unos pocos ejemplos. Antes de que Copérnico, Kepler y Newton concibieran un esquema mucho mejor, existía una teoría maravillosamente elaborada del movimiento planetario que habían desarrollado los antiguos griegos, conocida como *sistema ptolemaico*. Según este esquema los movimientos de los planetas están gobernados por complicadas composiciones de movimientos circulares. Fue bastante eficaz para hacer predicciones, pero se hizo más y más complicado a medida que se necesitaba una mayor exactitud. Hoy día el sistema ptolemaico nos parece muy artificioso. Éste es un buen ejemplo de una teoría ÚTIL (¡durante unos veinte siglos, de hecho!) que posteriormente se *disolvió* como teoría física, aunque jugó un papel organizativo de clara importancia histórica. Como un buen ejemplo de una teoría ÚTIL del tipo finalmente *acertado* podemos considerar, en su lugar, la brillante concepción de Kepler del movimiento planetario elíptico. Otro ejemplo fue la tabla periódica de Mendeleev para los elementos químicos. Por sí mismas ellas no proporcionaban esquemas predictivos con el carácter «excepcional» exigido, pero posteriormente llegaron a ser deducciones «correctas» dentro de teorías SOBERBIAS que crecieron a partir de ellas (la dinámica newtoniana y la teoría cuántica, respectivamente).

En las secciones y capítulos siguientes no tendré mucho que decir sobre las teorías actuales que son simplemente ÚTILES o TENTATIVAS. Hay bastante que decir sobre las que son SOBERBIAS. Es realmente una suerte que tengamos tales teorías y podamos comprender el mundo en que vivimos de una forma tan notablemente completa. Con el tiempo debemos tratar de decidir si incluso estas teorías son suficientemente ricas para gobernar las acciones de nuestros cerebros y mentes. Traeré a colación esta cuestión a su debido tiempo; pero por ahora consideraremos las teorías SOBERBIAS tal como las conocemos y trataremos de sopesar su relevancia para nuestros propósitos.

Geometría euclídea

La geometría euclídea es simplemente esa materia que aprendemos como «geometría» en la escuela. Sin embargo, supongo que la mayoría de la gente la considera parte de las matemáticas, antes que como una teoría física. Por supuesto es también parte de las matemáticas, pero la geometría euclídea no es ni mucho menos la única geometría matemática concebible. La geometría particular que nos ha sido transmitida por Euclides describe con gran exactitud el espacio físico del mundo en que vivimos, pero *no* es una necesidad lógica; es sólo una característica (aproximadamente exacta) *observada* del mundo físico.

De hecho, hay otra geometría, llamada *lobachevskiana* * (o *hiperbólica*) que es muy similar a la geometría euclídea en muchos aspectos pero con algunas curiosas diferencias. Por ejemplo, recordamos que en la geometría euclídea la suma de los ángulos de cualquier triángulo es siempre 180° . En la geometría lobachevskiana esta suma es siempre *menor* de 180° , siendo la diferencia proporcional al área del triángulo (ver Fig. 5.1)

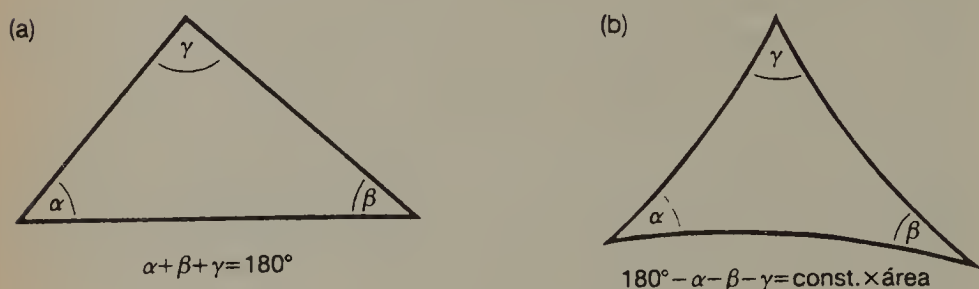


Fig. 5.1. (a) Un triángulo en un espacio euclídeo, (b) un triángulo en un espacio lobachevskiano.

El famoso artista holandés Maurits C. Escher ha concebido algunas representaciones muy hermosas y exactas de esta geometría. En la Figura 5.2 se reproduce uno de sus grabados. Cada pez negro debe ser imaginado, según la geometría lobachevskiana, del mismo tamaño

* Nicolai Ivanovich Lobachevsky (1792-1856) fue uno de los varios que, independientemente, descubrieron este tipo de geometría como alternativa a la de Euclides. Otros fueron Carl Friedrich Gauss (1777-1855), Ferdinand Schweickard y Janos Bolyai.



Fig. 5.2. Representación de Escher del espacio de Lobachevski. (Todos los peces negros deben considerarse congruentes, y lo mismo debe hacerse para todos los peces blancos.)

forma que cualquiera de los otros peces negros, y lo mismo es válido para los peces blancos. La geometría no puede representarse de forma completamente exacta en el plano euclídeo ordinario; de ahí el aparente apiñamiento en las proximidades del contorno circular. Imagínese usted mismo situado en el interior de la figura pero en algún lugar próximo a este contorno; se supone entonces que la geometría lobachevskiana parece la misma que si usted estuviera en el centro, o en cualquier otro lugar. Lo que parece ser el «contorno» de la estructura, según esta representación euclídea, está realmente «en el infinito» en la geometría lobachevskiana. El contorno circular real no debe considerarse en absoluto como parte del espacio lobachevskiano

—y tampoco debería ser considerada la región euclídea en el *exterior* de este círculo. (Esta ingeniosa representación del plano de Lobachevski se debe a Poincaré. Tiene la virtud especial de que las formas muy pequeñas no quedan distorsionadas en la representación; sólo los tamaños cambian.) Las «líneas rectas» de la geometría (a lo largo de algunas de las cuales apuntan los peces de Escher) son círculos que intersectan en ángulos rectos a este contorno circular.

Podría muy bien suceder que la geometría lobahevschiana fuera realmente verdadera en nuestro mundo a escala cosmológica (ver Capítulo 7, p. 297). Sin embargo, la constante de proporcionalidad entre el déficit de ángulo para un triángulo y su área tendría que ser *extraordinariamente* pequeña en este caso, y la geometría euclídea sería una excelente aproximación a esta geometría para cualquier escala ordinaria. De hecho, como veremos más adelante en este mismo capítulo, la teoría de la relatividad general de Einstein nos dice que la geometría de nuestro mundo *difiere* de la geometría euclídea (aunque de un modo «irregular» que es más complicado que la geometría lobahevschiana) en escalas considerablemente menos remotas que las cosmológicas, aunque las desviaciones son todavía extraordinariamente pequeñas en las escalas ordinarias de nuestra experiencia directa.

El hecho de que la geometría euclídea parezca tan precisa para reflejar la estructura del «espacio» de nuestro mundo nos ha engañado (o a nuestros predecesores) haciéndonos pensar que esta geometría es una necesidad lógica, o haciéndonos pensar que tenemos una comprensión intuitiva innata *a priori* de que la geometría euclídea *debe* aplicarse al mundo en que vivimos. (Incluso el gran filósofo Immanuel Kant afirmaba esto.) La ruptura real con la geometría euclídea sólo llegó con la teoría de la relatividad general de Einstein, que fue propuesta muchos años después. El que la geometría euclídea se aplique de forma tan precisa —aunque no suficientemente exacta— a la estructura de nuestro espacio físico, lejos de ser una necesidad lógica ¡es un *hecho de observación empírica*! La geometría euclídea fue realmente, desde el principio, una (SOBERBIA) teoría *física*. Lo era así además de ser un elegante y lógico elemento de la matemática pura.

En cierto sentido, esto no estaba tan alejado del punto de vista adoptado por Platón (c. 360 a.C.; esto es, unos cincuenta años antes de los *Elementos* de Euclides, el famoso libro sobre geometría). En opinión de Platón, los objetos de la geometría pura —líneas rectas, círculos, triángulos, planos, etc.— sólo se realizaban aproximadamente en el mundo de las cosas físicas reales. Los objetos matemáticamente precisos de la geometría pura no poblaban este mundo físico

sino un mundo diferente: *el mundo ideal de Platón* de los conceptos matemáticos. El mundo de Platón consta no sólo de objetos tangibles sino de «objetos matemáticos». Este mundo no nos es accesible del modo físico ordinario sino por la vía del *intelecto*. Nuestra mente entra en contacto con el mundo de Platón cada vez que contempla una verdad matemática, percibiéndola mediante el ejercicio de razonamiento e intuición matemática. El mundo ideal se consideraba diferente y más perfecto que el mundo material de nuestra experiencia externa, pero tan real como éste. (Recordemos nuestras discusiones de los Capítulos 3 y 4, pp. 134, 153 sobre la realidad platónica de los conceptos matemáticos). Así, mientras que los objetos de la geometría euclídea pura pueden ser estudiados por el pensamiento, y pueden derivarse de este modo muchas propiedades de este ideal, no hay necesidad de que el «imperfecto» mundo físico de la experiencia externa se adhiera exactamente a este ideal. Por alguna milagrosa intuición, y sobre la base de lo que debió haber sido una evidencia muy dispersa en ese tiempo, Platón parece haber anticipado esto: por una parte, las matemáticas deben ser estudiadas y comprendidas por sí mismas, y no debemos pedir una aplicabilidad completamente exacta a los objetos de la experiencia física; por otra parte, el funcionamiento del mundo externo real puede ser entendido finalmente sólo en términos de matemáticas exactas, lo que, en términos del mundo ideal de Platón, quiere decir «accesible por la vía del intelecto».

Platón fundó la Academia en Atenas destinada a fomentar tales ideas. Entre la élite que surgió de sus miembros estaba el extraordinariamente influyente y famoso filósofo Aristóteles. Pero aquí nos interesaremos en otro miembro de esta Academia (algo menos conocido que Aristóteles pero, en mi opinión, mucho mejor científico), uno de los más grandes pensadores de la antigüedad: el matemático y astrónomo Eudoxo.

Existe un ingrediente profundo y sutil en la geometría euclídea —en realidad el más esencial— ¡que hoy en día apenas consideramos en absoluto como geometría! (Los matemáticos tenderán a llamar a este elemento «análisis» más que «geometría»). Éste era la introducción efectiva de los *números reales*. La geometría euclídea trabaja con longitudes y ángulos. Para comprender esta geometría debemos estimar qué tipo de «números» son necesarios para describir estas longitudes y ángulos. La nueva idea central fue expuesta en el siglo IV a.C. por Eudoxo (c. 408-355 a.C.) * La geometría griega había pasado por

* Eudoxo fue también el originador de la, durante 2.000 años, ÚTIL teoría del

una «crisis» debido al descubrimiento por los pitagóricos de que números como $\sqrt{2}$ (necesarios para expresar la relación entre la longitud de la diagonal de un cuadrado y su lado) no puede expresarse como una fracción (cf. Capítulo 3, p. 114). Había sido importante para los griegos el poder formular sus medidas (razones) geométricas en términos de (razones de) enteros para que las magnitudes geométricas pudieran ser estudiadas de acuerdo con las leyes de la aritmética. Básicamente, la idea de Eudoxo fue proporcionar un método de describir razones de longitudes (esto es, ¡números reales!) en términos de *enteros*. Él fue capaz de dar criterios, establecidos en términos de operaciones enteras, para decidir cuándo una razón es mayor que otra, o si las dos deben considerarse exactamente iguales.

La idea era aproximadamente la siguiente: Si a , b , c y d son cuatro longitudes, entonces un criterio para verificar que la razón a/b es *mayor* que la razón c/d es que existan números enteros M y N tales que a sumada N veces consigo misma sea mayor que b sumada M veces consigo misma, al tiempo que d sumada M veces consigo misma sea mayor que c sumada N veces consigo misma *. Un criterio análogo puede utilizarse para verificar que a/b es *menor* que c/d . ¡El criterio buscado para la *igualdad* $a/b = c/d$ es entonces simplemente que ninguno de esos otros dos criterios pueda satisfacerse!

Una teoría matemática abstracta completamente exacta de los números reales no fue desarrollada hasta el siglo XIX, por matemáticos como Dedekind y Weierstrass. Pero sus métodos seguían realmente líneas muy similares a las que Eudoxo ya había descubierto ¡unos veintidós siglos antes! No hay necesidad aquí de describir este desarrollo moderno. Esta teoría moderna fue vagamente insinuada en p. 5 del Capítulo 3 pero entonces preferí, por facilidad de presentación, basar la discusión de los números reales en las más conocidas expansiones decimales. (Estas expansiones fueron introducidas por Stevin en 1585.) Debe tenerse en cuenta que las expansiones decimales, aunque familiares para nosotros, eran desconocidas para los griegos.

Hay una diferencia importante, sin embargo, entre la propuesta de Eudoxo y las de Dedekind y Weierstrass. Los antiguos griegos

movimiento planetario, posteriormente desarrollada con más detalle por Hiparco y Ptolomeo, y conocida en consecuencia como sistema ptolemaico.

* En notación moderna esto afirma la existencia de una fracción, a saber, M/N , tal que $a/b > M/N > c/d$. Siempre existirá tal fracción que yace entre los dos números a/b y c/d con tal de que $a/b > c/d$, de modo que el criterio de Eudoxo se satisface efectivamente.

pensaban en los números reales como cosas *dadas* —en términos de (razones de) magnitudes geométricas— es decir, como propiedades del espacio «real». Era necesario para los griegos el poder describir las magnitudes geométricas en términos aritméticos para poder razonar rigurosamente sobre ellas, y también sobre sus sumas y productos —ingredientes esenciales de muchos de los maravillosos teoremas geométricos de los antiguos. (En la Fig. 5.3 he dado, a modo de ilustración, el famoso *teorema de Ptolomeo* —aunque Ptolomeo lo descubrió en una época muy posterior a Eudoxo— que relaciona las distancias entre cuatro puntos en una circunferencia, que ilustra muy bien cómo son necesarias ambas sumas y productos.) El criterio de Eudoxo se mostró extraordinariamente fructífero y, en particular, capacitó a los griegos para calcular áreas y volúmenes de un modo riguroso.

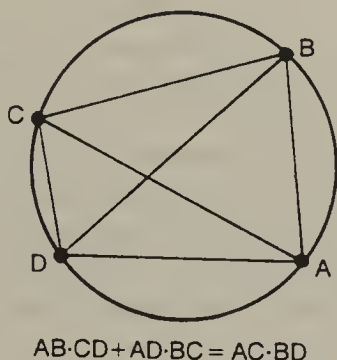


Fig. 5.3. El teorema de Ptolomeo.

Sin embargo, para los matemáticos del siglo XIX —y, ciertamente, para los matemáticos de hoy— el papel de la geometría ha cambiado. Para los antiguos griegos, y para Eudoxo en particular, los números «reales» eran cosas que había que *extraer* de la geometría del espacio físico. Ahora preferimos pensar en los números reales como lógicamente más primitivos que la geometría. Esto nos permite construir toda clase de tipos *diferentes* de geometría, *partiendo* para cada uno del concepto de número. (La idea clave fue la de geometría de coordenadas, introducida en el siglo XVII por Fermat y Descartes. Las coordenadas pueden utilizarse para *definir* otros tipos de geometría.) Cualquiera de estas «geometrías» debe ser lógicamente consistente, pero no es necesario que tenga relevancia directa para el espacio físico de nuestra experiencia. La geometría física particular que *nos parece*

percibir es una *idealización* de la experiencia (v.g. dependiente de nuestras extrapolaciones a tamaños indefinidamente grandes o pequeños, cf. Capítulo 3, p. 121), pero los experimentos no son por ahora suficientemente precisos para que debamos asumir que nuestra geometría «experimentada» *difiere* realmente del ideal euclídeo (cf. p. 272), y es consistente con lo que la teoría de la relatividad general de Einstein nos dice que debería ser. Sin embargo, a pesar de los cambios que han tenido lugar en nuestra visión de la geometría del mundo físico, el concepto de Eudoxo de número real, con veintitrés siglos de edad, ha permanecido sin cambios esenciales y constituye un ingrediente tan esencial en la teoría de Einstein como en la de Euclides. ¡En realidad es un ingrediente esencial de cualquier teoría física sería hasta nuestros días!

El libro quinto de los *Elementos* de Euclides era básicamente una exposición de la «teoría de las proporciones», descrita arriba, que introdujo Eudoxo. Ésta era profundamente importante para la obra en conjunto. En realidad, los *Elementos* en su totalidad, publicados por primera vez alrededor del año 300 a.C., deben situarse como una de las obras con influencia más profunda de todos los tiempos. Ellos sientan el escenario de casi todo el pensamiento científico y matemático a partir de entonces. Sus métodos eran deductivos, partiendo de axiomas claramente enunciados que se suponían propiedades «autoevidentes» del espacio; de éstos se derivaban numerosas consecuencias, muchas de las cuales eran sorprendentes e importantes, y no autoevidentes en absoluto. No hay duda de que la obra de Euclides tuvo una significación profunda para el desarrollo del pensamiento científico posterior.

El matemático más grande de la antigüedad fue indudablemente Arquímedes (287-212 a.C.). Utilizando de modo ingenioso la teoría de las proporciones de Eudoxo, calculó las áreas y volúmenes de muchas formas geométricas diferentes, como la esfera, u otras más complejas que incluyen parábolas y espirales. Hoy utilizaríamos el cálculo integral para hacerlo, ¡pero esto era unos diecinueve siglos antes de la introducción del cálculo infinitesimal, como tal, por Newton y Leibniz! (¡Se podría decir que una buena mitad —la mitad «integral»— del cálculo ya era conocida para Arquímedes!) El grado de rigor matemático que Arquímedes alcanzó en sus argumentos era impecable, incluso para los niveles modernos. Sus escritos influyeron profundamente sobre muchos matemáticos y científicos posteriores, muy en especial Galileo y Newton. Arquímedes introdujo también la (¿SOBERBIA?) teoría física de la estática (es decir, las leyes que

gobiernan los cuerpos en equilibrio, como la ley de la palanca y las leyes de los cuerpos flotantes) y la desarrolló como una ciencia deductiva, de un modo semejante a como Euclides había desarrollado la ciencia de la geometría del espacio y la geometría de los cuerpos rígidos.

Un contemporáneo de Arquímedes a quien también debo mencionar es Apolonio (c. 262-200 a.C.), un geómetra de profunda intuición e ingenio, cuyo estudio de la teoría de las secciones cónicas (esto es, elipses, parábolas e hipérbolas) tuvo una influencia muy importante sobre Kepler y Newton. ¡Precisamente estas figuras geométricas resultaron ser, de forma bastante notable, las que se necesitaban para las descripciones de las órbitas planetarias!

La dinámica de Galileo y Newton

El profundo cambio que el siglo XVII aportó a la ciencia fue la comprensión del *movimiento*. Los antiguos griegos tenían una maravillosa comprensión de la estática —formas geométricas rígidas, o cuerpos en *equilibrio* (es decir, cuando todas las fuerzas están compensadas de modo que no hay movimiento)— pero no tenían una buena concepción de las leyes que gobiernan los cuerpos que se *mueven*. Lo que les faltaba era una buena teoría de la *dinámica*, esto es, una teoría del modo en que la Naturaleza controla el cambio de posición de los cuerpos de un instante al siguiente. Parte (pero no todas ni mucho menos) de las razones para esto era la ausencia de cualquier medio suficientemente preciso para medir el tiempo, es decir, de un «reloj» razonablemente bueno. Un reloj así es necesario para poder cronometrar exactamente los cambios en posición, y de este modo comprobar las velocidades y aceleraciones de los cuerpos. Por ello, la observación de Galileo, en 1583, de que un péndulo podía ser un medio fiable de medir el tiempo tuvo para él (¡y para el desarrollo de la ciencia moderna en general!) una enorme importancia puesto que se podía hacer un cronometraje preciso del movimiento. Unos cincuenta años más tarde, con la publicación de los *Discorsi* de Galileo en 1638, nacía la nueva ciencia de la dinámica ¡y empezaba la transformación del antiguo misticismo en ciencia moderna!

Seleccionaré sólo *cuatro* de las ideas físicas más importantes que introdujo Galileo. La primera era que una fuerza que actúa sobre un cuerpo determina la *aceleración*, y no la velocidad. ¿Qué significan realmente los términos «aceleración» y «velocidad»? La *velocidad* de

una partícula —o de un punto de algún cuerpo— es el ritmo de cambio, con respecto al tiempo, de la posición de dicho punto. Normalmente se toma la velocidad como una cantidad *vectorial*, lo que quiere decir que se debe considerar tanto su *dirección* como su magnitud (de lo contrario utilizamos el término «celeridad»; ver Fig. 5.4). La aceleración (de nuevo una cantidad vectorial) es el ritmo de cambio de esta velocidad con respecto al tiempo —de modo que la aceleración es realmente ¡el ritmo de cambio del ritmo de cambio de la posición con respecto al tiempo! (Hubiera sido difícil para los antiguos entender esto, al faltarles los «relojes» adecuados y las ideas matemáticas relevantes en relación con los «ritmos de cambio».) Galileo comprobó que la fuerza que actúa sobre un cuerpo (en su caso, la fuerza de la gravedad) controla la aceleración de dicho cuerpo pero *no* controla directamente su velocidad— tal como los antiguos, por ejemplo Aristóteles, habían creído.



Fig. 5.4. Velocidad, celeridad y aceleración.

En particular, no hay fuerza cuando la velocidad es constante —y por lo tanto, en *ausencia* de fuerzas resultará un movimiento uniforme en línea recta (lo que constituye la primera ley de Newton). Los cuerpos en movimiento libre continúan uniformemente su camino, y no se necesita ninguna fuerza que mantenga su marcha. De hecho, una consecuencia de las leyes dinámicas que establecieron Galileo y Newton es que el movimiento rectilíneo uniforme es físicamente indistinguible del estado de reposo (es decir, de ausencia de movimiento): ¡no existe modo de distinguir localmente el movimiento uniforme

del reposo! Galileo fue especialmente claro en este punto (incluso más claro que Newton) y dio una descripción muy gráfica en términos de un barco en el mar (cf. Drake, 1953, pp. 186-187):

Encerraos con un amigo en la cabina principal bajo la cubierta de un gran barco, llevando con vos moscas, mariposas y otros pequeños animales voladores. Llevad un gran recipiente con agua y algún pez dentro; colgad una botella que se vacía gota a gota en alguna vasija que hay debajo de ella. Con el barco aún en reposo, observad cuidadosamente cómo vuelan los pequeños animales con igual velocidad hacia todos los lados de la cabina. El pez nada indistintamente en todas las direcciones; las gotas caen en la vasija inferior;... Cuando hayáis observado cuidadosamente todas estas cosas... haced avanzar el barco con la velocidad que queráis, de forma que el movimiento sea uniforme y no haya oscilaciones en un sentido u otro. No descubriréis el menor cambio en ninguno de los efectos mencionados, ni podríais decir a partir de ellos si el barco está moviéndose o está quieto... Las gotas caerán como antes en la vasija inferior sin desviarse hacia la popa, aunque el barco haya avanzado mucho mientras las gotas están en el aire. El pez en el agua nadará hacia la parte delantera de su recipiente sin mayor esfuerzo que hacia la parte trasera, y se dirigirá con la misma facilidad hacia un cebo colocado en cualquier parte del borde del recipiente. Finalmente, las mariposas y moscas continuarán su vuelo indistintamente hacia cualquier lado, y no sucederá que se concentren hacia la popa como si se cansaran de seguir el curso del barco, del que hubieran quedado separadas una gran distancia si se hubieran mantenido en el aire.

Este hecho notable, llamado *principio de relatividad galileana*, es realmente crucial para que tenga sentido dinámico el punto de vista *copernicano*. Nicolás Copérnico (1473-1543, y el antiguo astrónomo griego Aristarco, c. 310-230 a.C. —¡no debe confundirse con Aristóteles!— dieciocho siglos antes que él) había presentado la imagen en la que el Sol permanece en reposo mientras que la Tierra, al mismo tiempo que gira sobre su propio eje, se mueve en una órbita en torno al Sol. ¿Por qué no somos conscientes de este movimiento que sería de unos 100.000 kilómetros por hora? Antes de que Galileo presentase su teoría dinámica, este hecho planteaba un verdadero y profundo enigma para el punto de vista copernicano. Si hubiera sido correcta la anterior visión «aristotélica» de la dinámica, en la que la *velocidad* real de un sistema en su movimiento a través del espacio afectaría a su comportamiento dinámico, entonces el movimiento de la Tierra sería en verdad algo muy directamente evidente para nosotros. La relatividad galileana pone en claro cómo puede estar la Tierra en

movimiento aunque su movimiento no es algo que podamos percibir directamente *.

Nótese que, según la relatividad galileana, no se puede asociar ningún significado físico local al concepto de estar «en reposo». Esto tiene ya notables consecuencias para el modo de considerar el espacio y el tiempo. La imagen que intuitivamente tenemos sobre el espacio y el tiempo es que el «espacio» constituye una especie de arena en la que tienen lugar los sucesos físicos. Un objeto físico puede estar en un punto del espacio en un instante y en el mismo o un punto diferente del espacio en un instante posterior. Imaginemos que los puntos del espacio persisten de alguna manera entre un instante y el siguiente, de modo que tenga significado decir que el objeto ha cambiado o no su posición espacial. Ahora bien, la relatividad galileana nos dice que no hay significado absoluto para el «estado de reposo», así que no se puede asociar ningún significado a «el mismo punto del espacio en dos instantes diferentes». ¿Qué punto del espacio euclídeo tri-dimensional de la experiencia física en un instante dado es el «mismo» punto de nuestro espacio euclídeo tri-dimensional en otro instante? No hay manera de decirlo. ¡Parece como si debiéramos tener un espacio euclídeo completamente *nuevo* para cada instante de tiempo! La forma de que esto tenga sentido es considerar una imagen de la realidad física en un *espacio-tiempo tetra-dimensional* (ver Fig. 5.5). Los espacios euclídeos tridimensionales correspondientes a los diferentes instantes de tiempo se consideran realmente como independientes uno de otro, pero todos estos espacios están unidos para formar la imagen completa de nuestro espacio-tiempo tetra-dimensional. Las historias de las partículas que se mueven con movimiento rectilíneo uniforme se describen mediante líneas rectas (llamadas líneas de universo) en el espacio-tiempo. Volveré a la cuestión del espacio-tiempo, y la relatividad del movimiento, más adelante en el contexto de la relatividad einsteniana. Encontraremos que en este caso el argumento a favor de la tetra-dimensionalidad tiene una fuerza considerablemente mayor.

La tercera de estas grandes intuiciones de Galileo fue el inicio de una comprensión de la *conservación de la energía*. Galileo estaba in-

* Estrictamente hablando, esto se refiere sólo al movimiento de la Tierra en tanto que pueda considerarse como aproximadamente *uniforme* y, en particular, sin rotación. El movimiento rotacional de la Tierra tiene realmente efectos dinámicos (relativamente pequeños) que pueden ser detectados, siendo el más digno de mención la desviación de los vientos en sentidos diferentes en los hemisferios Norte y Sur. Galileo pensaba que esta no uniformidad era responsable de las mareas.

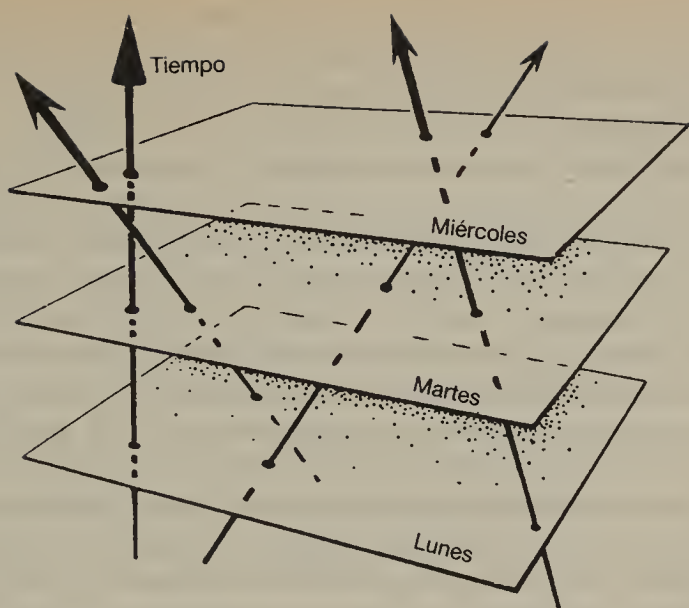


Fig. 5.5. Espacio-tiempo galileano: las partículas en movimiento uniforme se representan como líneas rectas.

interesado principalmente en el movimiento de los objetos sometidos a la gravedad. Él notó que si un cuerpo queda liberado a partir del reposo, entonces ya caiga libremente, o cuelgue de un péndulo de longitud arbitraria, o se deslice por un plano inclinado, su velocidad de movimiento depende siempre *sólo* de la distancia que ha alcanzado por debajo del punto en que se ha soltado. Además su velocidad es siempre la justa para volver a la altura de la que partió. Como diríamos ahora, la energía almacenada en su altura sobre el suelo (energía potencial gravitatoria) puede transformarse en la energía de su movimiento (energía cinética que depende de la *celeridad* del cuerpo) y viceversa, pero la energía en conjunto no se gana ni se pierde.

La ley de la conservación de la energía es un principio físico muy importante. No es un requisito físico independiente, sino una *consecuencia* de las leyes dinámicas de Newton a las que llegaremos en breve. En el curso de los siglos se hicieron formulaciones cada vez más generales de esta ley por Descartes, Huygens, Leibniz, Euler y Kelvin. Volveremos a ello más tarde y en el Capítulo 7. Resulta que, cuando se combina con el principio de relatividad de Galileo, la conservación de la energía da nuevas leyes de conservación de considerable importancia: conservación de *masa* y de *momento*. El momento

de una partícula es el producto de su masa por su velocidad. Ejemplos conocidos de conservación del momento ocurren en los cohetes a reacción, en donde el incremento de momento hacia adelante del cohete compensa exactamente el momento hacia atrás de los gases expulsados (de menor masa pero, en compensación, mucho más rápidos). El retroceso de un fusil es también una manifestación de una conservación del momento. Otra consecuencia de las leyes de Newton es la conservación del *momento angular* que describe la persistencia de la rotación de un sistema. Tanto la rotación de la Tierra en torno a su eje como la de una pelota de tenis se mantienen a través de la conservación de su momento angular. Cada partícula constituyente de un cuerpo contribuye al momento angular del mismo, para el que la magnitud de la contribución de una partícula es el producto de su momento por su distancia perpendicular al centro. (En consecuencia, puede incrementarse la velocidad angular de un objeto que rota libremente si se le hace más compacto. Esto conduce a la sorprendente, aunque familiar, acción que realizan a menudo los patinadores y trapecistas. El acto de recoger los brazos o las piernas, según sea el caso, provoca un incremento espontáneo de la velocidad de rotación, ¡debido simplemente a la conservación del momento angular!) Veremos que masa, energía, momento y momento angular son conceptos que tendrán importancia para nosotros más adelante.

Finalmente, recordaré al lector la profética intuición de Galileo de que, en ausencia de fricción atmosférica, todos los cuerpos sometidos a la gravedad caen con la misma velocidad. (Quizá recuerde el lector la famosa historia de Galileo dejando caer varios objetos simultáneamente desde la Torre Inclinada de Pisa.) Tres siglos después, esta misma intuición condujo a Einstein a generalizar el principio de relatividad a los sistemas de referencia acelerados, y proporcionó la piedra angular de su extraordinaria teoría de la relatividad general para la gravitación, como veremos cerca del final de este capítulo.

Sobre los impresionantes cimientos que Galileo había colocado, Newton pudo erigir una catedral de soberbia grandeza. Newton dio las tres leyes que gobiernan el comportamiento de los objetos materiales. La primera y segunda leyes eran esencialmente las dadas por Galileo: si sobre un cuerpo no actúa ninguna fuerza, éste continuará moviéndose uniformemente en línea recta; si una fuerza actúa sobre él, entonces el producto de su masa por su aceleración (es decir, el ritmo de cambio de su momento), es igual a dicha fuerza. Una de las intuiciones propias de Newton fue el darse cuenta de la necesidad de una tercera ley: la fuerza que un cuerpo A ejerce sobre un cuerpo B

es exactamente igual y opuesta a la fuerza que el cuerpo B ejerce sobre el cuerpo A («para toda acción existe una reacción igual y opuesta»). Esto proporciona el marco básico. El «universo newtoniano» consta de partículas que se mueven en un espacio que está sujeto a las leyes de la geometría euclídea. Las aceleraciones de dichas partículas están determinadas por las fuerzas que actúan sobre ellas. La fuerza sobre cada partícula se obtiene sumando (utilizando la *ley de suma vectorial*; ver Fig. 5.6) todas las contribuciones separadas a la fuerza sobre esa partícula debidas a todas las *demás* partículas. Para que el sistema esté bien definido se necesita alguna regla precisa que nos diga cuál sería la fuerza que aparece sobre la partícula A debida a otra partícula B. Normalmente se exige que esta fuerza actúe a lo largo de una línea recta entre A y B (ver Fig. 5.7). Si la fuerza es una fuerza gravitatoria, entonces actúa atractivamente entre A y B y su intensidad es proporcional al producto de las dos masas y a la inversa del cuadrado de la distancia entre ellas: la *ley de la inversa del cuadrado*. Para otros tipos de fuerza podría haber una dependencia de la posición distinta de ésta, y la fuerza podría depender de las partículas de acuerdo con alguna cualidad que posean distinta de sus masas.

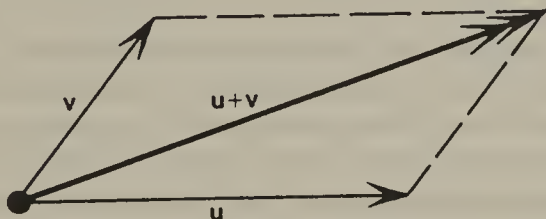


Fig. 5.6. Ley del paralelogramo para la suma vectorial.

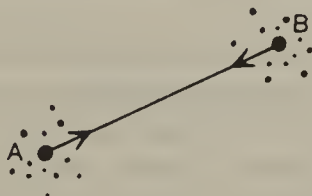


Fig. 5.7. La fuerza entre dos partículas se considera dirigida a lo largo de la línea recta que las une (y por la tercera ley de Newton, la fuerza sobre A debida a B es siempre igual y opuesta a la fuerza sobre B debida a A).

El gran Johannes Kepler (1571-1630), un contemporáneo de Galileo, había notado que las órbitas de los planetas en torno al Sol eran *elípticas* más que circulares (con el Sol estando siempre en un foco, no en el centro, de la elipse) y formuló otras dos leyes que gobiernan los ritmos con que se describen las elipses. Newton pudo demostrar que las tres leyes de Kepler se deducen de su propio esquema general (con una ley de fuerza atractiva inversa del cuadrado). No sólo esto, sino que también obtuvo todo tipo de correcciones detalladas a las órbitas elípticas de Kepler, así como otros efectos, tales como la precesión de los equinoccios (un lento movimiento de la dirección del eje de rotación, que había sido notado por los griegos a lo largo de los siglos). Para lograr todo esto, Newton tuvo que desarrollar muchas técnicas matemáticas —además del cálculo diferencial. El éxito excepcional de sus esfuerzos debió mucho a sus supremas habilidades matemáticas y a sus igualmente soberbias intuiciones físicas.

El mundo mecanicista de la dinámica newtoniana

Con una ley de fuerzas específica (como la ley de la inversa del cuadrado para la gravitación) el esquema newtoniano se traduce en un preciso y determinado sistema de ecuaciones dinámicas. Si se especifican las posiciones, velocidades y masas de las diversas partículas en un instante, entonces sus posiciones y velocidades (y sus masas, pues éstas se consideran *constantes*) están matemáticamente determinadas para todos los instantes posteriores. Esta forma de *determinismo*, satisfecha por el mundo de la mecánica newtoniana, tuvo (y aún la tiene) una profunda influencia sobre el pensamiento filosófico. Tratemos de examinar un poco más estrechamente la naturaleza de este determinismo newtoniano. ¿Qué puede decirnos sobre la cuestión del «libre albedrío»? ¿Podría contener mentes un mundo estrictamente newtoniano? ¿Puede un mundo newtoniano contener siquiera máquinas computadoras?

Tratemos de ser razonablemente concretos sobre este modelo «newtoniano» del mundo. Podemos suponer, por ejemplo, que todas las partículas constituyentes de la materia se consideran como puntos materiales, esto es, sin ninguna extensión espacial. Como alternativa podríamos considerarlas como bolas esféricas rígidas. En uno u otro caso tendremos que suponer que nos son conocidas las leyes de fuerzas, como la ley de atracción inversa del cuadrado de la teoría gravitatoria de Newton. Queremos modelar también otras fuerzas de la

Naturaleza, tales como las fuerzas *eléctrica* y *magnética* (estudiadas por primera vez en detalle por William Gilbert en 1600), o las fuerzas nucleares *fuertes* que ahora sabemos son las que ligan partículas (protones y neutrones) para formar núcleos atómicos. Las fuerzas eléctricas se parecen a las gravitatorias en que también satisfacen una ley de la inversa del cuadrado, pero aquí las partículas semejantes se *repelen* entre sí (en lugar de atraerse, como en el caso gravitatorio) y no son las masas de las partículas las que gobiernan la intensidad de las fuerzas eléctricas mutuas sino sus *cargas eléctricas*. Las fuerzas magnéticas son también del tipo de la «inversa del cuadrado» como las eléctricas *, pero las fuerzas nucleares tienen una dependencia de la distancia bastante diferente, siendo extremadamente grandes para las distancias muy cortas que separan las partículas dentro de los núcleos atómicos pero despreciables a mayores distancias.

Supongamos que adoptamos la imagen de la esfera rígida, exigiendo que cuando colisionan dos de las esferas rebotan simplemente de forma perfectamente *elástica*; es decir, se separan de nuevo sin ninguna pérdida de energía (ni de momento total) como si fueran perfectas bolas de billar. También tenemos que especificar exactamente cómo son las *fuerzas* que actúan entre una bola y otra. Por simplicidad, podemos suponer que la fuerza que cada bola ejerce sobre cada otra bola está dirigida a lo largo de la línea recta que une sus centros, y su magnitud es una determinada función de la longitud de esta línea. (Para la *gravitación* newtoniana esta suposición es automáticamente cierta por un famoso teorema debido a Newton; y para otras leyes de fuerza puede imponerse como un requisito consistente.) Con tal de que las bolas colisionen sólo en pares, y no ocurran colisiones triples o de mayor orden, entonces todo está bien definido y el resultado depende de manera continua del estado inicial (es decir, cambios suficientemente pequeños en el estado inicial conducen sólo a cambios pequeños en el resultado). Hay continuidad entre el comportamiento de las colisiones rasantes y el comportamiento de las bolas cuando apenas pasan una al lado de la otra. Hay, sin embargo, un problema en qué hacer con las colisiones triples o de mayor orden. Por ejemplo, si tres bolas A, B y C chocan al mismo tiempo, es diferente que consideremos que A y B chocan primero y C colisiona con B inme-

* La diferencia entre los casos eléctrico y magnético es que no parece que existan «cargas magnéticas» aisladas (es decir, polos Norte y Sur) en la Naturaleza, sino que las partículas magnéticas son lo que se denomina «dipolos», esto es, minúsculos imanes (con polos Norte y Sur inseparables).

diatamente después, o que consideremos que A y C chocan primero y B colisiona con A inmediatamente después (ver Fig. 5.8). ¡En nuestro modelo existe *indeterminismo* cuando ocurren colisiones triples exactas! Si queremos podemos simplemente *excluir* las colisiones triples o de mayor orden como «infinitamente improbables». Esto proporciona un esquema razonablemente consistente, pero el problema potencial de las colisiones triples significa que el comportamiento resultante puede *no* depender de forma continua del estado inicial.

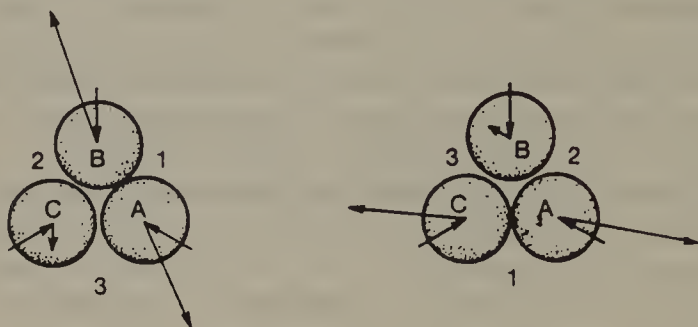


Fig. 5.8. Una colisión triple. El comportamiento resultante depende críticamente de qué partículas chocan primero, de modo que el resultado depende de manera discontinua de la situación inicial.

Esto no es totalmente satisfactorio, y podemos preferir una imagen en términos de partículas *puntuales*. Pero para evitar ciertas dificultades teóricas planteadas por el modelo de partículas puntuales (fuerzas infinitas y energías infinitas cuando las partículas llegan a coincidir) debemos hacer otras suposiciones, como la de que las fuerzas entre las partículas se hacen siempre fuertemente repulsivas a cortas distancias. De esta manera podemos asegurar que nunca colisionará realmente un par de partículas. (¡Esto también nos permite evitar el problema de decidir cómo *se supone* que se comportan las partículas cuando colisionan!) Sin embargo, por facilidad de visualización, prefiero expresar la discusión que sigue completamente en términos de esferas rígidas. ¡Parece que este tipo de imagen de «bola de billar» es esencialmente el modelo de *realidad* que mucha gente tiene en el fondo de su mente!

Ahora (pasando por alto el problema de las colisiones múltiples), la imagen de la bola de billar newtoniana ⁴ de la realidad constituye efectivamente un modelo *determinista*. Debemos tomar la palabra «de-

terminista» en el sentido de que el comportamiento físico está matemáticamente determinado en su totalidad para cualquier instante en el futuro (o pasado) por las posiciones y velocidades de todas las bolas (supuestas en número finito, pongámoslo así, para evitar ciertos problemas) en un instante *cualquiera*. Parece, entonces, que en un mundo de bolas de billar no hay lugar para una «mente» que influya en el comportamiento de las cosas materiales mediante la acción de su «libre albedrío». Si creemos en el «libre albedrío» parece que estuviéramos obligados a dudar de que nuestro mundo *real* pueda estar formado de este modo.

La controvertida cuestión del «libre albedrío» ronda en el trans-fondo de este libro —aunque, para la mayor parte de lo que tendré que decir, quedará solo en el transfondo. Tendrá un papel específico, aunque menor, que jugar más adelante en este mismo capítulo (en relación con el tema del envío de señales más-rápidas-que-la-luz en relatividad). La cuestión del libre albedrío se aborda directamente en el Capítulo 10, y allí el lector quedará sin duda desilusionado por lo que tengo que aportar. Creo realmente que hay aquí un problema real, y no sólo uno supuesto, pero es profundo y difícil de formular adecuadamente. El problema del *determinismo* en la teoría física es importante pero creo que es sólo una parte de la historia. El mundo podría ser, por ejemplo, determinista pero *no-computable*. En tal caso, el futuro podría estar determinado por el presente de un modo que *en principio* es no-calculable. En el Capítulo 10 trataré de presentar argumentos para demostrar que la acción de nuestras mentes conscientes es realmente no-algorítmica (es decir, no computable). En consecuencia, el libre albedrío del que nosotros mismos nos creemos dotados tendría que estar íntimamente ligado a algún ingrediente no-computable en las leyes que gobiernan el mundo en que realmente vivimos. Es una cuestión interesante —ya aceptemos o no este punto de vista respecto al libre albedrío— la de si una teoría física dada (como la de Newton) es en realidad *computable*, y no ya sólo si es determinista. Computabilidad es una cuestión diferente de determinismo —y el hecho de que *sea* una cuestión diferente es algo que estoy tratando de subrayar en este libro.

¿Es computable la vida en el mundo de las bolas de billar?

Déjenme ilustrar primero, con un ejemplo en verdad absurdamente artificial, el hecho de que computabilidad y determinismo *son* di-

ferentes, exhibiendo un «modelo de universo de juguete» que es determinista pero no computable. Describamos el «estado» del sistema en cualquier «instante» mediante un par de números naturales (m, n) . Sea T_u una máquina universal de Turing bien determinada, por ejemplo la definida concretamente en el Capítulo 2, pp. 86-87). Para decidir cuál es el estado de este universo en el próximo «instante de tiempo» debemos preguntar si la acción de T_u sobre m se llegará a parar o no (es decir, si $T_u(m) \neq \square$ o $T_u(m) = \square$, en la notación del Capítulo 2, p. 90). Si se para, el estado en este próximo «instante» es $(m+1, n)$. Si no se para, será $(n+1, m)$. Vimos en el Capítulo 2 que no existe algoritmo para el problema de la detención de máquinas de Turing. De ello se sigue que no puede haber algoritmo para predecir el «futuro» en este universo modelo, ¡a pesar de que es completamente determinista!

Por supuesto, éste no es un modelo para tomar en serio, pero muestra que existe un problema a resolver. Podemos preguntar a *cualquier* teoría física determinista si es o no computable. De hecho, ¿es computable el mundo de las bolas de billar newtonianas?

El tema de la computabilidad física depende en parte del tipo de pregunta que nos propongamos plantear al sistema. Puedo imaginar cierto número de preguntas que podrían plantearse para las que mi *conjetura* sería que *no* son computables (esto es, no es asunto algorítmico el verificar la respuesta) en un modelo de bolas de billar newtonianas. Una de estas cuestiones podría ser: ¿colisiona alguna vez la bola A con la bola B? La idea es que se nos darían, como *datos iniciales*, las posiciones y velocidades de todas las bolas en cierto instante de tiempo ($t=0$) y el problema consiste en calcular a partir de estos datos si las bolas A y B colisionarán o no alguna vez en cualquier instante posterior ($t>0$). Para hacer el problema más concreto (aunque no particularmente realista) podemos suponer que todas las bolas son del mismo radio y la misma masa y que hay, por ejemplo, una ley de fuerzas del tipo de la inversa del cuadrado actuando entre cada par de bolas. Una razón para conjeturar que esta pregunta concreta no es de las que pueden resolverse algorítmicamente es que el modelo es en cierto modo semejante a un «modelo de bolas de billar para una computación» que fue introducido por Edward Fredkin y Tommaso Toffoli (1982). En su modelo (en lugar de tener una ley de fuerzas de la inversa del cuadrado) las bolas están limitadas por varias «paredes», pero rebotan elásticamente una sobre otra de un modo semejante a las bolas newtonianas que acabo de describir (ver Fig. 5.9). En el modelo de Fredkin-Toffoli todas las operaciones ló-

gicas básicas de un ordenador pueden ejecutarse mediante las bolas. Cualquier computación de una máquina de Turing puede ser imitada: la elección particular de la máquina de Turing T_n define la configuración de las «paredes», etc., de la máquina de Fredkin-Toffoli; una vez hecho esto, un estado inicial de bolas codifica la información de la cinta input, y la cinta output de la máquina de Turing queda codificada en el estado final de las bolas. Así puede plantearse, en particular, la pregunta: ¿se parará alguna vez tal-o-cual máquina computadora de Turing? «Parada» puede parafrasearse en términos de que la bola A colisione finalmente con la bola B. El hecho de que esta cuestión no pueda ser resuelta algorítmicamente (pp. 96-97) *sugiere* al menos que la pregunta newtoniana «¿colisiona alguna vez la bola A con la bola B?», que plantee inicialmente, no podrá ser tampoco respondida algorítmicamente.

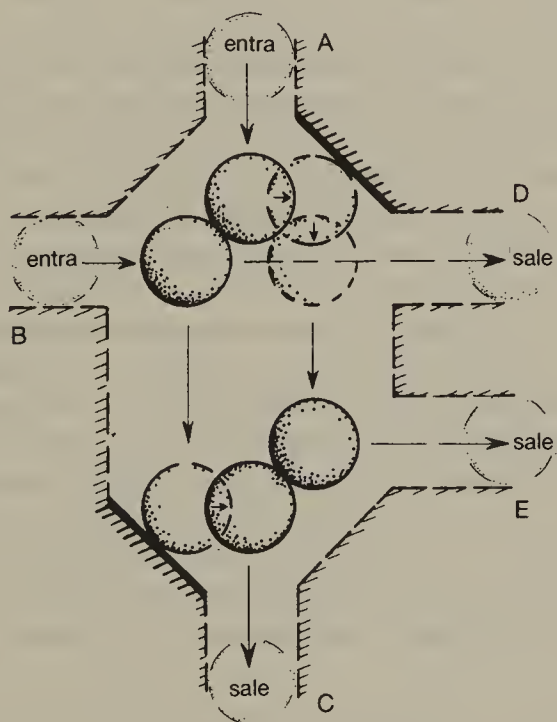


Fig. 5.9. Un «conmutador» (sugerido por A. Resler) en el computador de bolas de billar de Fredkin-Toffoli. Si una bola entra por B, una bola saldrá a continuación por D o por E dependiendo de si otra bola entra por A (en donde las entradas por A y B se suponen simultáneas).

En realidad, el problema newtoniano es mucho más complicado que el desarrollado por Fredkin y Toffoli. Éstos pudieron especificar los estados de su modelo en términos de parámetros *discretos* (es decir, en términos de enunciados «sí o no» como «o la bola está en el canal o no lo está»). Pero en el problema newtoniano completo las posiciones y velocidades de las bolas tienen que especificarse con precisión infinita en términos de coordenadas que son *números reales*, en lugar de en forma discreta. De este modo, nos enfrentamos otra vez con todos los problemas que tuvimos que considerar cuando en el Capítulo 4 abordábamos la cuestión de si el conjunto de Mandelbrot es o no recursivo. ¿Qué *significa* «computable» cuando se admiten como datos de entrada y salida parámetros que varían de forma continua? Por el momento, el problema puede atenuarse suponiendo que todas las coordenadas de posición y velocidad iniciales vienen dadas por *números racionales* (aunque no podemos esperar que tales coordenadas permanezcan racionales para posteriores valores racionales del tiempo t). Recordemos que un número racional es un cociente entre dos enteros, y por consiguiente está definido en términos discretos finitos. Utilizando números racionales podemos aproximar, tanto como queramos, cualesquiera conjuntos de datos iniciales que hayamos decidido examinar. No es completamente irrazonable conjeturar que, con datos iniciales racionales, pueda no existir algoritmo para decidir si las bolas A y B colisionan finalmente o no.

Sin embargo, esto no es realmente lo que queremos decir con una afirmación tal como «el mundo de las bolas de billar newtonianas no es computable». El modelo particular que he estado comparando con nuestro mundo de bolas de billar newtonianas, a saber, el «computador de bolas de billar» de Fredkin-Toffoli, no actúa realmente de acuerdo con un cálculo. Éste, después de todo, era el punto esencial de la idea de Fredkin y Toffoli —¡que su modelo se comportara como un computador (universal)! El tipo de problema que estoy tratando de plantear es el de si es concebible que un cerebro humano pueda, aprovechando algunas leyes físicas «no-computables», «superar» en algún sentido a una máquina de Turing. No sirve tratar de aprovechar algo como:

«Si la bola A nunca choca con la bola B entonces la respuesta a su problema es “No”.»

¡Podríamos tener que esperar indefinidamente para asegurar que las bolas en cuestión no chocan nunca! Éste es, por supuesto, el modo en que se *comporta* una máquina de Turing.

Parece que, en efecto, hay claros indicios de que, en un sentido apropiado, el mundo de las bolas de billar newtonianas es computable (al menos si prescindimos del problema de las colisiones múltiples). Normalmente trataríamos de computar este comportamiento haciendo algunas aproximaciones. Podríamos imaginar que se especifica que los centros de las bolas están en una malla de puntos, en donde los puntos nodales de la malla son precisamente aquellos que miden las coordenadas (en centésimas de unidad, por ejemplo). El tiempo se considera también «discreto»: todos los instantes permitidos son múltiplos de alguna pequeña unidad (denotada por Δt , por ejemplo). Esto plantea algunas posibilidades discretas para las «velocidades» (diferencias en los valores de la posición de los nodos en dos instantes permitidos sucesivos divididas por Δt). Las aproximaciones apropiadas para las aceleraciones se calculan utilizando la ley de fuerzas, y estas aceleraciones se utilizan para obtener las «velocidades» a partir de las que se calculan, con el grado de aproximación requerido, las nuevas posiciones de los nodos en el próximo instante permitido. El cálculo continúa durante tantos intervalos temporales como sea posible mientras se mantenga la precisión deseada. Puede ser que no se puedan computar muchos instantes antes de que se pierda toda precisión. El método consiste entonces en empezar con una malla espacial considerablemente más fina y una algo más fina división de los instantes de tiempo permitidos. Esto permite alcanzar una mayor precisión y el cálculo puede llevarse más lejos que en el caso anterior antes de que se pierda la precisión. Con una malla espacial aún más fina, y una división más fina de los intervalos de tiempo, la precisión puede aumentarse aún más y llevar el cálculo aún más lejos en el futuro. De esta forma, el mundo de las bolas de billar newtonianas puede computarse tan estrechamente como queramos (pasando por alto las colisiones múltiples) —y, en este sentido, podemos decir que el mundo newtoniano es realmente computable.

Existe un sentido, no obstante, en el que este mundo es «no-computable» *en la práctica*. Esto surge del hecho de que la precisión con que pueden *conocerse* los datos iniciales es siempre limitada. De hecho, existe una «inestabilidad» muy considerable inherente a este tipo de problemas. Un pequeñísimo cambio en los datos iniciales puede dar lugar rápidamente a un cambio absolutamente enorme en el comportamiento resultante. (¡Cualquiera que haya tratado de entronerar una bola de billar americano, golpeándola con una bola intermedia que, a su vez, debe ser golpeada antes sabrá lo que quiero decir!) Esto es particularmente evidente cuando están implicadas colisiones

(sucesivas), pero tales inestabilidades en el comportamiento pueden ocurrir también con la acción a distancia gravitatoria de Newton (con más de dos cuerpos). A menudo se utiliza el término «caos», o «comportamiento caótico», para este tipo de inestabilidad. El comportamiento caótico es importante, por ejemplo, en relación con el clima. ¡Aunque se conocen las ecuaciones newtonianas que gobiernan los elementos, las predicciones del tiempo a largo plazo son notoriamente poco fiables!

Éste no es en absoluto el tipo de «no-computabilidad» que pueda ser «aprovechada» de forma alguna. Consiste simplemente en que, puesto que hay un límite a la precisión con que puede conocerse el estado inicial, el estado futuro no puede ser computado fiablemente a partir del inicial. Un *elemento aleatorio* se ha introducido en efecto en el comportamiento futuro, pero esto es todo. Si el cerebro humano está recurriendo realmente a elementos no-computables *útiles* que existen en las leyes físicas, éstos deben ser de un tipo completamente diferente y de un carácter mucho más positivo. Consiguientemente, no me referiré a este tipo de comportamiento «caótico» como «no-computabilidad» en absoluto, prefiriendo utilizar el término «impredecibilidad». La presencia de impredecibilidad es un fenómeno muy general en el tipo de leyes deterministas que aparecen en la física (clásica), como pronto veremos. ¡La impredecibilidad es ciertamente algo que nos gustaría *minimizar* más que «aprovechar» en la construcción de una máquina pensante!

Para discutir las cuestiones de computabilidad e impredecibilidad será conveniente adoptar un punto de vista más amplio que antes respecto a las leyes físicas. Esto nos facilitará el considerar no sólo el esquema de la mecánica newtoniana sino también las teorías posteriores que han venido a reemplazarla. Tendremos que echar una ojeada a la notable formulación *hamiltoniana* de la mecánica.

Mecánica hamiltoniana

Los éxitos de la mecánica newtoniana se derivaban no sólo de su soberbia aplicabilidad al mundo físico sino también de la riqueza de la teoría matemática a que dio lugar. Resulta notable que *todas* las teorías SOBERBIAS de la Naturaleza han resultado ser extraordinariamente fértiles como fuente de ideas matemáticas. Hay un profundo y bello misterio en el hecho de que estas teorías soberbiamente precisas son también extraordinariamente fructíferas como simples *mate-*

máticas. Sin duda esto nos está diciendo algo profundo sobre la conexión entre el mundo real de nuestra experiencia física y el mundo platónico de las matemáticas. (Intentaré abordar este tema más tarde, en el Capítulo 10 p. 533.) La mecánica newtoniana ocupa tal vez un lugar supremo a este respecto puesto que su nacimiento dio lugar al cálculo infinitesimal. Además, el esquema newtoniano propiamente dicho ha dado lugar a un notable cuerpo de ideas matemáticas conocido como *mecánica clásica*. Los nombres de muchos de los grandes matemáticos de los siglos XVIII y XIX están asociados a este desarrollo: Euler, Lagrange, Laplace, Liouville, Poisson, Jacobi, Ostrogradski, Hamilton. Lo que se conoce como «teoría hamiltoniana» resume gran parte de esta obra, y para nuestro propósito bastará una muestra de ella. El polifacético y original matemático irlandés William Rowan Hamilton (1805-1865) —a quien también se deben los circuitos hamiltonianos discutidos en la p. 191— había desarrollado esta formulación de la teoría de una manera que realizaba una analogía con la propagación de ondas. Esta idea de una relación entre ondas y partículas —y la forma de las propias ecuaciones de Hamilton— fue muy importante para el desarrollo posterior de la *mecánica cuántica*. Volveré a este aspecto de las cosas en el próximo capítulo.

Un ingrediente novedoso del esquema hamiltoniano reside en las «variables» que se utilizan en la descripción de un sistema físico. Hasta ahora, las *posiciones* de las partículas se consideraban como primarias, siendo las velocidades simplemente los ritmos de cambio de las posiciones con respecto al tiempo. Recuérdese (p. 218) que en la especificación del estado inicial de un sistema newtoniano necesitábamos las posiciones y las velocidades de todas las partículas para que quede determinado el comportamiento subsiguiente. Con la formulación hamiltoniana debemos seleccionar los *momentos* de las partículas más que las velocidades. (Señalamos en p. 215 que el momento de una partícula es simplemente el producto de su velocidad por su masa.) Esto podría parecer en sí mismo un cambio pequeño, pero lo importante es que la posición y el momento de cada partícula van a ser tratados como si fueran variables *independientes*, más o menos en pie de igualdad. De este modo se «pretende», en primer lugar, que los momentos de las diversas partículas no tengan nada que ver con los ritmos de cambio de sus variables de posición respectivas sino que son simplemente un conjunto separado de variables, de modo que podemos imaginar que «podrían» haber sido completamente independientes de los movimientos de posición. En la formulación hamiltoniana tenemos ahora *dos* conjuntos de ecuaciones. Uno de éstos nos

dice cómo cambian con el tiempo los *momentos* de las diversas partículas, y el otro nos dice cómo cambian con el tiempo las *posiciones*. En ambos casos los ritmos de cambio están determinados por las diversas posiciones y momentos *en* dicho instante.

Hablando en términos generales, el primer conjunto de ecuaciones de Hamilton establece la crucial segunda ley de movimiento de Newton (fuerza = ritmo de cambio del momento) mientras que el segundo conjunto de ecuaciones nos dice qué *son* realmente los momentos en términos de las velocidades (en efecto, ritmo de cambio de la posición = momento $\div$ masa). Recordemos que las leyes de movimiento de Galileo-Newton venían descritas en términos de aceleraciones, es decir, ritmos de cambio de ritmos de cambio de posición (esto es, ecuaciones de «segundo orden»). Ahora sólo necesitamos hablar de ritmos de cambio de objetos (ecuaciones de «primer orden») en lugar de ritmos de cambio de ritmos de cambio de objetos. Todas estas ecuaciones se derivan simplemente de una cantidad importante: la *función hamiltoniana* H , que es la expresión para la *energía total* del sistema en términos de todas las variables de posición y momento.

La formulación hamiltoniana proporciona una descripción de la mecánica muy elegante y simétrica. Sólo para ver qué aspecto tiene, escribiremos aquí las ecuaciones aun cuando muchos lectores no estén familiarizados con las notaciones del cálculo que se requieren para una comprensión completa —que no será necesaria aquí. Todo lo que realmente tenemos que comprender, en lo que se refiere al cálculo, es que el «punto» que aparece en el primer miembro de cada ecuación representa el *ritmo de cambio con respecto al tiempo* (del momento, en el primer caso, y de la posición, en el segundo):

$$\dot{p}_i = - \frac{\partial H}{\partial x_i}, \quad \dot{x}_i = \frac{\partial H}{\partial p_i} .$$

Aquí el índice i se utiliza simplemente para distinguir las diferentes coordenadas de momento $p_1, p_2, p_3, p_4, \dots$ y las de posición $x_1, x_2, x_3, x_4, \dots$. Para n partículas sin ligaduras tendremos $3n$ coordenadas de momento y $3n$ coordenadas de posición (una para cada una de las tres direcciones independientes del espacio). El símbolo ∂ denota la «derivada parcial» («tomar derivadas mientras se mantienen constantes todas las demás variables»), y H es la función hamiltoniana descrita arriba. (Si usted no sabe nada sobre «diferenciación» no se preocupe. Piense sólo en los segundos miembros de estas ecuaciones como ex-

presiones matemáticas perfectamente bien definidas escritas en términos de las x_i y las p_i .)

Las coordenadas $x_1, x_2, \dots$ y $p_1, p_2, \dots$ pueden ser realmente cosas más generales que las simples coordenadas cartesianas de las partículas (esto es, con las x_i siendo distancias medidas en tres direcciones diferentes que forman ángulos rectos). Algunas de las coordenadas x_i podrían ser *ángulos*, por ejemplo (en cuyo caso las correspondientes p_i serían momentos *angulares*, cf. p. 216, en lugar de momentos), o alguna otra magnitud completamente general. Curiosamente, las ecuaciones hamiltonianas aún mantienen exactamente la misma forma. De hecho, con elecciones apropiadas de H , las ecuaciones de Hamilton siguen siendo verdaderas para *cualquier* sistema de ecuaciones clásicas, no sólo para las ecuaciones de Newton. Éste será el caso, en particular, para la teoría de Maxwell(-Lorentz) que consideraremos dentro de poco. Las ecuaciones de Hamilton siguen siendo válidas también para la relatividad especial. Incluso la relatividad general, si se pone el cuidado debido, puede ser incluida en el marco hamiltoniano. Además, como veremos más adelante con la ecuación de Schrödinger (p. 363), este marco hamiltoniano proporciona el punto de partida para las ecuaciones de la mecánica cuántica. ¡Semejante unidad formal en la estructura de las ecuaciones dinámicas, a pesar de todos los cambios revolucionarios que han ocurrido en las teorías físicas durante el siglo pasado más o menos, es algo verdaderamente notable!

Espacio de fases

La forma de las ecuaciones hamiltonianas nos permite «visualizar» de un modo muy potente y general la evolución de un sistema físico. Tratemos de imaginar un «espacio» de un gran número de dimensiones, una para cada una de las coordenadas $x_1, x_2, \dots, p_1, p_2, \dots$ (Los espacios matemáticos tienen con frecuencia muchas más de tres dimensiones.) Este espacio se llama *espacio de fases* (ver Fig. 5.10). Para n partículas sin ligaduras, éste será un espacio de $6n$ dimensiones (tres coordenadas de posición y tres coordenadas de momento por cada partícula). ¡El lector puede lamentarse de que, incluso para una *sola* partícula, estas dimensiones son ya el doble de las dimensiones que ella o él utilizarían normalmente para visualizar! El secreto está en no dejarse asustar por esto. Aunque seis dimensiones son realmente más dimensiones de las que podemos imaginar fácilmente (!) no

sería tampoco de mucha utilidad que pudiéramos representarlas efectivamente. Sólo para una habitación llena de moléculas de aire, el número de dimensiones del espacio de fases podría ser algo como

10 000 000 000 000 000 000 000 000 000 .

¡No hay muchas esperanzas en tratar de obtener una visualización precisa de un espacio *tan* grande! Por ello, el truco es no intentarlo siquiera —aun en el caso del espacio de fases para una sola partícula. Pensemos simplemente en algún vago tipo de región tri-dimensional (o incluso solamente bi-dimensional). Echemos otra ojeada a la Figura 5.10. Eso bastará.

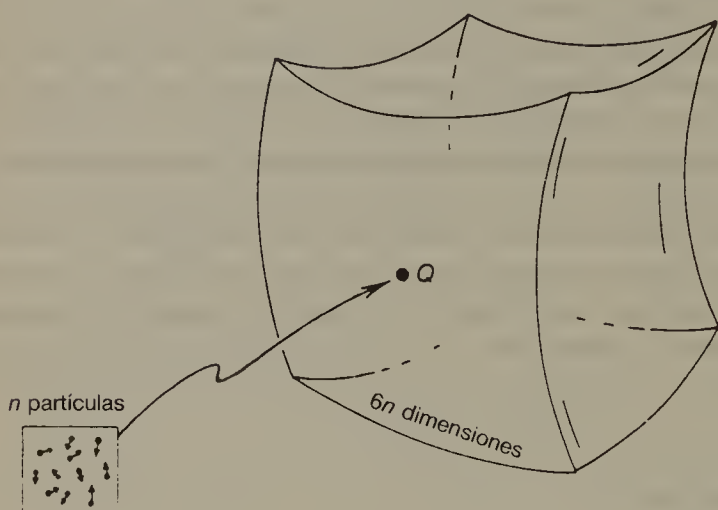


Fig. 5.10. Espacio de fases. Un simple punto Q del espacio de fases representa el estado global de algún sistema físico, incluyendo los movimientos instantáneos de cada una de sus partes.

Ahora bien, ¿cómo vamos a visualizar las ecuaciones de Hamilton en términos del espacio de fases? En primer lugar, tenemos que tener en cuenta lo que realmente representa un *simple punto* Q en el espacio de fases. Corresponde a un conjunto concreto de valores para todas las coordenadas de posición $x_1, x_2, \dots$ y para todas las coordenadas de momento $p_1, p_2, \dots$. Es decir, ' Q representa nuestro *sistema físico completo*, con un estado de movimiento particular especificado para cada una de sus partículas simples constituyentes. Las ecuaciones

de Hamilton nos dicen cuáles son los ritmos de cambio de todas estas coordenadas, una vez que sabemos sus valores presentes; es decir, gobiernan cómo van a moverse todas las partículas individuales. Traducido en lenguaje del espacio de fases, las ecuaciones nos están diciendo cómo debe moverse un simple punto Q del espacio de fases, dada la localización actual de Q en dicho espacio de fases. Así, en cada punto del espacio de fases tenemos una pequeña flecha —más correctamente, un *vector*— que nos dice cómo se está moviendo Q , para describir la evolución en el tiempo de nuestro sistema completo. La disposición global de flechas constituye lo que se conoce como un «campo vectorial» (Fig. 5.11). Por consiguiente, las ecuaciones de Hamilton definen un campo vectorial en el espacio de fases.

Veamos cómo debe interpretarse el *determinismo* físico en términos del espacio de fases. Para datos iniciales en el instante $t = 0$ tendremos un conjunto particular de valores especificados para todas las coordenadas de posición y momento; es decir, tendremos una elección particular del punto Q en el espacio de fases. Para hallar la evolución del sistema en el tiempo seguimos simplemente las flechas. Así, la evolución global de nuestro sistema con el tiempo —no importa cuán complicado pueda ser este sistema— se describe en el espacio de fases como un simple punto que se mueve siguiendo las flechas particulares que encuentra. Podemos pensar que las flechas indican la «velocidad» de nuestro punto Q en el espacio de fases. Si

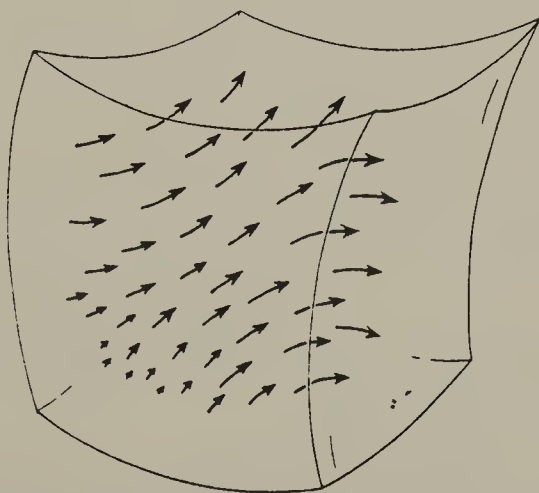


Fig. 5.11. Un campo de vectores del espacio de fases que representa la evolución temporal según las ecuaciones de Hamilton.

la flecha es «larga» Q se mueve rápidamente en su dirección, pero si la flecha es «corta» el movimiento de Q se hace más lento. Para ver lo que está haciendo nuestro sistema físico en el instante t , miramos simplemente hacia dónde se ha movido Q , en ese momento, siguiendo las flechas de esta manera. Evidentemente éste es un procedimiento determinista. La forma en que se mueve Q está completamente determinada por el campo vectorial hamiltoniano.

¿Qué sucede con la computabilidad? Si empezamos en un punto computable en el espacio de fases (esto es, en un punto en el que todas sus coordenadas de posición y momento son números computables, cf. Capítulo 3, p. 116) y esperamos a un tiempo computable t ¿terminaremos necesariamente en un punto que puede ser obtenido computablemente a partir de t y los valores de las coordenadas del punto de partida? La respuesta dependerá ciertamente de la elección de la función hamiltoniana H . De hecho, habrá *constantes físicas* que aparecen en H , como son la constante gravitatoria de Newton o la velocidad de la luz —cuyos valores exactos dependerán del sistema de unidades elegido, aunque otras podrían ser puros números— y sería necesario asegurar que estas constantes son *números computables* para tener esperanzas de obtener una respuesta afirmativa. Si suponemos que éste es el caso, la respuesta sería realmente afirmativa. Sin embargo, esto es simplemente una conjetura y la pregunta es una cuestión interesante que confío en que será examinada más a fondo en el futuro.

Por otra parte, tengo la impresión de que, por razones análogas a las que planteé hace poco en relación con el mundo de las bolas de billar, éste no es el resultado más relevante. Se requeriría una *precisión infinita* para las coordenadas de un punto del espacio de fases —es decir, *todas* las cifras decimales— para que tuviera sentido decir que el punto es no-computable. (Un número descrito con *finitos* decimales es siempre computable.) Una porción finita de una expansión decimal de un número no nos dice nada sobre la computabilidad de la expansión completa de dicho número. Pero todas las medidas físicas tienen un límite definido a la precisión con que pueden ser realizadas y sólo pueden dar información sobre un número finito de cifras decimales. ¿Anula esto el concepto global de «número computable» cuando se aplica a medidas físicas?

En realidad, un dispositivo que pudiera, de cualquier modo *útil*, sacar provecho de algún (hipotético) elemento no-computable en las leyes físicas no tendría probablemente que depender del hecho de hacer medidas de precisión ilimitada. Pero pudiera ser que esté adop-

tando aquí una postura demasiado estricta. Supongamos que tenemos un dispositivo físico que, por razones teóricas conocidas, imita algún interesante proceso matemático no-algorítmico. El comportamiento exacto del dispositivo, si este comportamiento pudiera ser exactamente comprobado, daría entonces las respuestas correctas a una serie de interesantes preguntas matemáticas del tipo sí/no para las que no puede haber un algoritmo (como los considerados en el Capítulo 5). Cualquier algoritmo *dado* fallaría en alguna etapa, y en *dicha* etapa, el dispositivo nos daría algo nuevo. El dispositivo podría implicar el examen de algún parámetro físico con una precisión cada vez mayor, en el que se necesitaría cada vez más precisión para ir cada vez más lejos en la lista de preguntas. Sin embargo, *no* obtenemos nada nuevo de nuestro dispositivo en una etapa *finita* en precisión, al menos hasta que encontremos un algoritmo mejorado para la secuencia de preguntas; entonces tendríamos que ir a una mayor precisión para poder conseguir algo que nuestro algoritmo *mejorado* no pudiera decirnos.

De todas formas, parecería aún que una precisión siempre creciente en un parámetro físico es una manera incómoda e insatisfactoria de codificar información. Mucho más preferible sería adquirir nuestra información en una forma *discreta* (o «digital»). Podrían conseguirse respuestas a preguntas cada vez más avanzadas en una lista examinando más unidades discretas cada vez, o quizá examinando un conjunto *fijo* de unidades discretas una y otra vez, en donde la ilimitada información requerida se extendería sobre intervalos de tiempo cada vez mayores. (Podemos imaginar estas unidades discretas constituidas por partes, cada una de ellas susceptible de un estado «sí» o «no», como los 0s y 1s de las descripciones de las máquinas de Turing dadas en el Capítulo 2.) Para esto parece que necesitamos dispositivos de algún tipo que puedan adoptar (distinguiblemente) estados discretos y que, después de evolucionar de acuerdo con las leyes dinámicas, adoptarán de nuevo un estado de un conjunto de estados discretos. Si así fuera podríamos evitar el tener que examinar cada dispositivo con una precisión arbitrariamente alta.

Ahora bien, ¿realmente se comportan de esta forma los sistemas hamiltonianos? Un ingrediente necesario sería algún tipo de estabilidad del comportamiento, de modo que pudiera comprobarse claramente en cuál de estos estados discretos está nuestro dispositivo. Una vez que está en uno de estos estados queremos que siga en él (al menos durante un intervalo de tiempo significativo) y no se deslice de uno de estos estados a otro. Además, si el sistema llega a estos estados con cierta imprecisión no queremos que estas imprecisiones

crezcan; antes bien, lo que realmente exigimos es que estas imprecisiones se *atenúen* con el tiempo. Ahora nuestro supuesto dispositivo tendría que estar constituido por partículas (u otras sub-unidades) que deben ser descritas en términos de parámetros continuos, y cada estado «discreto» distinguible tendrá que cubrir algún *rango* de estos parámetros continuos. (Por ejemplo, una posible manera de representar alternativas discretas sería tener una partícula que pueda estar en una caja o en otra. Para especificar que la partícula está realmente en una de las cajas necesitamos decir que las coordenadas de posición de la partícula están dentro de un cierto rango.) Lo que esto significa, en términos del espacio de fases, es que cada una de nuestras alternativas «discretas» debe corresponder a una *región* del espacio de fases, de modo que puntos diferentes del espacio de fases que están en la misma región corresponderán a la *misma* alternativa para nuestro dispositivo (Fig. 5.12).

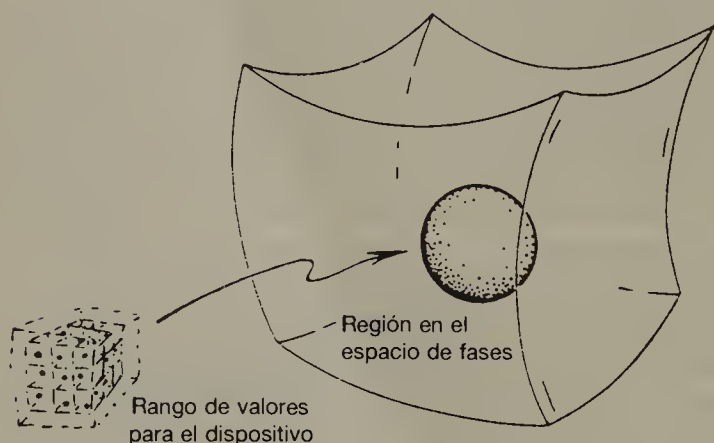


Fig. 5.12. Una *región* en el espacio de fases corresponde a un rango de posibles valores de las posiciones y momentos de todas las partículas. Una de estas regiones podría representar un estado distinguible (es decir, una «alternativa») de algún dispositivo.

Supongamos ahora que el dispositivo comienza con su punto en el espacio de fases dentro de alguna región R_0 que corresponde a una de estas alternativas. Pensamos en R_0 siendo arrastrada a lo largo del campo de vectores hamiltoniano a medida que transcurre el tiempo, hasta que, en el instante t , la región se ha convertido en R_t . Al representar esto estamos imaginando al mismo tiempo la evolución tem-

poral de nuestro sistema para *todos* los posibles estados de partida que corresponden a esta misma alternativa. (Ver Fig. 5.13.) El problema de la *estabilidad* (en el sentido que aquí nos interesa) consiste en si la región R_t permanece localizada, a medida que t crece, o si empieza a extenderse por el espacio de fases. Si estas regiones permanecen localizadas conforme avanza el tiempo, entonces tenemos una medida de estabilidad para nuestro sistema. Puntos del espacio de fases que están estrechamente próximos (y por lo tanto corresponden a estados físicos detallados del sistema que se parecen estrechamente entre sí) permanecerán estrechamente próximos en el espacio de fases, y las imprecisiones en su especificación no se amplificarán con el tiempo. Cualquier dispersión indebida implicaría una impredecibilidad efectiva en el comportamiento del sistema.

¿Qué se puede decir sobre los sistemas hamiltonianos en general? ¿Tienden a dispersarse con el tiempo las regiones en el espacio de fases, o no lo hacen? Parecería que muy poco se puede decir sobre un problema de tal generalidad. Resulta, sin embargo, que existe un teorema muy bello, debido al famoso matemático francés Joseph Liouville (1809-1882), que nos dice que el *volumen* de cualquier región del espacio de fases debe permanecer constante en cualquier evolución hamiltoniana. (Por supuesto, puesto que nuestro espacio de fases tiene una dimensión alta, éste tiene que ser un «volumen» en el sentido

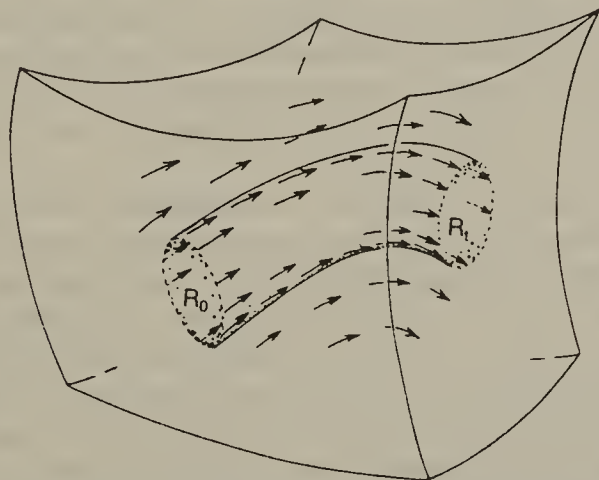


Fig. 5.13. A medida que pasa el tiempo, una región de estados de fase R^0 es arrastrada a lo largo del campo de vectores hacia una nueva región R_t . Esto representará la evolución temporal de una alternativa particular para nuestro dispositivo.

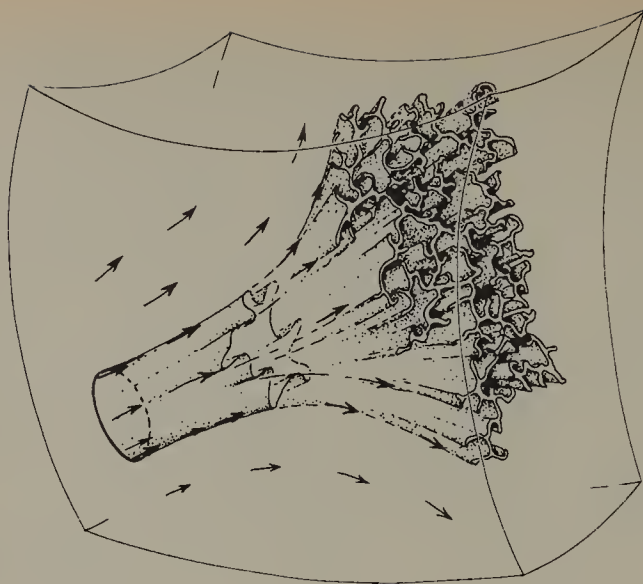


Fig. 5.14. Pese al hecho de que el teorema de Liouville nos dice que el volumen del espacio de fases no cambia durante la evolución temporal, normalmente este volumen se dispersará *efectivamente* debido a la extrema complejidad de su evolución.

apropiado para alta-dimensión). Por consiguiente, el volumen de cada R_i debe ser *igual* que el volumen de nuestra R_0 original. A primera vista esto parece responder afirmativamente a nuestra pregunta acerca de la estabilidad. En efecto, el *tamaño* —en el sentido de este volumen en el espacio de fases— de nuestra región *no puede crecer*, así que parece que nuestra región no puede dispersarse por el espacio de fases.

¡Sin embargo, esto es engañoso y reflexionando vemos que probablemente suceda el caso contrario! En la Figura 5.14 he tratado de indicar la clase de comportamiento que uno esperaría en general. Podemos imaginar que la región inicial R_0 es una región de forma «razonablemente» pequeña, más bien redondeada que espigada —indicando que los estados que pertenecen a R_0 pueden ser caracterizados de alguna manera que no requiere una precisión irrazonable. Sin embargo, a medida que transcurre el tiempo, la región R_i comienza a distorsionarse y estirarse— siendo al principio quizá como una ameba, pero estirándose luego hasta grandes distancias en el espacio de fases y contorsionándose de una forma muy complicada. El volumen sigue siendo el mismo pero este pequeño volumen puede haberse

extendido finalmente sobre enormes regiones del espacio de fases. Para una situación algo análoga piénsese en una pequeña gota de tinta colocada en un gran recipiente de agua. Aunque el volumen real del material que constituye la tinta permanece invariable, acabará por difundirse finalmente por todo el recipiente. En el caso del espacio de fases es probable que la región R_i se comporte de una manera semejante. Puede que no se extienda por la *totalidad* del espacio de fases (que es la situación extrema que se conoce como «ergódica») pero es probable que se extienda sobre una región enormemente mayor que la de partida. (Para discusiones adicionales, ver Davies, 1974.)

El problema es que la conservación del volumen no implica en absoluto conservación de *forma*: pequeñas regiones tenderán a distorsionarse, y esta distorsión les hace llegar a grandes distancias. El problema es mucho más grave en alta dimensión que en baja dimensión puesto que hay muchas más «direcciones» en las que la región puede estirarse localmente. ¡De hecho, lejos de ser una «ayuda» al mantener la región R_i bajo control, el teorema de Liouville nos coloca realmente ante un problema fundamental! Sin el teorema de Liouville podríamos suponer que esta indudable tendencia de una región a extenderse por el espacio de las fases podría quedar compensada, en circunstancias apropiadas, por una reducción del volumen global. Sin embargo, el teorema nos dice que esto es *imposible*, y que tenemos que afrontar esta sorprendente consecuencia —¡una característica universal de todos los sistemas dinámicos (hamiltonianos) clásico de tipo normal! ⁷

Podemos preguntar, en vista de esta difusión por todo el espacio de fases, ¿cómo es posible hacer alguna predicción en mecánica clásica? Esta es, en verdad, una buena pregunta. Lo que esta difusión nos dice es que, independientemente de la precisión con que conozcamos el estado inicial de un sistema (dentro de límites razonables), las imprecisiones tenderán a crecer con el tiempo y nuestra información inicial puede hacerse casi inútil. En este sentido, la mecánica clásica es esencialmente *impredecible*. (Recuérdese el concepto de «caos» considerado antes.)

¿Cómo es entonces que la mecánica clásica ha resultado ser tan acertada? En el caso de la mecánica celeste (esto es, el movimiento de los cuerpos celestes bajo la acción de la gravedad) las razones parecen ser, *en primer lugar*, que nos interesamos en un número relativamente pequeño de cuerpos coherentes (el Sol, los planetas y los satélites) que presentan grandes diferencias en cuanto a su masa —de modo que, en primera aproximación, podemos prescindir de los efectos perturbativos de los cuerpos de menor masa y tratar los mayores

como si fueran sólo unos *pocos* cuerpos que actúan bajo la influencia de los demás— y, *en segundo lugar*, que las leyes dinámicas que se aplican a las partículas individuales que constituyen estos cuerpos pueden verse operando también en el nivel de los propios cuerpos —¡de modo que, en una muy buena aproximación, el Sol, los planetas y los satélites pueden ser considerados como partículas y no tenemos que preocuparnos por todos los movimientos detallados de las partículas individuales que realmente componen estos cuerpos celestes! De nuevo salimos del paso considerando sólo unos «pocos» cuerpos, y la difusión en el espacio de fases no es importante.

Aparte de la mecánica celeste y del comportamiento de los proyectiles (que en realidad es simplemente un caso particular de la mecánica celeste), y del estudio de los sistemas simples que involucran un pequeño número de partículas, los principales modos de utilización de la mecánica newtoniana no parecen ser, ni mucho menos, de este tipo detallado «predictivo de forma determinista». Más bien, se utiliza el esquema general newtoniano para hacer modelos a partir de los cuales se pueden inferir propiedades globales de comportamiento. Algunas consecuencias precisas de las leyes, como la conservación de la energía, el momento y el momento angular tienen realmente importancia en todas las escalas. Además, existen propiedades estadísticas que pueden combinarse con las leyes dinámicas que gobiernan las partículas individuales y pueden utilizarse para hacer predicciones globales acerca del comportamiento. (Ver la discusión de la termodinámica en el Capítulo 7; el efecto de difusión en el espacio de fases que hemos estado discutiendo tiene una íntima relación con la segunda ley de la termodinámica, y teniendo el cuidado debido se pueden utilizar estas ideas de maneras verdaderamente predictivas.) El mismo cálculo notable de Newton para la velocidad del sonido en el aire (sutilmente corregido unos cien años más tarde por Laplace) fue un buen ejemplo de esto. Sin embargo, es ciertamente muy raro que se utilice realmente el determinismo inherente en la dinámica newtoniana (o, con más generalidad, hamiltoniana).

El efecto de difusión en el espacio de las fases tiene otra curiosa implicación. Nos dice, en efecto, que *¡la mecánica clásica no puede ser realmente verdadera en nuestro mundo!* Estoy exagerando algo esta implicación, pero quizá no tanto. La mecánica clásica puede dar cuenta perfectamente del comportamiento de los cuerpos fluidos —en particular de los gases, pero también de los líquidos en buena medida— en donde sólo nos interesan propiedades globales «promedio» de sistemas de partículas, pero tiene problemas para dar cuenta de la es-

estructura de los sólidos, en los que es necesaria una organización más estructurada y detallada. Existe un problema en cómo puede mantener su forma un sólido cuando está compuesto de miríadas de partículas puntuales cuya organización está reduciéndose continuamente debido a la difusión en el espacio de fases. Como ahora sabemos, se necesita la teoría cuántica para poder comprender adecuadamente la estructura real de los sólidos. Los efectos cuánticos pueden impedir de algún modo esta difusión en el espacio de fases. Éste es un tema importante al que volveré más tarde (ver Capítulos 8 y 9).

Este asunto tiene también importancia para la cuestión de la construcción de una «máquina computadora». La difusión en el espacio de fases es algo que debe ser controlado. No se debe permitir que una región del espacio de fases que corresponde a un estado «discreto» de un dispositivo de computación (tal como la R_0 antes descrita) se disperse indebidamente. Recordemos que incluso el «computador de bolas de billar» de Ferdkin-Toffoli requiere algunas *paredes sólidas* externas para que pueda funcionar. La «solidez» de un objeto que está compuesto de muchas partículas es algo que realmente tiene necesidad de la mecánica cuántica. ¡Parece que incluso una máquina de computar «clásica» debe apropiarse de los efectos de la física cuántica si tiene que funcionar eficazmente!

La teoría electromagnética de Maxwell

En la imagen newtoniana del mundo pensamos en minúsculas partículas que interaccionan mediante fuerzas que actúan a distancia —en donde las partículas, si no enteramente puntuales, pueden considerarse rebotando de vez en cuando unas en otras por contacto físico real. Como he dicho antes (p. 215), las fuerzas de la electricidad y el magnetismo (cuya existencia había sido conocida desde la antigüedad, y fueron estudiadas con algún detalle por William Gilbert en 1600 y Benjamin Franklin en 1752), actúan de un modo semejante a las fuerzas gravitatorias en que también decrecen de forma inversamente proporcional al cuadrado de la distancia, aunque son repulsivas en lugar de atractivas —es decir, actúan como si las partículas semejantes se repelieran— y es la carga eléctrica (y la intensidad del polo magnético), en lugar de la masa, la que mide la intensidad de la fuerza. En este nivel no hay dificultad en incorporar la electricidad y el magnetismo en el esquema newtoniano. También el comportamiento de la luz puede ser acomodado aproximadamente (aunque con ciertas difi-

cultades), bien considerando la luz como compuesta de partículas individuales («fotones», como se las llama ahora) o bien considerando la luz como un movimiento ondulatorio en algún medio, en cuyo caso este medio («éter») debe considerarse él mismo compuesto de partículas.

El hecho de que cargas eléctricas en movimiento puedan dar lugar a fuerzas magnéticas provocó alguna complicación adicional, pero no alteró el esquema general. Numerosos matemáticos y físicos (incluyendo a Gauss) habían propuesto sistemas de ecuaciones para los efectos de las cargas eléctricas en movimiento que habían parecido satisfactorias dentro del marco general newtoniano. El primer científico en desafiar seriamente la imagen «newtoniana» parece haber sido el gran teórico y experimentador inglés Michael Faraday (1791-1867).

Para comprender la naturaleza de este desafío debemos entender primero el concepto de *campo* físico. Consideremos un campo magnético en primer lugar. Muchos lectores se habrán tropezado con el comportamiento de las limaduras de hierro colocadas en una hoja de papel sobre un imán. Las limaduras se alinean de una manera sorprendente a lo largo de lo que se denominan «líneas de fuerza magnética». Imaginemos que las líneas de fuerza siguen estando aún presentes cuando las limaduras no estén ahí. Constituyen lo que se llama un *campo magnético*. En cada punto del espacio este «campo» está orientado en una cierta dirección, a saber, la dirección de la línea de fuerza que pasa por dicho punto. En realidad, lo que tenemos es un *vector* en cada punto de modo que el campo magnético nos proporciona un ejemplo de campo vectorial. (Podemos comparar esto con el campo de vectores hamiltoniano que consideramos en la sección anterior, pero ahora se trata de un vector en el espacio ordinario en lugar del espacio de fases.) Análogamente, un cuerpo eléctrico cargado estará rodeado por un tipo de campo diferente, conocido como un *campo eléctrico*, y un *campo gravitatorio* rodea del mismo modo cualquier cuerpo con masa. También éstos son campos de vectores en el espacio.

Tales ideas eran conocidas mucho tiempo antes de Faraday y habían llegado a ser una buena parte del arsenal de los teóricos en mecánica newtoniana. Pero el punto de vista prevaleciente no consideraba que tales «campos» constituyeran en sí mismos una sustancia física real. Más bien debería pensarse que proporcionaban el necesario «balance de cuentas» para las fuerzas que actuarían si se colocara una partícula adecuada en varios puntos diferentes. Sin embargo, los profundos hallazgos experimentales de Faraday (con conductores en

movimiento, imanes y similares) le llevaron a creer que los campos eléctricos y magnéticos son «substancia» física *real* y, además, que campos eléctricos y magnéticos variables podrían a veces ser capaces de «empujarse» a través del, aparte de ellos, espacio vacío ¡produciendo un tipo de onda incorpórea! Conjeturó que la propia luz podría consistir en dichas ondas. Semejante punto de vista estaría en desacuerdo con el «saber newtoniano» prevaleciente, en el que dichos campos no se consideraban «reales» en ningún sentido sino simplemente unos auxiliares matemáticos convenientes para la «verdadera» imagen newtoniana de la «realidad» como acción-a-distancia-entre-partículas-puntuales.

Confrontado con los hallazgos experimentales de Faraday, junto con algunos anteriores a cargo del notable físico francés André Marie Ampère (1775-1836) y otros, el gran físico y matemático escocés James Clerk Maxwell (1831-1879) se interrogó sobre la forma matemática de las ecuaciones que surgían de aquellos hallazgos. Con un notable golpe de intuición propuso un cambio en las ecuaciones —quizá pequeño en apariencia, pero fundamental en sus consecuencias. Este cambio no era sugerido en absoluto por los hechos experimentales conocidos (aunque era consistente con ellos); fue un resultado de las propias exigencias teóricas de Maxwell, en parte físicas, en parte matemáticas y en parte estéticas. Una consecuencia de las ecuaciones de Maxwell era que los campos eléctrico y magnético se «empujan» realmente a través del espacio vacío. Un campo magnético oscilante daría lugar a un campo eléctrico oscilante (esto estaba implícito en los descubrimientos experimentales de Faraday) y este campo eléctrico oscilante daría lugar, a su vez, a un campo magnético oscilante (por una inferencia teórica de Maxwell), y éste daría lugar de nuevo a un campo eléctrico y así sucesivamente. (Ver Fig. 6.26, 6.27 en pp. 342 y 343 para representaciones detalladas de estas ondas.) Maxwell pudo calcular la velocidad con la que se propagaría este efecto a través del espacio ¡y encontró que sería la velocidad de la luz! Estas llamadas ondas *electromagnéticas* mostrarían también la propiedad de interferencia y la enigmática propiedad de polarización de la luz que eran conocidas desde hacía mucho tiempo (volveremos a estas en el Capítulo 6, pp. 248, 342). Además de explicar las propiedades de la luz visible, para las que las ondas tendrían un rango concreto de longitudes de onda ($4-7 \times 10^{-7}$ m), se predecía la existencia de ondas electromagnéticas de otras longitudes de onda que se podrían producir mediante corrientes eléctricas en hilos conductores. La existencia de tales ondas fue establecida experimentalmente por Heinrich Hertz

en 1888. La inspirada esperanza de Faraday había puesto una base firme para las maravillosas ecuaciones de Maxwell.

Aunque no será necesario que apreciemos los detalles de las ecuaciones de Maxwell no será malo echarles una ojeada:

$$\frac{1}{c^2} \cdot \frac{\partial \mathbf{E}}{\partial t} = \text{rot } \mathbf{B} - 4\pi \mathbf{j}, \quad \frac{\partial \mathbf{B}}{\partial t} = - \text{rot } \mathbf{E}$$

$$\text{div } \mathbf{E} = 4\pi \rho, \quad \text{div } \mathbf{B} = 0$$

Aquí, $\mathbf{E}$, $\mathbf{B}$ y $\mathbf{j}$ son campos de vectores que describen el campo eléctrico, el campo magnético y la corriente eléctrica, respectivamente; ρ describe la densidad de carga eléctrica, y c es simplemente una constante: la velocidad de la luz. No hay que preocuparse por los términos «rot» y «div» que simplemente se refieren a diferentes tipos de variación espacial. (Son ciertas combinaciones de los operadores de derivación parcial respecto a las coordenadas espaciales. Recuérdese la operación «derivada parcial», simbolizada por ∂ , que ya encontramos a propósito de las ecuaciones de Hamilton.) Los operadores $\partial/\partial t$ que aparecen en el primer miembro de las dos primeras ecuaciones son, de hecho, iguales que el «punto» que utilizábamos en las ecuaciones de Hamilton, la diferencia es sólo técnica. Así, $\partial \mathbf{E}/\partial t$ significa «el ritmo de variación del campo eléctrico» y $\partial \mathbf{B}/\partial t$ significa «el ritmo de variación del campo magnético». La primera ecuación * dice cómo está variando el campo eléctrico con el tiempo, en función de lo que hacen el campo magnético y la corriente eléctrica en ese momento; mientras que la segunda ecuación dice cómo está variando el campo magnético con el tiempo en función de lo que está haciendo el campo eléctrico en ese momento. La tercera ecuación es, hablando en términos generales, una forma codificada de la ley de la inversa del cuadrado que nos dice cómo debe relacionarse (en cada instante) el campo eléctrico con la distribución de cargas; mientras que la cuarta ecuación dice lo que sería lo equivalente para el campo magnético, salvo que en este caso no existen «cargas magnéticas» (partículas aisladas «polo Norte» o «polo Sur»).

Estas ecuaciones son semejantes a las de Hamilton en cuanto que dicen cuál debe ser el ritmo de variación con el tiempo de las canti-

* Fue la presencia de $\partial \mathbf{E}/\partial t$ en esta ecuación la que constituyó el golpe maestro de la inferencia teórica de Maxwell. Todos los demás términos en las ecuaciones ya eran conocidos, de hecho, por evidencia experimental directa. El coeficiente $1/c^2$ es muy pequeño y por esta razón no se había observado experimentalmente aquel término.

dades relevantes (aquí los campos eléctrico y magnético) en función de cuáles son sus valores en cualquier instante dado. Por lo tanto, las ecuaciones de Maxwell son deterministas, igual que lo son las teorías hamiltonianas ordinarias. La única diferencia —y es una diferencia importante— es que las ecuaciones de Maxwell son ecuaciones de *campo* en lugar de ecuaciones para partículas, lo que significa que se necesita un *infinito* número de parámetros para describir el estado del sistema (los vectores del campo en cada punto del espacio), en lugar del número simplemente finito que se necesita en una teoría de partículas (tres coordenadas de posición y tres de momento para cada partícula). Por consiguiente, el espacio de fases para la teoría de Maxwell ¡es un espacio de un *infinito* número de dimensiones! (Como mencioné antes, las ecuaciones de Maxwell pueden ser englobadas en el marco general hamiltoniano, pero este marco debe ser ligeramente ampliado debido a esta dimensionalidad infinita.)

Además de lo que se ha visto previamente, el *nuevo* ingrediente fundamental para nuestra imagen de la realidad física que presenta la teoría de Maxwell consiste en que ahora los *campos* deben tomarse en serio por propio derecho y no pueden considerarse como simples apéndices matemáticos a lo que son las partículas «reales» de la teoría newtoniana. En realidad, Maxwell demostró que cuando los campos se propagan como ondas electromagnéticas transportan con ellos cantidades definidas de *energía*. Él pudo proporcionar una expresión explícita para esta energía. El hecho notable de que la energía pueda transportarse realmente de un lugar a otro por estas ondas electromagnéticas «incorpóreas» fue, en efecto, confirmado experimentalmente mediante la detección de tales ondas por Hertz. ¡Ahora es algo familiar para nosotros —aunque todavía un hecho bastante chocante— que las ondas de radio *pueden* transportar realmente energía!

Computabilidad y ecuación de ondas

Maxwell fue capaz de deducir directamente de sus ecuaciones que, en regiones del espacio en donde no hay cargas o corrientes (es decir, si $\mathbf{j} = 0$, $\rho = 0$ en las ecuaciones anteriores), todas las componentes de los campos eléctrico y magnético deben satisfacer una ecuación conocida como la *ecuación de ondas* *. La ecuación de ondas puede

* La ecuación de ondas (o ecuación de D'Alembert) puede escribirse $\{(1/c^2)(\partial/\partial t)^2 - (\partial/\partial x)^2 - (\partial/\partial y)^2 - (\partial/\partial z)^2\}\varphi = 0$.

considerarse como una «versión simplificada» de las ecuaciones de Maxwell puesto que es una ecuación para una *sola* cantidad en lugar de las seis componentes de los campos eléctrico y magnético. Su solución ejemplifica el comportamiento ondulatorio sin complicaciones añadidas, tales como la «polarización» de la teoría de Maxwell (dirección del vector del campo eléctrico, ver p. 342).

La ecuación de ondas tiene un interés adicional para nosotros debido a que ha sido estudiada explícitamente en relación con sus propiedades de *computabilidad*. En efecto, Marian Boykan Pour-El e Ian Richards han podido demostrar que incluso si las soluciones de la ecuación de ondas se comportan de modo *determinista* en el sentido ordinario —es decir, los datos proporcionados para un instante inicial dado determinarán la solución para cualquier otro instante— existen datos iniciales *computables*, de cierto tipo «peculiar», con la propiedad de que, para un tiempo posterior computable, el valor determinado del campo resulta ser *no computable*. Por ello, las ecuaciones de una teoría de campos plausible (aunque no exactamente la teoría de Maxwell que es realmente válida en nuestro mundo) puede dar lugar, en el sentido de Pour-El y Richards, ¡a una evolución no-computable!

A primera vista éste es un resultado bastante sorprendente —y parece contradecir lo que he estado conjeturando en la última sección acerca de la probable computabilidad de sistemas hamiltonianos «razonables». Sin embargo, aunque el resultado de Pour-El-Richards es ciertamente sorprendente y matemáticamente relevante, no contradice realmente dicha conjetura de una manera que tenga sentido físico aceptable. La razón es que su tipo «peculiar» de datos iniciales no es de «variación suave»¹⁰ como se requiere normalmente para un campo físico razonable. En realidad, Pour-El y Richards demuestran que la no-computabilidad *no puede aparecer* para la ecuación de ondas si rechazamos este tipo de campos. En cualquier caso, incluso si se permitieran campos de este tipo, sería difícil ver cómo un «dispositivo» físico (¿quizá un cerebro humano?) pudiera hacer uso de tal «no-computabilidad». Solamente podría tener relevancia cuando se admitieran medidas de precisión arbitrariamente alta, lo que, como describí antes, no es físicamente muy realista. De todas formas, el resultado de Pour-El-Richards supone un comienzo intrigante para un importante área de investigación en la que se ha trabajado poco hasta ahora.

La ecuación de movimiento de Lorentz; partículas desbocadas

Las ecuaciones de Maxwell, tal como están, no son suficientemente completas como sistema de ecuaciones. Ellas proporcionan una maravillosa descripción del modo en que se propagan los campos eléctrico y magnético una vez *dadas* las distribuciones de cargas y corrientes eléctricas. Estas cargas nos vienen dadas físicamente en forma de *partículas cargadas* —principalmente electrones y protones, como ahora sabemos— y las corrientes se deben al movimiento de dichas partículas. Si sabemos dónde están estas partículas y cómo se están moviendo, entonces las ecuaciones de Maxwell nos dicen cómo se comportará el campo electromagnético. Lo que *no* nos dicen las ecuaciones de Maxwell es cómo se comportarán las propias partículas. Una respuesta parcial a esta pregunta era conocida en los días de Maxwell, pero no se había establecido un sistema satisfactorio de ecuaciones hasta que, en 1895, el famoso físico holandés Hendrick Antoon Lorentz utilizó ideas próximas a las de la teoría de la relatividad para derivar las que ahora se conocen como *ecuaciones de movimiento de Lorentz* para una partícula cargada (cf. Whittaker 1910, pp. 310, 395). Estas ecuaciones nos dicen cómo varía continuamente la velocidad de una partícula cargada debido a los campos eléctrico y magnético en el punto en que está localizada la partícula. Cuando se añaden las ecuaciones de Lorentz a las de Maxwell se obtienen reglas para la evolución temporal *tanto* de las partículas cargadas *como* del campo electromagnético.

Sin embargo, no todo marcha bien con este sistema de ecuaciones. Ellas proporcionan excelentes resultados si los campos son muy uniformes hasta la escala del tamaño de las propias partículas (tomando este tamaño como el «radio clásico» del electrón, alrededor de 10^{-15} m), y siempre que los movimientos de las partículas no sean demasiado violentos. Sin embargo, existe aquí una dificultad *de principio* que puede ser importante en otras circunstancias. Lo que las ecuaciones de Lorentz nos dicen que debemos hacer es examinar el campo electromagnético en el punto *exacto* en el que está localizada la partícula (y, en efecto, nos proporcionan una «fuerza» en dicho punto). ¿Qué punto debe tomarse si la partícula tiene un tamaño finito? ¿Debemos tomar el «centro» de la partícula, o bien debemos promediar el campo (de la «fuerza») sobre todos los puntos de la superficie? Esto podría suponer una diferencia si el campo *no* es uniforme a la escala de la partícula. Hay otro problema más serio: ¿cuál es realmente el campo en la superficie de la partícula (o en el centro)?

Recuérdese que estamos considerando una partícula *cargada*. Existirá un campo electromagnético *debido a la propia partícula* que debe añadirse al «campo de fondo» en el que está situada la partícula. El campo de la propia partícula se hace enormemente intenso muy cerca de su «superficie» y supera con mucho a todos los demás campos en su vecindad. Además, el campo de la partícula apuntará más o menos directamente hacia fuera (o hacia dentro) en cada punto, de modo que el campo *real* resultante, al que se supone que está respondiendo la partícula, no será ni mucho menos uniforme sino que apuntará en direcciones diferentes en diferentes lugares de la «superficie» de la partícula, y ya no digamos en su «interior» (Fig. 5.15). Ahora tenemos que empezar a preocuparnos de si las diferentes fuerzas sobre la partícula tenderán a girarla o distorsionarla, y debemos preguntarnos qué propiedades elásticas tiene, etc. (y hay aquí cuestiones especialmente problemáticas en relación con la *relatividad* con las que no abrumaré al lector). Evidentemente el problema es mucho más complicado de lo que parecía ser antes.

Quizá sea mejor considerar la partícula como una partícula *puntual*. Pero esto lleva a otro tipo de problemas pues entonces el propio campo eléctrico de la partícula se hace *infinito* en su inmediata vecin-

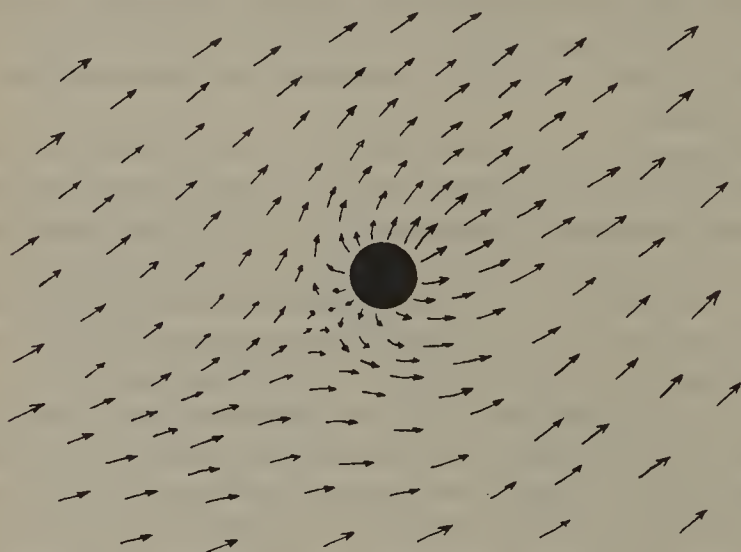


Fig. 5.15. ¿Cómo aplicamos estrictamente las ecuaciones de movimiento de Lorentz? Las fuerzas sobre una partícula cargada no pueden obtenerse simplemente examinando el campo «en» la localización de la partícula ya que allí es dominante el propio campo de la partícula.

dad. Si, según las ecuaciones de Lorentz, la partícula tiene que responder al campo electromagnético en el que está inmersa, entonces ¡parece que tendría que responder a un campo infinito! Para que la ley de la fuerza de Lorentz tenga sentido se necesita una manera de *restar el propio campo de la partícula* de modo que sólo quede un campo de fondo finito al que la partícula puede responder de modo inequívoco. El problema de cómo hacer esto fue resuelto en 1938 por Dirac (de quien volveremos a tener noticia más tarde). Sin embargo, la solución de Dirac le llevó a algunas conclusiones alarmantes. Encontró que para que el comportamiento de las partículas y campos esté determinado por sus datos iniciales es necesario que sean conocidas no sólo la posición y velocidad inicial de cada partícula sino que también debe ser conocida su *aceleración* inicial (una situación algo anómala en el contexto de las teorías dinámicas estándar). Para la mayoría de los valores de esta aceleración inicial la partícula se comporta finalmente de un modo totalmente disparatado, acelerándose espontáneamente hacia una velocidad ¡que se aproxima muy rápidamente a la de la luz! Éstas son las «soluciones desbocadas» de Dirac y no corresponden a nada de lo que realmente sucede en la Naturaleza. Debemos encontrar un modo de descartar las soluciones desbocadas escogiendo las aceleraciones iniciales del modo correcto. Esto puede hacerse siempre, pero sólo si aplicamos «presciencia» —es decir, debemos especificar las aceleraciones iniciales de una manera que anticipe qué soluciones se harán finalmente soluciones desbocadas y evitarlas. No es así ni mucho menos como se especifican los datos iniciales en un problema físico determinista estándar. Con el determinismo convencional aquellos datos pueden darse arbitrariamente, libres de cualquier requisito sobre cómo vaya a ser el comportamiento futuro. Aquí, no sólo está el futuro completamente determinado por los datos que pueden especificarse en un instante del pasado sino que ¡la propia especificación de dichos datos está limitada de forma muy precisa por el requisito de que el comportamiento futuro sea «razonable»!

Esto es todo lo lejos que podemos llegar con las ecuaciones fundamentales clásicas. El lector se habrá dado cuenta que el tema del determinismo y la computabilidad en las leyes de la física clásica se ha hecho preocupantemente confuso. ¿Tenemos realmente un elemento *teleológico* en las leyes físicas, en donde el futuro influye de algún modo sobre lo que está permitido que suceda en el pasado? Ciertamente los físicos no aceptan normalmente estas implicaciones de la *electrodinámica clásica* (la teoría de las partículas cargadas clásicas).

sicas, y los campos eléctrico y magnético) como descripciones serias de la realidad. Su respuesta normal a las dificultades anteriores consiste en decir que con partículas cargadas individuales se está propiamente en el dominio de la *electrodinámica cuántica*, y no se puede confiar en obtener respuestas razonables utilizando un procedimiento estrictamente clásico. Esto es indudablemente cierto pero, como veremos más adelante, la *propia* teoría cuántica tiene problemas a este respecto. De hecho, Dirac había considerado el problema clásico de la dinámica de una partícula cargada precisamente *porque* pensaba que podría dar intuiciones para resolver las dificultades fundamentales aún mayores del (más apropiado físicamente) problema cuántico. ¡Más adelante debemos enfrentarnos a los problemas de la teoría *cuántica*!

La relatividad especial de Einstein y Poincaré

Recordemos el principio de relatividad galileana que nos dice que las leyes físicas de Galileo y Newton permanecen totalmente invariantes si pasamos de un sistema de referencia en reposo a un sistema en movimiento. Esto implica que no podemos distinguir, mediante un simple examen del comportamiento dinámico de los objetos próximos a nosotros, si estamos en reposo o si nos estamos moviendo con velocidad uniforme en alguna dirección. (Recuérdese el barco en el mar de Galileo, p. 213.) Pero supongamos que añadimos las leyes de Maxwell a estas leyes: ¿sigue siendo cierta la relatividad galileana? Recuérdese que las ondas electromagnéticas de Maxwell se propagan a una velocidad fija c , la velocidad de la luz. El sentido común parece decirnos que si estuviéramos viajando muy rápidamente en una dirección, entonces la velocidad de la luz en dicha dirección debería aparecerse *reducida* por debajo de c (debido a que estamos moviéndonos para «alcanzar a la luz» en dicha dirección) y la aparente velocidad de la luz en la dirección contraria debería en consecuencia estar *incrementada* por encima de c (debido a que nos estamos alejando de la luz) —que es diferente del valor *fijo* c de la teoría de Maxwell. De hecho, el sentido común tendría razón: las ecuaciones de Newton y Maxwell combinadas no *satisfacen* la relatividad galileana.

Fue la preocupación por estos temas la que llevó a Einstein, en 1905 —como, en efecto había llevado a Poincaré antes de él (en 1898-1905)— a la teoría de la relatividad especial. Poincaré y Einstein descubrieron independientemente que las ecuaciones de Maxwell *también* satisfacen un principio de relatividad (cf. Pais, 1982); es decir,

las ecuaciones tienen una propiedad análoga a la de permanecer invariantes si pasamos de un sistema de referencia en reposo a uno en movimiento, ¡aunque las reglas para esto son *incompatibles* con las de la física de Galileo-Newton! Para hacer ambas compatibles sería necesario modificar uno u otro conjunto de ecuaciones —o abandonar el principio de relatividad.

Einstein no tenía intención de abandonar el principio de relatividad. Su soberbio instinto físico era insistente en que este principio debe ser válido para las leyes físicas de nuestro mundo. Además, él sabía que, prácticamente en todos los fenómenos conocidos, la física de Galileo-Newton había sido verificada sólo para velocidades muy pequeñas comparadas con la velocidad de la luz, donde esta incompatibilidad no era significativa. La única cosa que se conocía que involucrara velocidades suficientemente grandes para que tales discrepancias fueran importantes era la *propia luz*. Sería, por lo tanto, el comportamiento de la luz el que nos informara sobre qué principio de relatividad debíamos adoptar —y las ecuaciones que gobiernan la luz son las de Maxwell. Por lo tanto el principio de la relatividad de la teoría de Maxwell era el principio a mantener; y, consiguientemente, ¡deben modificarse las leyes de Galileo-Newton!

Lorentz, antes que Poincaré y Einstein, también había abordado y parcialmente respondido estas cuestiones. Hacia 1895 Lorentz había aceptado la idea de que las fuerzas que ligán la materia eran de naturaleza electromagnética (como realmente resultaron ser) de modo que el comportamiento de los cuerpos materiales reales debería satisfacer leyes derivadas de las ecuaciones de Maxwell. Una consecuencia de ello resultaba ser que un cuerpo que se mueva con una velocidad comparable a la de la luz se contraerá ligeramente en la dirección del movimiento (la «contracción de Fitzgerald-Lorentz»). Lorentz había utilizado esto para explicar un enigmático descubrimiento experimental, el de Michelson y Morley en 1887, que parecía indicar que los fenómenos electromagnéticos no pueden ser utilizados para determinar un sistema de referencia «absoluto». (Michelson y Morley demostraron que, contrariamente a lo que se esperaba, la velocidad aparente de la luz en la superficie de la Tierra no está influenciada por el movimiento de la Tierra alrededor del Sol.) ¿Siempre se comporta la materia de modo que no pueda detectarse localmente su movimiento (uniforme)? Ésta fue la conclusión *aproximada* de Lorentz; además él se limitaba a una concreta teoría de la materia, en la que ninguna fuerza distinta de las electromagnéticas se consideraba significativa. Poincaré, siendo como era un matemático sobresaliente, pudo demos-

trar (en 1905) que la materia debe comportarse de una forma *precisa*, según el principio de relatividad subyacente a las ecuaciones de Maxwell, para que el movimiento uniforme no pueda ser detectado localmente. También obtuvo una buena comprensión de las implicaciones físicas de este principio (incluyendo la «relatividad de la simultaneidad» que consideraremos dentro de poco). Parece haberla considerado como sólo *una* posibilidad, y no compartía la convicción de Einstein de que algún principio de relatividad debe ser *válido*.

El principio de relatividad que satisfacen las ecuaciones de Maxwell —que ha llegado a ser conocido como *relatividad especial*— es algo difícil de captar, y tiene muchas características no intuitivas que son difíciles de aceptar, al principio, como propiedades reales del mundo en que vivimos. De hecho, la relatividad especial no puede entenderse propiamente sin el ingrediente *adicional*, introducido en 1908 por el altamente original e intuitivo geómetra ruso/germano Hermann Minkowski (1864-1909). Minkowski había sido uno de los profesores de Einstein en el Instituto Politécnico de Zurich. Su idea nueva fundamental era que el espacio y el tiempo tenían que ser considerados en conjunto como una sólo entidad: un *espacio-tiempo tetra-dimensional*. En 1908 Minkowski anunció en una famosa conferencia en la universidad de Göttingen:

En lo sucesivo el espacio por sí mismo, y el tiempo por sí mismo, están condenados a desvanecerse en meras sombras, y sólo una especie de fusión entre los dos mantendrá una realidad independiente.

Tratemos de comprender las bases de la relatividad especial en términos de este magnífico espacio-tiempo de Minkowski.

Una de las dificultades para comprender el concepto de espacio-tiempo es que éste es *tetra-dimensional*, lo que lo hace difícil de visualizar. Sin embargo, ¡después de haber sobrevivido a nuestro encuentro con el espacio de fases no tendremos ningún problema con unas simples cuatro dimensiones! Como antes, haremos «trampa» y representaremos un espacio de dimensión menor —pero ahora el grado de la trampa es incomparablemente menos grave y, correspondientemente, nuestra representación será más aproximada. Dos dimensiones (una espacial y una temporal) bastarán para muchos propósitos, pero espero que el lector me permitirá ser un poco más temerario y subir a tres (dos espaciales y una temporal). Esto proporcionará una representación muy buena y no será difícil aceptar que, en principio, las ideas se extiendan, sin muchos cambios, a la completa situación

tetra-dimensional. Lo que hay que tener en mente, a propósito de un diagrama espacio-tiempo, es que cada punto en la imagen representa un *suceso* —es decir, un punto en el espacio en un simple momento, un punto que tiene sólo una existencia *instantánea*. El diagrama completo representa toda la historia: pasado, presente y futuro. Una partícula, en tanto que persiste en el tiempo, viene representada no por un punto sino por una línea, llamada la *línea-de-universo*. Esta línea-de-universo —recta si la partícula se mueve uniformemente y curva si se acelera (es decir, se mueve *no-uniformemente*)— describe toda la historia de la existencia de la partícula.

En la Figura 5.16 se ha representado un espacio-tiempo con dos dimensiones espaciales y una temporal. Imaginamos que hay una coordenada-temporal estándar t , medida en la dirección vertical, y dos coordenadas-espaciales x/c y z/c , medidas horizontalmente *. El cono en el centro es el *cono de luz* (futuro) del origen espacio-temporal O . Para darnos cuenta de su significado imaginemos que una explosión tiene lugar en el suceso O . (Por lo tanto, la explosión ocurre en el origen espacial en el instante $t = 0$.) La historia de la luz emitida por la explosión es este cono de luz. En términos de dos dimensiones espaciales, la historia del destello de luz sería un círculo que se mueve hacia arriba con la velocidad fundamental de la luz c . En el espacio completo tri-dimensional sería una *esfera* que se expande a velocidad c —el frente de onda esférico de la luz— pero aquí estamos *suprimiendo* la dirección espacial y , de modo que sólo tenemos un círculo, como las ondas circulares que emanan del punto de caída de una piedra en un estanque. Podemos ver este círculo en la imagen espacio-temporal si hacemos cortes horizontales sucesivos a través del cono que se mueve uniformemente hacia arriba. Estos planos horizontales representan distintas descripciones espaciales a medida que la coordenada-temporal t crece. Ahora bien, una de las características de la teoría de la relatividad es que es imposible para una partícula material viajar más rápido que la luz (más adelante diremos más sobre esto). Esto significa, en términos espacio-temporales, que las líneas-de-universo de todas las partículas emitidas en la explosión deben estar *dentro* del cono de luz.

Es, a menudo, más conveniente describir la luz en términos de *partículas* —llamadas *fotones*— que en términos de ondas electromag-

* La razón para dividir las coordenadas espaciales por c —la velocidad de la luz— es que las líneas-de-universo de los fotones tengan una inclinación conveniente: 45° respecto a la vertical; ver más adelante.

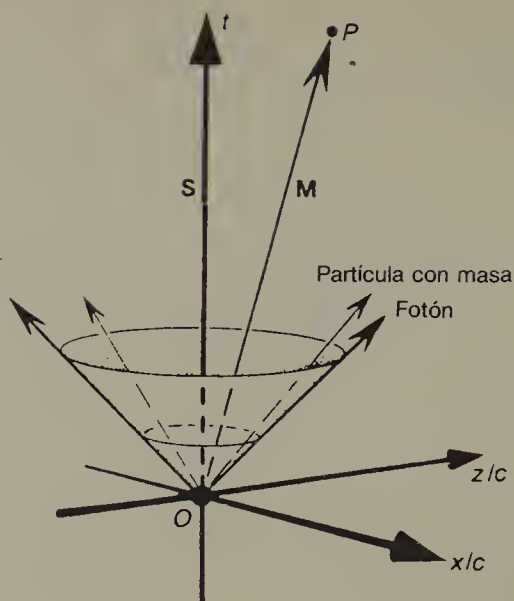


Fig. 5.16. Un cono de luz en el espacio-tiempo de Minkowski (con sólo dos dimensiones espaciales), que describe la historia del destello de una explosión que tiene lugar en el suceso O , origen del espacio-tiempo.

néticas. Por el momento podemos pensar en un «fotón» como un pequeño «paquete» de oscilaciones de alta frecuencia del campo electromagnético. El término es más apropiado físicamente en el contexto de las descripciones *cuánticas* que consideraremos en el próximo capítulo, pero los fotones «clásicos» también nos serán útiles aquí. En el espacio libre los fotones viajan siempre en línea recta con la velocidad fundamental c . Esto significa que en la imagen del espacio-tiempo de Minkowski la línea-de-universo de un fotón viene representada siempre como una línea recta inclinada a 45° de la vertical. Los fotones producidos en nuestra explosión en O describen el cono de luz centrado en O .

Estas propiedades deben ser válidas en general en todos los puntos del espacio-tiempo. No hay nada especial en el origen; el punto O no es diferente de ningún otro punto. Por lo tanto debe haber un cono de luz en cada punto del espacio-tiempo, con un significado idéntico al del cono de luz en el origen. La historia de cualquier flash de luz

—o las líneas-de-universo de los fotones, si preferimos utilizar una descripción de la luz como partículas— está siempre sobre la superficie del cono de luz en cada punto, mientras que la historia de cualquier partícula material debe estar siempre en el interior del cono de luz en cada punto. Esto se ilustra en la Figura 5.17. La familia de conos de luz en todos los puntos puede ser considerada como parte de la *geometría minkowskiana* del espacio-tiempo.

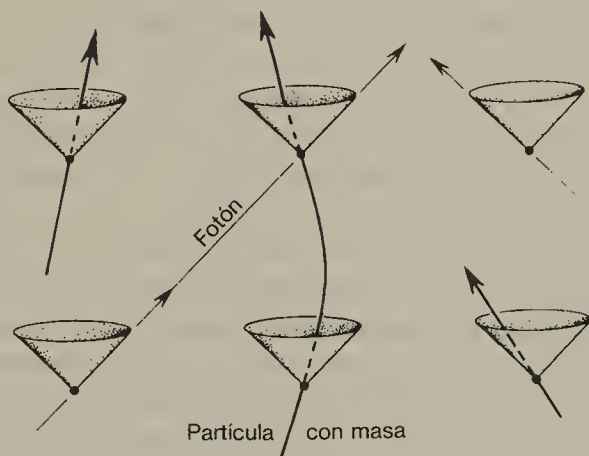


Fig. 5.17. Una imagen de la geometría minkowskiana.

¿Qué es la geometría minkowskiana? La estructura de conos-de-luz es su aspecto más importante, pero hay más que esto en la geometría minkowskiana. Existe un concepto de «distancia» que tiene algunas analogías notables con la distancia en la geometría euclídea. En la geometría euclídea tri-dimensional, la distancia r de un punto al origen, en términos de las coordenadas cartesianas estándar, está dada por

$$r^2 = x^2 + y^2 + z^2$$

(Ver Fig. 5.18a. Esto es precisamente el teorema de Pitágoras —siendo quizá más familiar el caso de dos dimensiones.) En nuestra geometría minkowskiana tri-dimensional la expresión es formalmente muy parecida (Fig. 5.18b), siendo la diferencia esencial que ahora tenemos dos *signos menos*:

$$s^2 = t^2 - (x/c)^2 - (z/c)^2.$$

Más correctamente, por supuesto, tendremos la geometría minkowskiana *tetra*-dimensional, y la expresión para la «distancia» es

$$s^2 = t^2 - (x/c)^2 - (y/c)^2 - (z/c)^2$$

¿Cuál es el significado físico de la «distancia» s en ésta expresión? Supongamos que el punto en cuestión —esto es, el punto P , de coordenadas $\{t, x/c, y/c, z/c\}$ (o $\{t, x/c, z/c\}$, en el caso tridimensional; ver Fig. 5.16)— está dentro del cono de luz (futuro) de O . Entonces el segmento de línea recta OP puede representar parte de la historia de alguna partícula material —por ejemplo una partícula emitida por nuestra explosión. La «longitud» minkowskiana s del segmento OP tiene una interpretación física directa: ¡es el *intervalo temporal* experimentado realmente por la partícula entre los sucesos O y P ! Dicho de otro modo, si hubiera un reloj robusto y preciso unido a la partícula ¹², entonces la diferencia entre los tiempos que registra en O y en P es precisamente s . En contra de lo que se podría esperar, la magnitud de la coordenada t *no* describe en sí misma el tiempo medido por un reloj preciso a menos que esté «en reposo» en nuestro sistema de coordenadas (es decir, con valores fijos $x/c, y/c, z/c$ de las coordenadas), lo que quiere decir que el reloj tendría una línea-de-universo «vertical» en el diagrama. Por lo tanto, « t » significa «tiempo» sólo para observadores «en reposo» (esto es, con líneas-de-universo «verticales»). La medida *correcta* del tiempo para un observador en

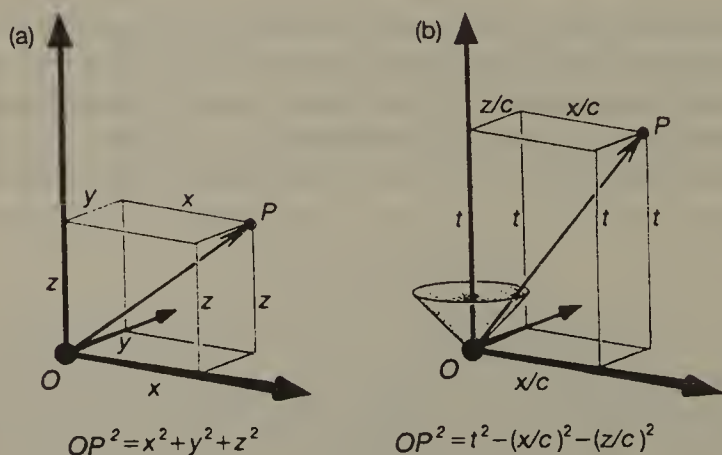


Fig. 5.18. Una comparación entre la «distancia» medida en (a) geometría euclídea y (b) geometría minkowskiana (donde «distancia» significa «tiempo experimentado»).

movimiento (que se aleja uniformemente del origen O), según la relatividad, viene proporcionada por la cantidad s .

Esto es muy curioso —y bastante distinto de la medida de tiempo galileo-newtoniana de «sentido común», que sería sencillamente el valor de la coordenada t . Nótese que la medida de tiempo relativista (minkowskiana) s es siempre algo *menor* que t si no hay movimiento en absoluto (puesto que s^2 es menor que t^2 , por la fórmula anterior, siempre que x/c , y/c , y z/c no sean todas nulas). El movimiento (es decir, si OP no está a lo largo del eje t) tenderá a «retardar», el reloj respecto a t —es decir, como se vería con respecto a nuestro sistema de coordenadas. Si la velocidad del movimiento es pequeña comparada con c , entonces s y t serán casi iguales, lo que explica por qué no somos conscientes directamente del hecho de que «los relojes en movimiento atrasan». En el otro extremo, cuando la velocidad es la de la misma luz, P está entonces *sobre* el cono de luz; y encontramos $s = 0$. El cono de luz es precisamente el conjunto de puntos cuya «distancia» minkowskiana (es decir, «tiempo») a O es realmente cero. Así, ¡un fotón no «experimentará» el paso del tiempo en absoluto! (No está permitido el caso aún *más* extremo, en el que P se mueve *fuera* del cono de luz, ya que ello conduciría a un s imaginario —la raíz cuadrada de un número negativo— y violaría la regla de que las partículas materiales o los fotones no pueden viajar más rápido que la luz) *.

Esta noción de «distancia» minkowskiana se aplica igualmente a *cualquier* par de puntos del espacio-tiempo tal que uno está dentro del cono de luz del otro —de modo que una partícula puede viajar de uno a otro. Consideremos simplemente que O se desplaza a algún otro punto del espacio-tiempo. De nuevo, la distancia minkowskiana entre los puntos mide el intervalo de tiempo que experimenta un reloj que se mueve uniformemente de uno a otro. Cuando la partícula es un fotón, y la distancia minkowskiana se hace cero, tenemos dos puntos uno de los cuales está *sobre* el cono de luz del otro —y este hecho sirve para *definir* el cono de luz de este punto.

La estructura básica de la geometría minkowskiana, con esta curiosa medida de «longitud» para las líneas de universo —interpretada como el *tiempo* medido (o «experimentado») por relojes físicos— contiene la esencia misma de la relatividad especial. Puede que el lector

* Sin embargo, para sucesos separados por valores negativos de s^2 , la cantidad $c\sqrt{(-s^2)}$ tiene un significado, a saber, el de distancia *ordinaria* —para aquel observador para quien los sucesos parecen simultáneos (cf. más adelante).

esté familiarizado, en particular, con lo que se conoce como «la paradoja de los gemelos» de la relatividad: uno de los hermanos gemelos permanece en la Tierra mientras que el otro hace un viaje a una estrella cercana, yendo y volviendo a una gran velocidad que se aproxima a la de la luz. A su regreso, se descubre que los dos hermanos han envejecido de manera diferente: el viajero se encuentra aún joven mientras que el hermano que se quedó en casa es un viejo. Esto se describe fácilmente en términos de la geometría de Minkowski —y se ve por qué, aunque es un fenómeno enigmático, no es realmente una paradoja. La línea de universo AC representa al gemelo que se queda en casa mientras que el viajero tiene una línea de universo compuesta de dos segmentos AB y BC que representan las etapas de ida y vuelta del viaje (ver Fig. 5.19). El gemelo que-queda-en-casa experimenta un tiempo medido por la distancia minkowskiana AC , mientras que el viajero experimenta un tiempo dado por la suma ¹³ de las dos

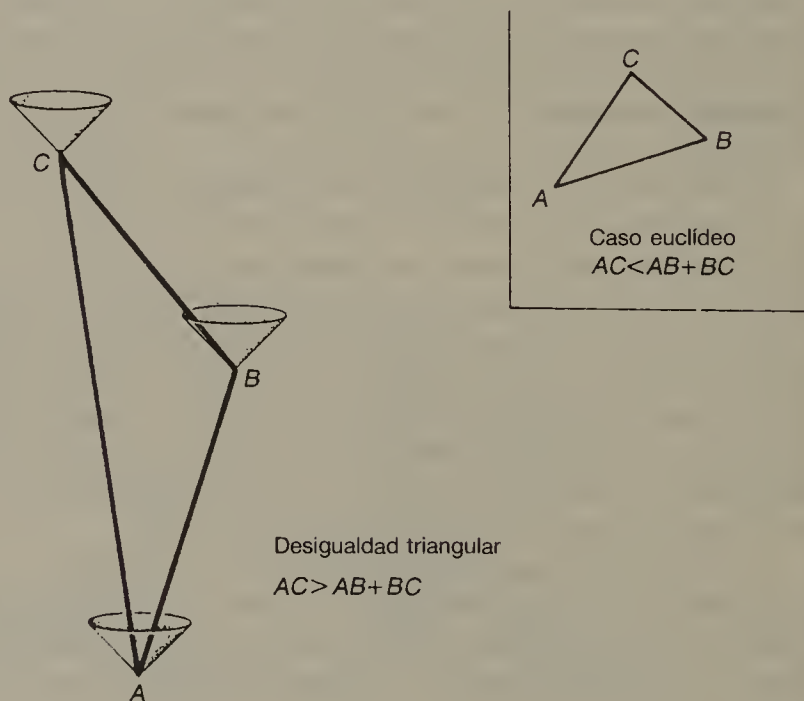


Fig. 5.19. La llamada «paradoja de los gemelos» de la relatividad especial se comprende en términos de una desigualdad triangular minkowskiana. (Se da también el caso euclídeo para comparación.)

distancias de Minkowski AB y BC . Estos tiempos no son iguales sino que encontramos

$$AC > AB + BC$$

lo que demuestra que realmente el tiempo experimentado por el que queda-en-casa es mayor que el del viajero.

La desigualdad superior se parece bastante a la bien conocida *desigualdad triangular* de la geometría euclídea ordinaria, a saber, (siendo ahora A , B y C tres puntos en el espacio *euclídeo*):

$$AC < AB + BC$$

que afirma que la suma de dos lados de un triángulo es siempre *mayor* que el tercero. ¡*Esto* no se considera una paradoja! Estamos perfectamente familiarizados con la idea de que la medida de la distancia euclídea a lo largo de un camino de un punto a otro (aquí de A a C) depende del camino concreto que tomemos. (En este caso, los dos caminos son AC y la ruta más larga y quebrada ABC .) Este ejemplo es un caso particular del hecho de que la distancia más corta entre dos puntos (aquí A y C) es la medida a lo largo de la línea recta que los une (la línea AC). El sentido contrario en el signo de la desigualdad en el caso *minkowskiano* aparece debido a los cambios de signo en la definición de «distancia», de modo que la AC minkowskiana es «más larga» que la ruta combinada ABC . Así mismo, esta «desigualdad triangular» minkowskiana es un caso particular de un resultado más general: la *más larga* (en el sentido de que se experimenta un tiempo mayor) de las líneas-de-universo entre dos puntos es la recta (es decir, no acelerada). Si dos gemelos empiezan en el mismo suceso A y terminan en el mismo suceso C , en donde el primer gemelo se mueve directamente de A a C sin aceleraciones pero el segundo acelera, entonces el primero experimentará siempre un intervalo de tiempo mayor cuando se encuentren de nuevo.

Puede parecer extravagante el introducir un concepto tan extraño de medida-de-tiempo en abierto contraste con nuestras nociones intuitivas. Sin embargo, hay ahora una enorme cantidad de evidencia experimental en su favor. Por ejemplo, existen muchas partículas subatómicas que se desintegran (es decir, se deshacen en otras partículas) en una escala-de-tiempo definida. A veces tales partículas viajan a velocidades muy próximas a la de la luz (v.g. en los rayos cósmicos que llegan a la Tierra desde el espacio exterior, o en los aceleradores de partículas contruidos por el hombre), y sus tiempos

de desintegración se alargan en la medida exacta que se deduce a partir de las consideraciones anteriores. Más impresionante aún es el hecho de que ahora pueden hacerse relojes («relojes nucleares») tan precisos que estos efectos de retardo-temporal son detectables directamente mediante relojes transportados en aviones rápidos en vuelo bajo —en concordancia con la medida de «distancia» minkowskiana s , ¡y *no* con t ! (Estrictamente hablando, si tomamos en cuenta la *altitud* del avión aparecerán involucrados pequeños efectos gravitatorios adicionales de la relatividad *general*, pero éstos concuerdan también con las observaciones; ver la próxima sección.) Además, existen muchos otros efectos íntimamente relacionados con el marco global de la relatividad especial, que están recibiendo continuamente verificación detallada. Uno de estos, la famosa relación de Einstein

$$E = mc^2$$

que equipara efectivamente energía y masa, tendrá algunas implicaciones seductoras para nosotros al final de este capítulo.

No se ha explicado aún de qué modo se incorpora realmente el principio de relatividad en este esquema de cosas. ¿Cómo es que observadores que se mueven con velocidades uniformes diferentes pueden ser *equivalentes* con respecto a la geometría minkowskiana? ¿Cómo puede el eje temporal de la Figura 5.16 («observador en reposo») ser completamente equivalente a alguna otra línea-de-universo recta, digamos la recta OP extendida («observador en movimiento»)? Pensemos primero en la geometría *euclídea*. Evidentemente dos líneas rectas individuales cualesquiera son equivalentes con respecto a la geometría en general. Podemos imaginar que «deslizamos rígidamente sobre sí mismo» todo el espacio euclídeo hasta que una de las líneas rectas toma la posición de la otra. Pensemos en el caso *bi*-dimensional, un *plano* euclídeo. Podemos imaginar que movemos una hoja de papel rígidamente sobre una superficie plana de modo que una línea recta cualquiera dibujada en el papel llegue a coincidir con una línea recta dada en la superficie. Este movimiento rígido conserva la estructura de la geometría. Algo similar es válido para la geometría minkowskiana, aunque esto es menos obvio y tenemos que tener cuidado en lo que entendemos por «rígido». Ahora, en lugar de en una hoja de papel deslizante debemos pensar en algún material de un tipo peculiar —tomando, por simplicidad, el caso bi-dimensional— en el que las líneas inclinadas 45° siguen estando a 45° mientras que el material puede estirarse en una dirección a 45° y contraerse en la otra dirección a 45° . Esto se ilustra en la Figura 5.20. En la Figura 5.21

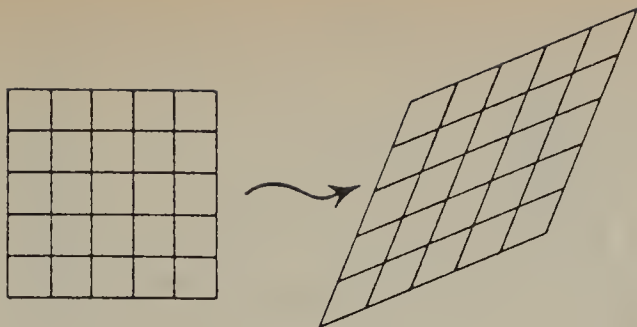


Fig. 5.20. Un movimiento de Poincaré en dos dimensiones espacio-temporales.

he tratado de indicar lo que está involucrado en el caso tri-dimensional. Este tipo de «movimiento rígido» del espacio de Minkowski —llamado un *movimiento de Poincaré* (o movimiento inhomogéneo de Lorentz) puede no parecer muy «rígido» pero conserva todas las distancias minkowskianas, y es precisamente «conservación de las distancias» lo que significa la palabra «rígido» en el caso euclídeo. El principio de relatividad espacial asegura que la física es invariante bajo tales movimientos de Poincaré del espacio-tiempo. En particular, el observador «en reposo» S , cuya línea-de-universo es el eje temporal de nuestra representación de Minkowski original Figura 5.16, tiene

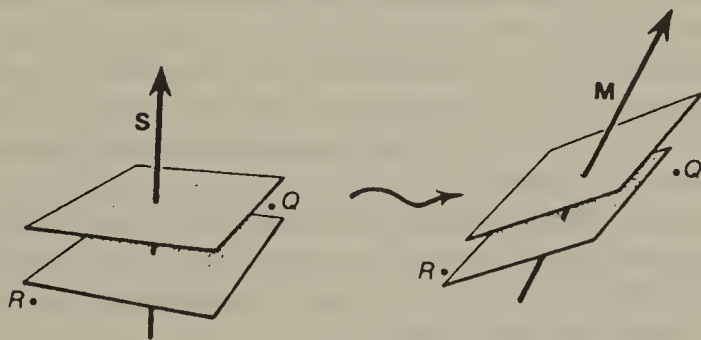


Fig. 5.21. Un movimiento de Poincaré en tres dimensiones espacio-temporales. El diagrama de la izquierda muestra espacios simultáneos para S , y el diagrama de la derecha, espacios simultáneos para M . Nótese que S piensa que R precede a Q , mientras que M piensa que Q precede a R . (El movimiento se considera aquí *pasivo*, es decir, que sólo afecta a las descripciones diferentes que los dos observadores S y M harían de un mismo espacio-tiempo.)

una física completamente equivalente a la del observador «en movimiento» M con una línea-de-universo a lo largo de OP .

Cada plano de coordenadas $t = \text{constante}$ representa el «espacio» en cada «instante» para el observador S , es decir, es una familia de sucesos que él consideraría *simultáneos* (esto es, que tienen lugar todos ellos en el «mismo instante»). Llamaremos a estos planos *espacios simultáneos* de S . Cuando pasamos a otro observador M debemos cambiar nuestra familia original de espacios simultáneos por una nueva familia mediante un movimiento de Poincaré que proporciona los espacios simultáneos para M ¹⁴. Nótese que los espacios simultáneos de M aparecen «inclinados» en la Figura 5.21. Podría parecer que esta inclinación está en la dirección errónea si pensamos en términos de movimientos rígidos en la geometría euclídea, pero es la que debemos esperar en el caso minkowskiano. Mientras S piensa que todos los sucesos de cualquier plano $t = \text{constante}$ ocurren simultáneamente, M tiene una visión diferente: ¡para él son los sucesos de cada uno de sus espacios simultáneos «inclinados» los que parecen ser simultáneos! La geometría minkowskiana no contiene, en sí misma, un único concepto de «simultaneidad» sino que cada observador en movimiento uniforme lleva con él su propia idea de lo que significa «simultáneo».

Consideremos los dos sucesos R y Q en la Figura 5.21. Según S , el suceso R tiene lugar antes que el suceso Q , debido a que R está en un espacio simultáneo anterior al de Q , pero según M es el revés, Q está en un espacio simultáneo anterior al de R . Así, para un observador el suceso R tiene lugar antes que Q , ¡pero para el otro es Q el que tiene lugar antes que R ! (Esto puede suceder sólo debido a que entre R y Q existe una *separación de tipo-espacio*, lo que significa que cada uno de ellos está fuera del cono de luz del otro, de modo que ninguna partícula material ni ningún fotón puede viajar de un suceso a otro.) Incluso con velocidades relativas pequeñas pueden ocurrir diferencias significativas en el orden temporal para sucesos a grandes distancias. Imaginemos que dos personas que caminan lentamente se cruzan en la calle. Los sucesos en la galaxia Andrómeda (la galaxia importante más cercana a nuestra propia Vía Láctea, a unos 20 000 000 000 000 000 000 kilómetros de distancia), que las dos personas juzgan simultáneos con el momento en que ellas se cruzan, podrían tener una diferencia de varios días (Fig. 5.22). Para una de las personas la flota espacial lanzada con el intento de acabar con la vida en el planeta Tierra está ya en camino; mientras que para la otra, la decisión sobre si enviar o no esa flota ¡todavía no se ha tomado!

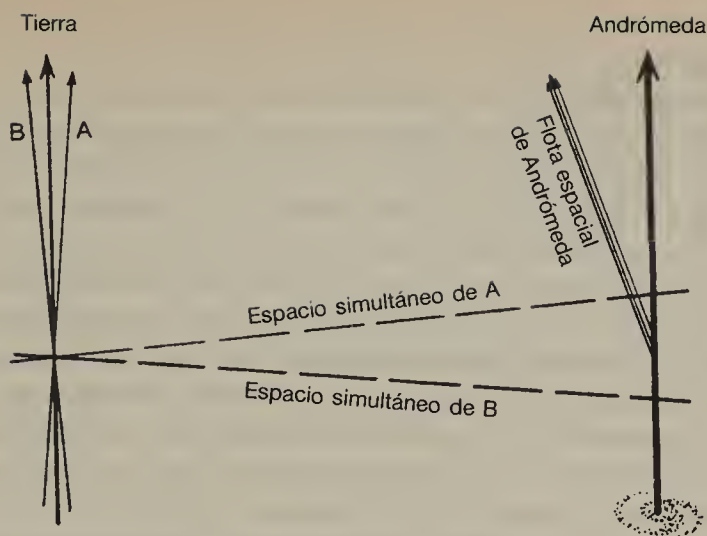


Fig. 5.22. Dos personas *A* y *B* pasean lentamente, pero tienen diferentes visiones acerca de si ya ha partido una flota espacial de Andrómeda en el momento en que ellos se cruzan.

La relatividad general de Einstein

Recuérdese la gran intuición de Galileo de que todos los cuerpos caen con la misma velocidad en un campo gravitatorio. (Fue una intuición y no precisamente una observación directa ya que, debido a la resistencia del aire, ¡las plumas y las piedras *no* caen a la vez! La intuición de Galileo fue el darse cuenta de que si se pudiera anular la resistencia del aire, ellas *caerían* a la vez.) Se necesitaron tres siglos para que la profunda significación de esta intuición fuese completamente entendida y constituyese la piedra angular de una gran teoría. Esta teoría fue la relatividad de Einstein —una extraordinaria descripción de la gravitación que, como comprenderemos en un momento, ¡necesita el concepto de un *espacio-tiempo curvo* para su realización!

¿Qué tiene que ver la intuición de Galileo con la idea de «curvatura del espacio-tiempo»? ¿Cómo es posible que tal idea, en apariencia tan diferente del esquema de Newton en el que las partículas se aceleran bajo la acción de las fuerzas gravitatorias ordinarias, pudiera reproducir, e incluso mejorar, toda la soberbia precisión de dicha teoría? Además, ¿realmente puede ser cierto que la antigua intuición

de Galileo contuviese algo que *no* fuera incorporado posteriormente a la teoría de Newton?

Permítaseme empezar por la última pregunta puesto que es la más fácil de responder. ¿Qué es, según la teoría de Newton, lo que gobierna la aceleración de un cuerpo sometido a la gravedad? En primer lugar tenemos la *fuerza* gravitatoria sobre dicho cuerpo, que la ley de atracción gravitatoria de Newton nos dice debe ser *proporcional a la masa del cuerpo*. En segundo lugar está la cantidad en que se acelera el cuerpo *dada* la fuerza que actúa sobre él, cantidad que, por la segunda ley de Newton, es *inversamente proporcional a la masa del cuerpo*. El hecho del que depende la intuición de Galileo es que la «masa» que aparece en la ley de la fuerza gravitatoria de Newton es la *misma* que la «masa» de la segunda ley de Newton. (Podría decirse «proporcional a» en lugar de «la misma que».) Esto es lo que asegura que la aceleración del cuerpo sometido a la gravedad es realmente *independiente* de su masa. No hay nada en el esquema general de Newton que exija que estos dos conceptos de masa sean el mismo. Newton sencillamente lo *postuló*. De hecho, las fuerzas eléctricas son semejantes a las fuerzas gravitatorias en que ambas son fuerzas del tipo de la inversa del cuadrado, pero la fuerza depende ahora de la *carga eléctrica* que es totalmente diferente de la *masa* de la segunda ley de Newton. La «intuición de Galileo» no se aplicaría a las fuerzas eléctricas: ¡los objetos (esto es, los objetos cargados) «arrojados» en un campo eléctrico *no* «caen» todos con la misma velocidad!

Por el momento, aceptemos simplemente la intuición de Galileo —para el movimiento bajo la acción de la *gravedad*— y busquemos sus consecuencias. Imaginemos a Galileo arrojando dos piedras desde la Torre Inclinada de Pisa. Si hubiera habido una cámara de vídeo en una de las piedras, apuntando hacia la otra, entonces la imagen que proporcionaría sería la de una piedra suspendida en el espacio, ¡aparentemente *inafectada* por la gravedad (Fig. 5.23)! Esto sucede precisamente *debido a* que todos los objetos sometidos a la gravedad caen con la misma velocidad.

Estamos ignorando aquí la resistencia del aire. Los vuelos espaciales nos ofrecen ahora una mejor verificación de estas ideas, puesto que no hay aire en el espacio exterior. Ahora, «caída» en el espacio significa sencillamente seguir en órbita apropiada bajo la gravedad. No es necesario que esta «caída» sea una línea recta hacia abajo, hacia el centro de la Tierra. Puede haber también una componente horizontal del movimiento. Si esta componente horizontal es suficientemente grande, entonces ¡se puede «caer» alrededor de la Tierra sin

aproximarse al suelo! Viajar en una órbita libre bajo la acción de la gravedad es solamente una sofisticada (y muy «costosa») manera de «caer». Ahora, igual que en la imagen de la vídeo-cámara anterior, un astronauta durante un «paseo espacial» ve su nave espacial ¡como si estuviera suspendida delante de él! (Ver Fig. 5.24.) Por consiguiente, podemos eliminar localmente los efectos de la gravedad pasando al «sistema de referencia acelerado» en caída libre.

La gravedad puede así *neutralizarse* mediante la caída libre debido a que los efectos del campo gravitatorio son precisamente iguales a los de una aceleración. De hecho, si usted está en el interior de un ascensor que acelera hacia arriba simplemente experimentará un incremento del campo gravitatorio aparente; y si acelera hacia abajo, una disminución. Si se rompiera el cable que suspende el ascensor, entonces (despreciando la resistencia del aire y los efectos de rozamiento) la aceleración resultante hacia abajo neutralizaría completamente el efecto de la gravedad, y los ocupantes del ascensor parecerían flotar libremente —como el astronauta anterior— ¡hasta que el



Fig. 5.23. Galileo dejando caer dos piedras (y una vídeo-cámara) desde la Torre Inclinada de Pisa.



Fig. 5.24. El astronauta ve su vehículo espacial suspendido ante él, aparentemente inafectado por la gravedad.

ascensor golpea el suelo! Incluso en un tren o en un avión las aceleraciones pueden ser tales que nuestras sensaciones sobre la intensidad y dirección de la gravedad pueden sugerirnos un «abajo» diferente del que nos sugiere la evidencia visual. Esto es debido a que los efectos aceleracionales y gravitatorios *son* exactamente iguales, de modo que las sensaciones son incapaces de distinguir uno del otro. Este hecho —que los efectos locales de la gravitación son equivalentes a los de un sistema de referencia acelerado— es lo que Einstein denominó *principio de equivalencia*.

Las consideraciones anteriores son «locales». Sin embargo, si se nos permite hacer medidas (no completamente locales) de precisión suficiente, podemos, en principio, descubrir una *diferencia* entre un «verdadero» campo gravitatorio y una pura aceleración. En la Figura 5.25 he mostrado, un poco exagerado, cómo una disposición de partículas que forman inicialmente una superficie esférica, cayendo libremente bajo la acción de la gravedad terrestre, empezará a ser afectada por la *no-uniformidad* del campo gravitatorio (newtoniano). El campo es no-uniforme en dos aspectos. En primer lugar, puesto que el centro de la Tierra está a una distancia finita, las partículas más próximas a la superficie de la Tierra se acelerarán hacia abajo más rápidamente que las que están más altas (recordemos la ley de la inversa del cuadrado de Newton). En segundo lugar, y por la mis-



Fig. 5.25. El efecto de marea. Las flechas dobles muestran la aceleración relativa (WEYL).

ma razón, habrá ligeras diferencias en la *dirección* de esta aceleración para diferentes desplazamientos horizontales de las partículas. Debido a esta no-uniformidad, la forma esférica comienza a distorsionarse ligeramente transformándose en un «elipsoide». Se alarga en dirección hacia el centro de la Tierra (y también en la dirección contraria), puesto que las partes más próximas al centro experimentan una aceleración ligeramente mayor que las partes más distantes; se estrecha en las direcciones horizontales, debido al hecho de que las aceleraciones dirigidas hacia el centro de la Tierra tienen una pequeña componente horizontal hacia adentro.

Este efecto distorsionante se conoce como el efecto *de marea* de la gravedad. Si reemplazamos el centro de la Tierra por la Luna, y la superficie esférica de partículas por la superficie de la Tierra, entonces tenemos precisamente la influencia de la Luna sobre las mareas en la Tierra, con abultamientos que se producen tanto en la dirección de la Luna como en la dirección contraria. El efecto de marea es una característica general de los campos gravitatorios que no pueda ser «eliminada» por la caída libre. Este efecto mide la no-uniformidad del campo gravitatorio. (La *magnitud* de la distorsión de marea decrece como el inverso del *cubo* de la distancia al centro de atracción, en lugar de como el inverso del cuadrado.)

La ley de la fuerza gravitatoria inversa del cuadrado de Newton resulta tener una interpretación simple en términos de este efecto de marea: el *volumen* del elipsoide en que se transforma la esfera inicial ¹⁵ es *igual* al de la esfera original —si se considera que la superficie esférica rodea a un vacío. Esta propiedad del volumen es característica de la ley del inverso del cuadrado; no se da para ninguna otra ley de fuerzas. Supongamos, a continuación, que la superficie esférica no rodea un vacío sino a una cierta materia de masa total *M*. Ahora habrá una componente adicional hacia adentro de la aceleración debido a la atracción gravitatoria de esta materia. El volumen del elipsoide en que se transforma la superficie esférica que forman las partículas inicialmente, *se contrae* —precisamente en una cantidad que es *proporcional a M*. Un ejemplo del efecto de reducción de volumen ocurriría si consideráramos esta superficie esférica rodeando la Tierra a una altura constante (Fig. 5.26). Entonces la aceleración hacia abajo (esto es, hacia adentro) ordinaria debida a la gravedad terrestre es la que causa la reducción del volumen de nuestra esfera. Esta propiedad de reducción-de-volumen codifica la parte restante de la ley de la fuerza gravitatoria de Newton, a saber, que la fuerza es proporcional a la masa del *cuerpo atractivo*.

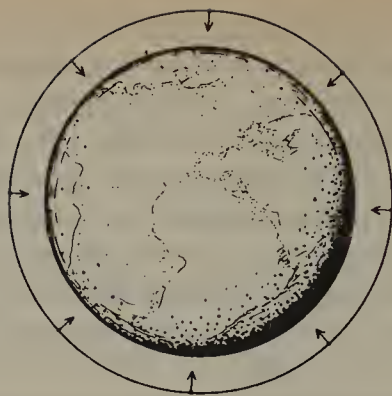


Fig. 5.26. Cuando la superficie esférica rodea materia (en este caso la Tierra) hay una aceleración neta hacia adentro (**RICCI**).

Tratemos de obtener una imagen espacio-temporal de esta situación. En la Figura 5.27 he indicado las líneas-de-universo de las partículas de nuestra superficie esférica (dibujada como un círculo en la Fig. 5.25), en donde he hecho la descripción en un sistema para el que el punto en el centro de la esfera parece estar en reposo («caída



Fig. 5.27. Curvatura espacio-temporal: el efecto de marea mostrado en el espacio-tiempo.

libre»). El punto de vista de la relatividad general consiste en considerar los movimientos de caída libre como «movimientos naturales» —los análogos del «movimiento uniforme en línea recta» que tenemos en la física sin gravedad. Así, *tratamos* de considerar la caída libre como descrita por líneas-de-universo «rectas» en el espacio-tiempo. Sin embargo, por el aspecto de la Figura 5.27, sería confuso utilizar la *palabra* «recta» para esto y, como cuestión de terminología, llamaremos *geodésicas* en el espacio-tiempo a las líneas-de-universo de las partículas en caída libre.

¿Es ésta una buena terminología? ¿Qué entendemos normalmente por una «geodésica»? Examinemos una analogía para una superficie curva bidimensional. Las geodésicas son las curvas en dicha superficie que (localmente) son los «camino más cortos». Así, si pensamos en un trozo de cuerda estirado sobre la superficie (y no demasiado largo o podría escurrirse) entonces la cuerda estará a lo largo de una geodésica de la superficie. En la Figura 5.28 he indicado dos ejemplos de superficies, la primera con lo que se llama «curvatura positiva» (como la superficie de una esfera) y la segunda con «curvatura negativa» (una superficie en forma de silla de montar). Para la superficie de curvatura positiva, dos geodésicas próximas que comienzan paralelas entre sí empezarán a curvarse *acercándose* si seguimos a lo largo de ellas; para curvatura negativa, empezarán a curvarse *apartándose* una de la otra. Si imaginamos que las líneas-de-universo de partículas en caída libre son en algún sentido análogas a las geodésicas en una superficie, entonces vemos que hay una estrecha analogía entre el efecto de marea gravitatorio, discutido arriba, y los efectos de curvatura de una superficie —pero ahora los efectos de curvatura positiva y negativa están *ambos* presentes. Miremos las Figuras 5.25, 5.27.

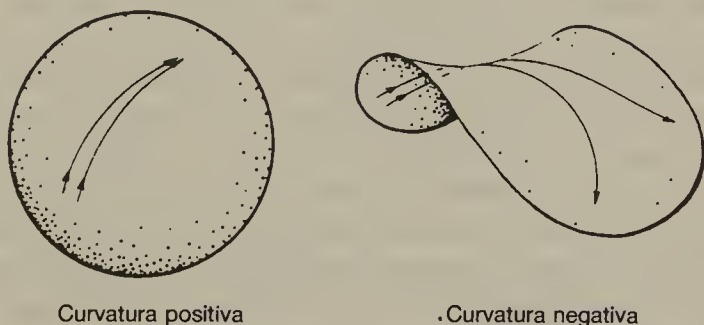


Fig. 5.28. Geodésicas en una superficie curva. Con curvatura positiva las geodésicas convergen, mientras que con curvatura negativa divergen.

Vemos que nuestras «geodésicas» espacio-temporales empiezan a *separarse* en una dirección (cuando están alineadas en la dirección de la Tierra) —como sucede con la superficie de curvatura *negativa* de la Fig. 5.28— y comienzan a *acercarse* en otras direcciones (cuando se desplazan horizontalmente con relación a la Tierra) —como sucede con la superficie de curvatura *positiva* de la Fig. 5.28. Así, nuestro espacio-tiempo parece poseer realmente una «curvatura», análoga a la de nuestras dos superficies pero más complicada debido a la mayor dimensionalidad, y mezclas de curvaturas tanto positivas como negativas están involucradas para diferentes desplazamientos.

Esto muestra cómo puede utilizarse un concepto de «curvatura» del espacio-tiempo para describir la acción de campos gravitatorios. La posibilidad de utilizar semejante descripción procede en última instancia de la intuición de Galileo (el principio de equivalencia) y nos permite eliminar la «fuerza» gravitatoria mediante la caída libre. De hecho, nada de lo que he dicho hasta ahora requiere que salgamos de la teoría newtoniana. Esta nueva imagen proporciona simplemente una *reformulación* de dicha teoría. Sin embargo, interviene nueva física cuando tratamos de combinar esta imagen con lo que aprendimos a partir de la descripción de Minkowski de la *relatividad especial* —la geometría del espacio-tiempo que ahora sabemos se aplica *en ausencia* de gravedad. La combinación resultante es la *relatividad general* de Einstein.

Recordemos lo que Minkowski nos ha enseñado. Tenemos (en ausencia de gravedad) un espacio-tiempo con un tipo peculiar de medida de «distancia» definida entre dos puntos: si tenemos una línea-de-universo en el espacio-tiempo, que describe la historia de alguna partícula, entonces la «distancia» minkowskiana medida a lo largo de la línea-de-universo describe el *tiempo* que es experimentado realmente por dicha partícula. (En realidad, en la sección previa considerábamos esta «distancia» sólo a lo largo de líneas de universo constituidas por tramos rectos, pero el enunciado se aplica también a líneas-de-universo curvas en las que la «distancia» se mide a lo largo de la curva.) La geometría de Minkowski se supone exacta si no hay campo gravitatorio —es decir, no hay curvatura del espacio-tiempo. Pero en presencia de gravedad consideramos la geometría de Minkowski como sólo aproximada— de la misma forma que una superficie plana da sólo una descripción aproximada de la geometría de una superficie curva. Si imaginamos que tomamos un microscopio cada vez más potente para examinar una superficie curva —de tal modo que la geometría de la superficie aparezca cada vez más estirada— entonces la

superficie aparece cada vez más plana. Decimos que una superficie curva es *localmente* semejante a un plano euclídeo¹⁷. De modo análogo, podemos decir que, en presencia de gravedad, el espacio-tiempo es *localmente* semejante a la geometría de Minkowski (que es espacio-temporalmente *plana*) pero permitimos alguna «curvatura» en una escala mayor (ver Fig. 5.29). En particular, cualquier punto del espacio-tiempo es el vértice de un *cono de luz*, igual que en el espacio de Minkowski, pero estos conos de luz no están dispuestos de la manera completamente uniforme en que lo estaban en el espacio de Minkowski. En el Capítulo 7 veremos algunos ejemplos de modelos de espacio-tiempo para los que esta no-uniformidad es manifiesta: cf. Fig. 7.13, 7.14 en p. 417. Las partículas materiales tienen como líneas-de-universo curvas que están siempre en el *interior* de los conos de luz, y los fotones tienen curvas dirigidas *a lo largo* de la superficie de los conos de luz. También hay un concepto de «distancia» minowskiana a lo largo de tales curvas que mide el tiempo experimentado por las partículas, igual que en el espacio de Minkowski. Como en el caso de una superficie curva, la medida de distancia define una *geometría* para la superficie que puede diferir de la geometría del caso plano.

Ahora puede darse a las geodésicas en el espacio-tiempo una interpretación similar a la de las geodésicas en superficies bidimensionales consideradas arriba, en donde debemos tener en cuenta las diferencias entre las situaciones minowskiana y euclídea. Así, en lugar

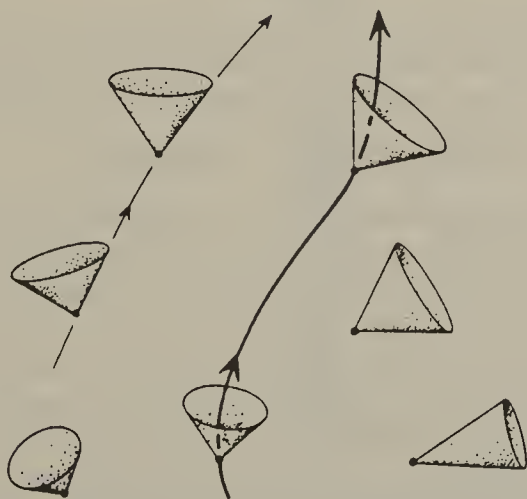


Fig. 5.29. Una representación del espacio-tiempo curvo.

de ser curvas de longitud mínima (localmente), nuestras líneas-de-universo geodésicas en el espacio-tiempo son curvas que *maximizan* (localmente) la «distancia» (es decir, el tiempo) a lo largo de la línea-de-universo. Las líneas-de-universo de las partículas en movimiento libre bajo gravedad *son* realmente geodésicas según esta regla. Así, en particular, los cuerpos celestes que se mueven en un campo gravitatorio están bien descritos por tales geodésicas. Además, los rayos de luz (líneas-de-universo de los fotones) en el espacio vacío son también geodésicas, pero esta vez geodésicas de «longitud» *cero*. Como ejemplo, he indicado esquemáticamente en la Figura 5.30 las líneas-de-universo de la Tierra y el Sol, siendo el movimiento de la Tierra en torno al Sol una geodésica parecida a un «sacacorchos» en torno a la línea-de-universo del Sol. He indicado también un fotón que llega a la Tierra desde una estrella lejana. Su línea-de-universo aparece ligeramente «curvada» debido al hecho de que la luz, según la teoría de Einstein, es *desviada* por el campo gravitatorio del Sol.

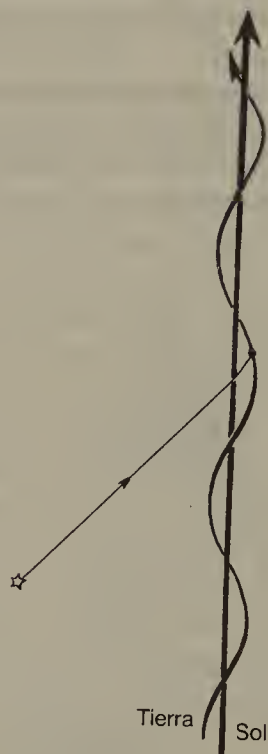


Fig. 5.30. Líneas-de-universo de la Tierra y el Sol, y un rayo de luz procedente de una estrella lejana que es desviado por el Sol.

Tenemos que ver aún cómo debe incorporarse la ley del inverso del cuadrado de Newton —y de qué modo apropiado debe modificarse de acuerdo con la relatividad de Einstein. Volvamos a nuestra esfera de partículas arrojadas en un campo gravitatorio. Recordemos que si la superficie esférica rodea sólo un vacío, entonces, según la teoría de Newton, el volumen de la esfera inicial no cambia; pero si la superficie rodea materia de masa total M , entonces hay una reducción de volumen proporcional a M . En la teoría de Einstein las reglas son exactamente las mismas (para una esfera pequeña) salvo que no es exactamente M la que determina el cambio de volumen; hay una contribución adicional (normalmente pequeñísima) de la *presión* en la materia rodeada.

La expresión matemática completa para la curvatura del espacio-tiempo tetra-dimensional (que debe describir los efectos de marea para partículas que viajan en cualquier dirección posible en cualquier punto dado) viene dada por algo llamado *tensor de curvatura de Riemann*. Éste es un objeto algo complicado que requiere para su especificación veinte números reales en cada punto. Estos veinte números reales se denominan *componentes* del tensor. Las diferentes componentes se refieren a las diferentes curvaturas en diferentes direcciones del espacio-tiempo. El tensor de curvatura de Riemann se escribe normalmente R_{ijkl} , pero, puesto que no deseo explicar aquí qué significa cada uno de estos subíndices (ni, en realidad, qué *es* realmente un tensor), lo escribiré simplemente como:

RIEMANN

Este tensor puede descomponerse en dos partes, llamadas el tensor de Weyl y el tensor de Ricci (con diez *componentes cada uno*). Escribiré este desdoblamiento esquemáticamente como

$$\text{RIEMANN} = \text{WEYL} + \text{RICCI}$$

(Las expresiones detalladas no nos serían especialmente útiles aquí.) El tensor de Weyl **WEYL** mide una *distorsión de marea* de nuestra esfera de partículas en caída libre (es decir, un cambio inicial en forma más que en tamaño), y el tensor de Ricci **RICCI** mide su *cambio de volumen* inicial¹⁹. Recordemos que la teoría de la gravitación de Newton exige que la *masa* rodeada por nuestra superficie esférica en caída sea proporcional a esta reducción del volumen inicial. Esto nos dice, hablando en términos generales, que la densidad de *masa* de la materia —o equivalentemente la densidad de *energía* (puesto que $E = mc^2$)— debería *igualarse* al tensor de Ricci.

De hecho, esto es básicamente lo que afirman las ecuaciones de campo de la relatividad general —a saber, las *ecuaciones de campo de Einstein*²⁰. Sin embargo, hay ciertas cuestiones técnicas sobre ello en las que será mejor que no entremos. Baste decir que existe un objeto llamado el tensor *energía-momento* que organiza toda la información relevante acerca de la energía, presión y momento de la materia y los campos electromagnéticos. Me referiré a este tensor como **ENERGÍA**. Entonces las ecuaciones de Einstein se escriben, muy esquemáticamente:

$$\text{RICCI} = \text{ENERGÍA}$$

(Es la presencia de la «presión» en el tensor **ENERGÍA**, junto con algunos requisitos de consistencia para las ecuaciones en general, la que exige que la presión contribuya también al efecto de reducción-de-volumen descrito arriba.)

Esta ecuación no parece decir nada sobre el tensor de Weyl. No obstante, es una cantidad importante. El efecto de marea que se experimenta en el espacio vacío se debe enteramente a **WEYL**. De hecho, las ecuaciones de Einstein anteriores implican que existen ecuaciones *diferenciales* que ligan **WEYL** con **ENERGÍA**, algo similares a las ecuaciones de Maxwell que encontramos antes²¹. En realidad, un punto de vista fructífero consiste en considerar **WEYL** como un tipo de análogo gravitatorio de la cantidad campo electromagnético (también de hecho un tensor —el tensor de Maxwell) descrito por el par $(\mathbf{E}, \mathbf{B})$. Por consiguiente, en cierto sentido **WEYL** mide realmente el campo *gravitatorio*. La «fuente» de **WEYL** es el tensor **ENERGÍA**, lo que es análogo al hecho de que la fuente del campo electromagnético $(\mathbf{E}, \mathbf{B})$ es $(\rho, \mathbf{j})$, el conjunto de cargas y corrientes de la teoría de Maxwell. Este punto de vista nos será útil en el Capítulo 7.

Puede parecer curioso, cuando tenemos en cuenta estas sorprendentes diferencias en la formulación e ideas subyacentes, que sea difícil descubrir diferencias observacionales entre la teoría de Einstein y la teoría que había desarrollado Newton dos siglos y medio antes. Pero siempre que las velocidades en consideración sean pequeñas comparadas con la velocidad de la luz c , y que los campos gravitatorios no sean demasiado intensos (de modo que las velocidades de escape sean mucho menores que c ; cf. Capítulo 7, p. 295) entonces la teoría de Einstein da resultados prácticamente idénticos a los de la de Newton. No obstante, la teoría de Einstein es más exacta en situaciones en las que las predicciones de ambas teorías difieren. Hay ahora varias verificaciones experimentales muy impresionantes, y la teoría más

reciente de Einstein está completamente justificada. Los relojes marchan más lentos en un campo gravitatorio, como sostenía Einstein, habiendo sido medido este efecto de muchas maneras diferentes. La luz y las señales de radio son desviadas por el Sol, y son ligeramente retardadas por este encuentro —de nuevo efectos de relatividad general perfectamente verificados. Las sondas espaciales y los planetas en movimiento requieren pequeñas correcciones a las órbitas newtonianas, como exige la teoría de Einstein; estas han sido también verificadas experimentalmente. (En particular, la anomalía del movimiento del planeta Mercurio, conocida como la «precesión del perihelio», que había preocupado a los astrónomos desde 1859, fue explicada por Einstein en 1915.) Quizá lo más impresionante de todo sea un conjunto de observaciones de un sistema llamado el *pulsar binario*, consistente en un par de estrellas pequeñas de mucha masa (presumiblemente dos «estrellas de neutrones», cf. p. 415), que concuerdan estrechamente con la teoría de Einstein y verifican indirectamente un efecto que está totalmente ausente en la teoría de Newton: la emisión de *ondas gravitatorias*. (Una onda gravitatoria es el análogo gravitatorio de una onda electromagnética, y viaja a la velocidad de la luz c .) No existe ninguna observación confirmada que contradiga la relatividad general de Einstein. ¡A pesar de toda su extrañeza inicial, la teoría de Einstein está definitivamente con nosotros!

Causalidad relativista y determinismo

Recuérdese que en la teoría de la relatividad los cuerpos materiales no pueden viajar más rápidos que la luz —en el sentido de que sus líneas-de-universo deben estar siempre dentro de los conos de luz (cf. Fig. 5.29). (En relatividad general, particularmente, necesitamos expresar las cosas de este modo local. Los conos de luz no están dispuestos uniformemente, de modo que no tendría mucho significado decir que la velocidad de una partícula muy *lejana* supera la velocidad de la luz *aquí*.) Las líneas-de-universo de los fotones están *a lo largo* de las superficies de los conos de luz, pero no se permite ninguna partícula con línea-de-universo *fuera* de los conos. De hecho debe satisfacerse un enunciado más general, a saber, que ninguna *señal* puede viajar fuera del cono de luz.

Para comprender por qué debe ser así, consideremos nuestra representación del espacio de Minkowski (Fig. 5.31). Supongamos que se ha construido algún dispositivo que pueda enviar una señal con

velocidad un poco mayor que la de la luz. Utilizando este dispositivo el observador W envía una señal desde un suceso A de su línea-de-universo a un suceso distante B que está justo por debajo del cono de luz de A . En la Figura 5.31a se ha dibujado esto desde el punto de vista de W , pero en la Figura 5.31b esto se ha redibujado desde el punto de vista de un segundo observador U que se aleja rápidamente de W (digamos desde un punto entre A y B), y para el que el suceso B parece haber ocurrido *antes* que A ! (Este «redibujado» es un movimiento de Poincaré, como el descrito arriba, p. 259.) Desde el punto de vista de W los espacios simultáneos de U parecen estar «inclinados», que es por lo que el suceso B puede parecer a U anterior a A . Así, para U , la señal transmitida por W ¡parece estar viajando hacia atrás en el tiempo!

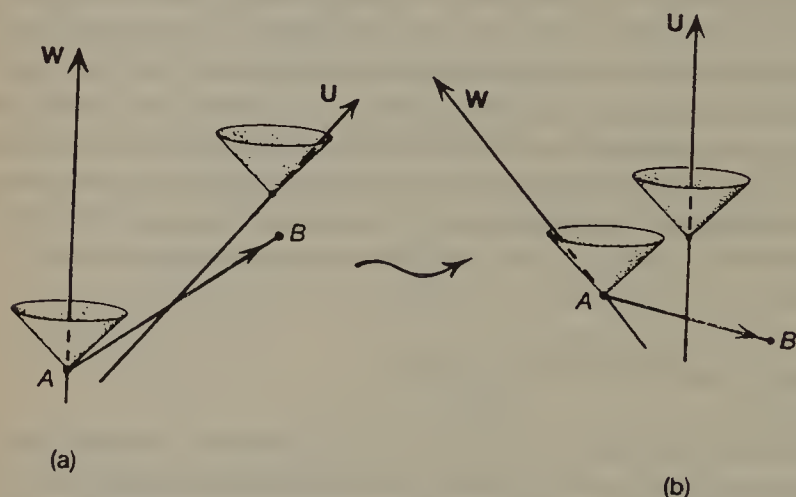


Fig. 5.31. Una señal que para el observador W es más rápida que la luz parecerá, para el observador U , que viaja hacia atrás en el tiempo. La figura de la derecha (b) es simplemente la figura de la izquierda (a) redibujada desde el punto de vista de U . (Este nuevo dibujo puede imaginarse como un movimiento de Poincaré. Compárese con la Fig. 5.21 —aunque en este caso la transformación de (a) a (b) debe tomarse en un sentido *activo* más que pasivo.)

Esto no es todavía una contradicción. Pero por simetría a partir del punto de vista de U (por el principio de relatividad especial), un *tercer* observador V , que se aleja de U en la dirección opuesta de W y equipado con un dispositivo idéntico al de W , podría también enviar

una señal un poco más rápida que la luz, desde *su* (esto es, el de V) punto de vista, de retorno en la dirección de U. Esta señal podría parecer, para U, estar viajando también hacia atrás en el tiempo, ahora en la dirección espacial opuesta. En realidad V transmitirá esta segunda señal de retorno a W en el momento (B) que reciba la original enviada por W. Esta señal alcanza a W en un suceso C que es anterior, en la estimación de U, al suceso A de la emisión original (Fig. 5.32). Pero, lo que es aún peor, el suceso C es realmente anterior al suceso A de emisión en la *propia línea-de-universo* de W, de modo que W *experimenta* realmente que el suceso C ocurre ¡antes de que él emita la señal en A! El mensaje que el observador V devuelve a W podría, por un acuerdo previo con W, repetir simplemente el mensaje que recibió en B. Así, W recibe, en un tiempo anterior en su línea-de-universo, ¡precisamente el mismo mensaje que él va a enviar más tarde! Separando a los dos observadores a una distancia suficientemente grande, podemos disponer que la cantidad en la que la señal de retorno precede a la señal original sea un intervalo de tiempo tan grande como queramos. Quizá el mensaje original de W es que él se ha roto la pierna. Podría recibir el mensaje de retorno *antes* de que el accidente haya ocurrido y entonces (presumiblemente), por la acción de su libre voluntad, ¡tomará las medidas para evitarlo!

Por lo tanto, las señales superlumínicas, junto con el principio de relatividad de Einstein, llevan a una patente contradicción con nuestras sensaciones normales de «libre albedrío». De hecho, la cuestión es aún más seria que eso. En efecto, podríamos imaginar que tal vez

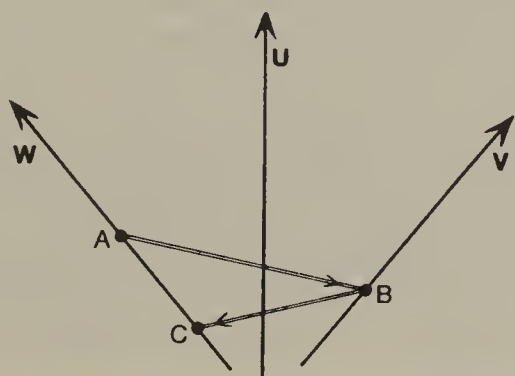


Fig. 5.32. Si V está equipado con un dispositivo de enviar señales más-rápidas-que-la-luz idéntico al de W, pero apuntando en la dirección opuesta, puede ser utilizado por W ¡para enviar un mensaje a su propio pasado!

«el observador W» es simplemente un dispositivo mecánico programado para enviar el mensaje «SÍ» si recibe «NO» y «NO» si recibe «SÍ». De hecho, V puede ser también un dispositivo mecánico pero programado esta vez para enviar de retorno «NO» si recibe «NO» y «SÍ» si recibe «SÍ». Esto conduce a la misma contradicción esencial que teníamos antes ²², ahora aparentemente independiente de la cuestión de si el observador W tiene o no «libre albedrío», y nos dice que un dispositivo de envío de señales más-rápidas-que-la-luz no es una «buena» posibilidad física. Esto tendrá más adelante para nosotros algunas implicaciones enigmáticas (Capítulo 6, p. 361).

Aceptemos entonces que *cualquier* tipo de señal —no simplemente las señales transportadas por partículas físicas ordinarias— deben estar limitadas por los conos de luz. Realmente el argumento superior utiliza la relatividad *especial*, pero en la relatividad general las reglas de la relatividad especial son válidas localmente. Es la validez local de esta relatividad especial la que nos dice que todas las señales están limitadas por los conos de luz, de modo que esto también se aplica en relatividad general. Veremos de qué forma afecta esto a la cuestión del *determinismo* en estas teorías. Recuértese que en el esquema newtoniano (o hamiltoniano, etc.), «determinismo» significa que los *datos iniciales* en *un instante particular* fijan completamente el comportamiento para todos los demás instantes. Si adoptamos una visión espacio-temporal en la teoría newtoniana, entonces el «instante particular» en el que especificamos los datos sería alguna «rebanada» tri-dimensional a través del espacio-tiempo tetra-dimensional (es decir, la totalidad del espacio en dicho instante). En la teoría de la relatividad no existe un concepto global de «tiempo» que pueda seleccionarse para esto. El procedimiento usual consiste en adoptar una actitud más flexible. El «tiempo» de cualquiera es bueno. En relatividad especial podemos tomar algún espacio simultáneo de un observador en donde especificar los datos iniciales, en lugar de la «rebanada» anterior. Pero en relatividad general, el concepto de un «espacio simultáneo» no está muy bien definido. En su lugar podemos utilizar la noción más general de *superficie de tipo-espacio* ²³. Una de estas superficies se representa en la Figura 5.33; está caracterizada por el hecho de que está completamente fuera del cono de luz en cada uno de sus puntos —de modo que *localmente* se asemeja a un espacio simultáneo.

El determinismo, en relatividad especial, puede formularse como el hecho de que los datos iniciales en cualquier espacio simultáneo dado S fija el comportamiento en la totalidad del espacio-tiempo.

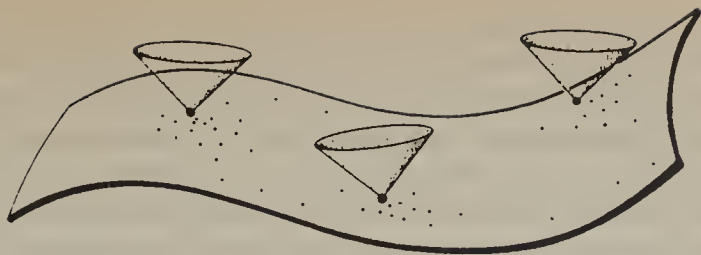


Fig. 5.33. Una superficie de tipo espacio para la especificación de nuestros datos iniciales en relatividad general.

(Esto será cierto, en particular, para la teoría de Maxwell —que es realmente una teoría «relativista-especial».) Sin embargo, existe un enunciado más fuerte que podemos hacer. Si queremos saber lo que va a suceder en algún suceso P que está en alguna parte en el futuro de S , entonces sólo necesitamos los datos iniciales en alguna región acotada (finita) de S , y no en la totalidad de S . Esto es debido a que la «información» no puede viajar más rápida que la luz, de modo que cualesquiera puntos de S que están demasiado alejados para que sus señales luminosas puedan alcanzar P , no pueden tener influencia en P (ver Fig. 5.34) *. Esto es realmente mucho más satisfactorio que la situación que surge en el caso newtoniano en donde, en principio, podríamos tener necesidad de saber qué está pasando en toda la «rebanada» *infinita* para hacer cualquier predicción sobre lo que va a suceder en cualquier punto en un momento posterior. No existe límite para la velocidad a la que puede propagarse la información newtoniana y, de hecho, las fuerzas newtonianas son instantáneas.

El «determinismo» en relatividad *general* es un asunto mucho más complicado que en relatividad especial, y sólo haré algunos comentarios sobre ello. En primer lugar, debemos utilizar una *superficie S de tipo espacio* para la especificación de datos iniciales (en lugar de una simple superficie simultánea). Entonces resulta que las ecuaciones de Einstein dan un comportamiento *localmente* determinista para el campo gravitatorio, suponiendo (como es usual) que los campos de materia que contribuyen al tensor **ENERGÍA** se comportan de forma determinista. Sin embargo, hay complicaciones considerables. La pro-

* Puede señalarse que la ecuación de ondas (cf. nota a pie de página en p. 243) es también, como las ecuaciones de Maxwell, una ecuación relativista. Por lo tanto, el «fenómeno de no-computabilidad» de Pour-El-Richards que consideramos antes es un efecto que también se refiere sólo a datos iniciales en una región acotada de S .

pia geometría del espacio-tiempo —incluyendo su estructura «causal» de conos de luz— es ahora parte de lo que hay que determinar. No conocemos esta estructura de conos-de-luz en el tiempo por venir, de modo que no podemos decir qué partes de S serían necesarias para determinar el comportamiento en algún suceso futuro P . En algunas situaciones extremas puede darse el caso de que incluso la S entera sea insuficiente y, consiguientemente, ¡se pierde el determinismo global! (Aquí están implícitas cuestiones difíciles que están relacionadas con un importante problema no resuelto en la teoría de la relatividad general llamado «censura cósmica» —que tiene que ver con la formación de *agujeros negros* (Tipler *et al.*, 1980); cf. Capítulo 7, p. 415; cf. también nota a pie de página en pp. 419, y p. 428.) Podría parecer altamente improbable que cualquiera de estos posibles «fallos de determinismo» que pudieran ocurrir con campos gravitatorios «extremos» tuviera una conexión directa con asuntos a la escala humana de las cosas, pero vemos a partir de esto que la cuestión del determinismo en relatividad general no es en absoluto tan clara como a uno le gustaría que fuera.

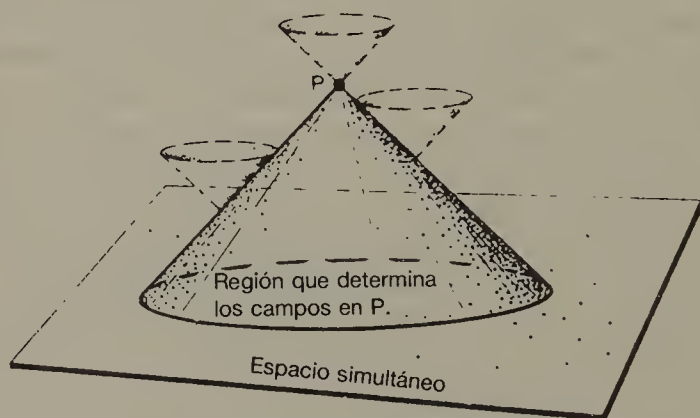


Fig. 5.34. En relatividad especial, lo que sucede en P depende solamente de los datos en una región finita de un espacio simultáneo. Esto es debido a que los efectos no pueden viajar hacia P más rápidos que la luz.

Computabilidad en física clásica: ¿dónde estamos?

A lo largo de este capítulo he tratado de no perder de vista el tema de la *computabilidad*, como distinta del determinismo, y he tratado de indicar que los temas de computabilidad pueden ser tan im-

portantes al menos como los del determinismo cuando afectan a las cuestiones del «libre albedrío» y los fenómenos mentales. Pero el propio determinismo ha resultado no ser ni mucho menos tan claro, en la teoría clásica, como nos habían llevado a pensar. Hemos visto que la ecuación de Lorentz clásica para el movimiento de una partícula cargada da lugar a algunos problemas molestos. (Recuérdese las «soluciones desbocadas» de Dirac.) Hemos señalado también que existen algunas dificultades para el determinismo en relatividad general. Cuando, en tales teorías, no existe determinismo tampoco hay ciertamente computabilidad. Pero en ninguno de los casos recién citados parece que la falta de determinismo tuviera mucha relevancia filosófica directa para nosotros mismos. No hay aún «lugar» para nuestras libres voluntades en dichos fenómenos: en el primer caso, debido a que la ecuación de Lorentz clásica para una partícula puntual (de la forma resuelta por Dirac) no se considera físicamente apropiada en el nivel en el que aparecen estos problemas; y en el segundo, debido a que las escalas en las que la relatividad general clásica podría llevar a tales problemas (agujeros negros, etc.) son completamente diferentes de las escalas de nuestros cerebros.

Ahora bien; ¿dónde estamos con respecto a la *computabilidad* en la teoría clásica? Es razonable conjeturar que, con la relatividad general, la situación no es significativamente diferente de la de la relatividad especial —aparte de las diferencias en causalidad y determinismo que acabo de presentar. Allí donde el comportamiento futuro del sistema físico esté determinado a partir de datos iniciales, dicho comportamiento futuro parecerá estar (por razonamientos similares a los que presenté en el caso de la teoría newtoniana) también *computablemente* determinado por estos datos ²⁴ (aparte del tipo «inútil» de no-computabilidad encontrado por Pour-El y Richards para la ecuación de ondas que consideramos arriba —y que no ocurre para datos que varían *suavemente*). En realidad, es difícil ver que en *cualquiera* de las teorías físicas que he discutido hasta aquí pueda haber elementos «no-computables» significativos. Debe esperarse ciertamente que pueda ocurrir comportamiento «caótico» en muchas de estas teorías, en donde cambios muy pequeños en los datos iniciales pueden dar lugar a diferencias enormes en el comportamiento resultante. (Éste parece ser el caso en relatividad general, cf. Misner 1969, Belinskii *et al.*, 1970.) Pero, como mencioné antes, es difícil ver cómo *este* tipo de no-computabilidad —esto es, «impredecibilidad»— podría ser de «utilidad» en un dispositivo que trate de «aprovechar» los posibles elementos no-computables en las leyes físicas. Si la «mente» puede

estar haciendo uso de algún modo de los elementos no-computables, entonces parece que deberían ser elementos que están fuera de la física clásica. Tendremos que re-examinar más tarde esta cuestión, después de haber echado una mirada a la teoría cuántica.

Masa, materia y realidad

Hagamos un breve inventario de esa imagen del mundo que nos ha presentado la física clásica. En primer lugar, existe un espacio-tiempo que juega un papel primordial como arena para toda la diversa actividad de la física. En segundo lugar, existen *objetos físicos* entregados a esta actividad, aunque limitados por leyes matemáticamente precisas. Los objetos físicos son de dos tipos: *partículas* y *campos*. Poco se ha dicho sobre la naturaleza real y propiedades distintivas de las partículas, salvo que cada una tiene su propia línea-de-universo y posee una masa (en reposo) individual y tal vez carga eléctrica, etc. Los campos, por otro lado, están muy específicamente dados —estando sujeto al campo electromagnético a las ecuaciones de Maxwell y el campo gravitatorio a las ecuaciones de Einstein.

Existe alguna ambigüedad sobre cómo deben ser tratadas las partículas. Si las partículas tienen masas tan minúsculas que podemos despreciar su propia influencia sobre los campos, entonces las partículas se llaman *partículas de prueba* —y en este caso su movimiento en *respuesta* a los campos es inequívoco. La ley de fuerzas de Lorentz describe la respuesta de las partículas de prueba al campo electromagnético, y la ley geodésica describe su respuesta al campo gravitatorio (en una combinación apropiada, cuando ambos campos están presentes). Para esto, las partículas deben considerarse como partículas *puntuales*, es decir, que tienen líneas-de-universo uni-dimensionales. Sin embargo, cuando los efectos de las partículas sobre los campos (y, por lo tanto, sobre otras partículas) tienen que ser considerados —es decir, las partículas actúan como *fuentes* de los campos— entonces las partículas deben considerarse como objetos con cierta extensión en el espacio. De otro modo los campos en el entorno inmediato de cada partícula se hacen infinitos. Estas fuentes extensas proporcionan la distribución de carga-corriente ($\rho, \mathbf{j}$) que se necesita en las ecuaciones de Maxwell y el tensor **ENERGIA** que se necesita para las ecuaciones de Einstein. Además de todo esto, el espacio-tiempo —en el que residen todas las partículas y campos— tiene una estructura variable que por sí misma describe directamente la gravi-

tación. ¡La «arena» interviene en la propia acción de la que ella misma es escenario!

Esto es lo que la física clásica nos ha enseñado sobre la realidad física. Es evidente que hemos aprendido mucho, aunque tampoco deberíamos confiar demasiado en que las imágenes que nos hemos formado no vayan a ser trastocadas por alguna visión posterior más profunda. Veremos en el próximo capítulo que incluso los cambios revolucionarios que ha operado la teoría de la relatividad palidecen casi hasta la insignificancia en comparación con los de la teoría cuántica. No obstante, no hemos acabado completamente con la teoría clásica y con lo que tiene que decirnos sobre la realidad material. ¡Guarda una sorpresa más para nosotros!

¿Qué es la materia? Es la substancia real de la que están compuestos los objetos físicos verdaderos, las «cosas» de este mundo. Es de lo que estamos hechos usted, yo y nuestras casas. ¿Cómo *cuantificamos* esta substancia? Nuestros libros de texto de física elemental nos proporcionan la respuesta clara y nítida de Newton. Es la *masa* de un objeto, o de un sistema de objetos, la que mide la cantidad de materia que contiene este objeto. Esto parece correcto —no hay ninguna otra cantidad física que pueda competir seriamente con la masa como verdadera medida de sustancia total. Además *se conserva*: la masa, y por consiguiente el contenido total de materia, de un sistema cualquiera debe seguir siendo siempre la misma.

Pero la fórmula de la relatividad especial

$$E = mc^2$$

nos dice que masa (m) y energía (E) son intercambiables. Por ejemplo, cuando se desintegra un átomo de uranio, descomponiéndose en partes más pequeñas, la masa total de todas estas partes, si pudieran ser llevadas al reposo, sería *menor* que la masa original del átomo de uranio; pero si se toma en consideración la *energía de movimiento* —energía *cinética*, cf. p. 215 *— de cada parte, y se convierte en valores de masa dividiéndola por c^2 (pues $E = mc^2$), entonces encontramos que el total es realmente *invariante*. La masa se conserva, pero al estar compuesta en parte de energía parece ahora menos evidente que sea la medida de substancia real. Después de todo, la energía depende de la velocidad a que esté viajando la substancia. La energía de movimiento de un tren expreso es considerable, pero si vamos

* En la teoría newtoniana la energía cinética es $\frac{1}{2}mv^2$, donde m es la masa y v la velocidad; pero en la relatividad especial la expresión es algo más complicada.

sentados en el tren, entonces según nuestro propio punto de vista el tren no tiene ningún movimiento en absoluto. La energía de ese movimiento (aunque no así la *energía térmica* de los movimientos aleatorios de las partículas individuales) ha sido «reducida a cero» mediante esta apropiada elección del punto de vista. Para un ejemplo sorprendente, en donde el efecto de la relación masa-energía de Einstein es más extremo, consideremos la desintegración de cierto tipo de partícula subatómica llamada mesón π^0 . Ésta es ciertamente una partícula *material* que tiene una masa (positiva) bien definida. Después de alrededor de 10^{-16} segundos se desintegra (como el átomo de uranio anterior pero mucho más rápidamente)— casi siempre en sólo *dos fotones* (Fig. 5.36). Para un observador en reposo respecto al mesón π^0 , cada fotón lleva la mitad de la energía y, por lo tanto, la mitad de la masa del mesón. Pero esta «masa» del fotón es de un tipo vago: *pura energía*. En efecto, si tuviéramos que viajar rápidamente en la dirección de uno de los fotones, entonces podríamos reducir su masa-energía a un valor tan pequeño como quisiéramos —siendo *cero* realmente la masa intrínseca (o masa *en reposo* como veremos en breve) de un fotón. Todo esto forma una imagen consistente de la masa conservada, pero no es ya lo que teníamos antes. La masa puede aún, en cierto sentido, medir la «cantidad de materia», pero ha habido un

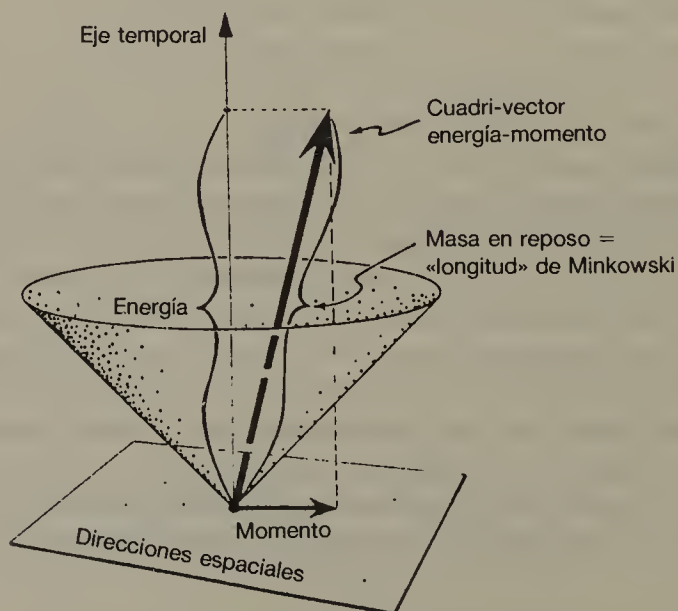


Fig. 5.35. El cuadrivector energía-momento.

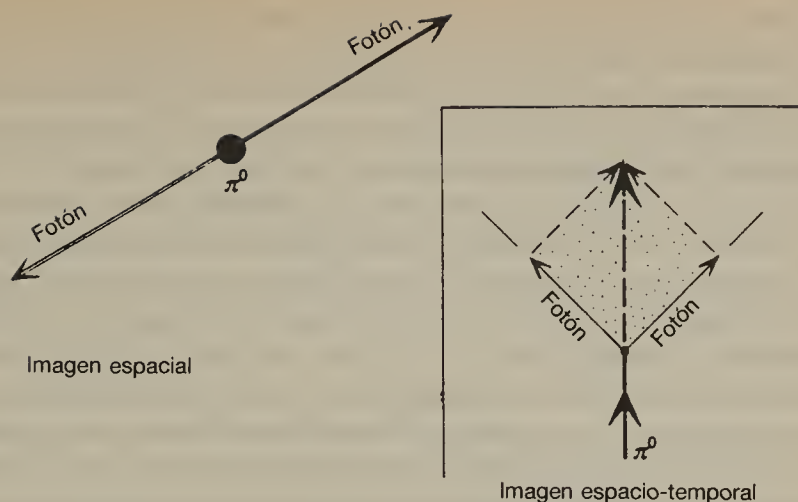


Fig. 5.36. Un mesón π^0 con masa se desintegra en dos fotones sin masa. La imagen espacio-temporal muestra como el cuadrivector energía-momento se conserva: el cuadrivector del mesón π^0 es la suma de los cuadrivectores de los dos fotones, sumados según la ley del paralelogramo (representado con trazos).

cambio de punto de vista: puesto que la masa es equivalente a la energía, ¡la masa de un sistema depende, como la energía, del movimiento del observador!

Merece la pena ser algo más explícitos acerca del punto de vista al que hemos sido llevados. La cantidad conservada que asume el papel de la masa es un objeto global llamado el *cuadrivector energía-momento*. Este puede representarse como una flecha (vector) en el origen O del espacio de Minkowski que apunta hacia el *interior* del cono de luz futuro en O (o, en el caso extremo de un fotón, *sobre* este cono); ver Figura 5.25. Esta flecha, que apunta en la misma dirección que la línea-de-universo del objeto, contiene toda la información sobre la energía, masa y momento. Así el «valor t » (o «altura») de la punta de esta flecha, medida en el sistema de referencia de algún observador, describe la *masa* (o *energía* dividida por c^2) del objeto, según dicho observador, mientras que las componentes espaciales proporcionan el *momento* (dividido por c).

La «longitud» minkowskiana de esta flecha es una cantidad importante conocida como la *masa-en-reposo*. Describe la masa para un observador en reposo respecto al objeto. Podríamos tratar de adoptar el punto de vista de que *esto* sería una buena medida de la «cantidad

de materia». Sin embargo, no es aditiva: si un sistema se desdobra en dos, entonces la masa-en-reposo original no es la suma de las dos masas-en-reposo resultantes. Recuérdese la desintegración del mesón π^0 considerada antes. El mesón π^0 tiene una masa-en-reposo positiva, mientras que las masas-en-reposo de cada uno de los dos fotones resultantes es nula. Sin embargo, la propiedad de aditividad es satisfecha realmente por la flecha global (cuadrivector), donde ahora debemos «sumar» en el sentido de la ley de la suma *vectorial* representada en la Figura 5.6. ¡Es ahora esta *flecha entera* nuestra medida de la «cantidad de materia»!

Consideremos ahora el campo electromagnético de Maxwell. Hemos resaltado que lleva energía. Por $E = mc^2$ debe tener también masa. ¡Por lo tanto, el campo de Maxwell es también materia! Debemos ciertamente aceptar esto ahora puesto que el campo de Maxwell está íntimamente involucrado en las fuerzas que ligan las partículas. Debe haber una contribución substancial ²⁵ a la masa de cualquier cuerpo que procede del campo electromagnético en su interior.

¿Qué hay del campo gravitatorio de Einstein? En muchos aspectos se parece al campo de Maxwell. De modo análogo a como, en la teoría de Maxwell, los objetos cargados en movimiento pueden emitir ondas *electromagnéticas*, los cuerpos con masa en movimiento pueden emitir (según la teoría de Einstein) ondas *gravitatorias* (cf. p. 273) —que, como las ondas electromagnéticas, viajan a la velocidad de la luz y transportan energía. Pero esta energía no se mide de la manera estándar, que sería mediante el tensor **ENERGÍA** antes mencionado. ¡En una onda gravitatoria (pura) el tensor es realmente *cero* en cualquier parte! Podríamos aceptar, de todas formas, la idea de que la *curvatura* del espacio-tiempo (ahora dada completamente por el tensor **WEYL**) puede representar de algún modo la «substancia» de las ondas gravitatorias. Pero la energía gravitatoria resulta ser *no-local*, lo que quiere decir que no podemos determinar cuál es la medida de la energía examinando simplemente la curvatura del espacio-tiempo en regiones limitadas. La energía —y por consiguiente la masa— de un campo gravitatorio es una anguila escurridiza y rehúsa estar quieta en ninguna localización clara. De todas formas, debe tomarse seriamente. Está ciertamente *ahí*, y tiene que ser tenida en cuenta para que el concepto de masa se conserve globalmente. Existe una buena (y positiva) medida de la masa (Bondi, 1960, Sachs, 1962) que se aplica a las ondas gravitatorias, ¡pero la no-localidad es tal que hace que esta medida pueda ser a veces *no-nula* en regiones *planas* del espacio-tiempo —entre dos estallidos de radiación (un poco como la

calma en el ojo de un huracán)— en las que el espacio-tiempo está totalmente *libre* de curvatura (cf. Penrose y Rindler, 1986, p. 427) (es decir, *tanto WEYL como RICCI* son nulos). En tales casos, parecemos obligados a deducir que si esta masa-energía tiene que estar localizada, debe estarlo en este *espacio vacío plano*— una región totalmente libre de materia o campos de cualquier tipo. En estas curiosas circunstancias, nuestra «cantidad de materia» está *allí*, en las regiones más vacías del espacio vacío, ¿o no está en ninguna parte!

Esto parece una pura paradoja. Pero es una clara consecuencia de lo que nos dicen nuestras mejores teorías clásicas —y son teorías realmente soberbias— sobre la naturaleza de la substancia «real» de nuestro mundo. La realidad material, según la teoría clásica, y ya no digamos en la teoría cuántica que estamos a punto de explorar, es algo mucho más vago de lo que hubiéramos imaginado. ¡Su cuantificación —e incluso su existencia o no— depende de puntos extraordinariamente sutiles y no puede verificarse localmente! Si esta no-localidad parece enigmática prepárense para las sorpresas mucho mayores que vienen ahora.

6. Magia cuántica y misterio cuántico

¿Necesitan los filósofos la teoría cuántica?

En física clásica existe, de acuerdo con el sentido común, un mundo objetivo «ahí fuera». El mundo evoluciona de un modo claro y determinista, gobernado por ecuaciones matemáticas formuladas exactamente. Esto es tan cierto para las teorías de Maxwell y Einstein como para el esquema original de Newton. Se considera que la realidad física existe independientemente de nosotros mismos; y el modo exacto de «ser» del mundo clásico no está afectado por el modo en que decidimos observarlo. Además, nuestros cuerpos y nuestros cerebros van a ser ellos mismos parte de este mundo. Se considera que también ellos evolucionan de acuerdo con las mismas ecuaciones clásicas exactas y deterministas. Todas nuestras acciones van a estar fijadas por estas ecuaciones —por mucho que podamos tener la sensación de que nuestras voluntades conscientes están influyendo en nuestro modo de comportarnos.

Esta imagen parece estar en el fondo de los argumentos filosóficos más serios ¹ concernientes a la naturaleza de la realidad, de nuestras percepciones conscientes y de nuestro aparente libre albedrío. Algunas personas no se quedarán tranquilas con esto pensando que también debería haber un papel para la *teoría cuántica* — ese fundamental pero perturbador esquema de cosas que surgió, en el primer cuarto de este siglo, a partir de las observaciones de discrepancias sutiles entre el comportamiento real del mundo y las descripciones de la física clásica. Para muchos, el término «teoría cuántica» evoca simplemente el concepto vago de un «principio de incertidumbre» que, en el nivel de las partículas, átomos o moléculas, prohíbe la precisión

en nuestras descripciones y da lugar a un comportamiento meramente probabilístico. La verdad es que las descripciones cuánticas *son* muy precisas, como veremos, aunque radicalmente diferentes de las familiares descripciones clásicas. Además encontraremos que, pese a la opinión común en contra, las probabilidades *no* surgen en el ínfimo nivel cuántico de las partículas, átomos o moléculas —éstos evolucionan de modo *determinista*— sino que, aparentemente, lo hacen por vía de cierta misteriosa acción a mayor escala ligada a la emergencia de un mundo clásico que podemos percibir conscientemente. Debemos tratar de comprender esto y de qué forma la teoría cuántica nos obliga a cambiar nuestra concepción de la realidad física.

Se tiende a pensar que las discrepancias entre la teoría cuántica y la clásica son muy insignificantes, pero de hecho subyacen también a muchos fenómenos físicos a escala ordinaria: la existencia misma de los cuerpos sólidos, la resistencia y propiedades físicas de los materiales, la naturaleza de la química, los colores de las sustancias, los fenómenos de congelación y ebullición, la fiabilidad de la herencia; estas y muchas otras propiedades familiares requieren la teoría cuántica para su explicación. Quizá el fenómeno de la consciencia sea también algo que no puede entenderse en términos enteramente clásicos. Tal vez nuestras mentes son cualidades arraigadas en alguna extraña y misteriosa característica de las leyes físicas que *realmente* gobiernan el mundo en que vivimos, en lugar de ser simples características de algún algoritmo ejecutado por los llamados «objetos» de una estructura física *clásica*. Quizá, en cierto sentido, éste sea el «porqué» de que debamos vivir, en tanto que seres sensibles, en un mundo cuántico en lugar de uno enteramente clásico, a pesar de toda la riqueza, y misterio que está ya presente en el universo clásico. ¿Sería *necesario* un mundo cuántico para que pudieran formarse a partir de sus sustancias criaturas pensantes y perceptivas, como nosotros mismos? ¡Esta pregunta parece más apropiada para un Dios, intentando construir un universo habitado, que para nosotros! No obstante, también tiene relevancia para nosotros. Si la consciencia no puede formar parte de un mundo clásico, entonces nuestras mentes deben depender de algún modo de las desviaciones concretas respecto de la física clásica. Ésta es una idea a la que volveré más adelante en este libro.

Debemos entender la teoría cuántica —la más exacta y misteriosa de las teorías físicas— si queremos ahondar profundamente en algunos temas mayores de la filosofía: ¿cómo se *comporta* nuestro mundo y qué constituye las «mentes» que, en realidad, son «nosotros»? No obstante, algún día la ciencia puede darnos una comprensión *más*

profunda de la que nos puede proporcionar la teoría cuántica. Mi opinión personal es que incluso la teoría cuántica es un callejón sin salida, inadecuado en ciertos puntos esenciales para proporcionar una imagen completa del mundo en el que realmente vivimos. Pero esto no nos sirve de excusa; para obtener algunas de las intuiciones filosóficas que deseamos es preciso que comprendamos la imagen del mundo según la teoría cuántica existente.

Por desgracia, diferentes teóricos tienden a tener puntos de vista muy diferentes (aunque observacionalmente equivalentes) sobre la *realidad* de esta imagen. Muchos físicos, encabezados por la figura señera de Niels Bohr, dirán que *no* hay imagen objetiva en absoluto. En el nivel cuántico nada hay realmente «ahí afuera». En cierto modo, la realidad emerge sólo en relación con los resultados de las «medidas». La teoría cuántica, según esta concepción, proporciona simplemente un procedimiento de cálculo, y no intenta describir el mundo como realmente «es». Esta actitud hacia la teoría me parece demasiado derrotista, y seguiré la línea más positiva que atribuye *realidad física objetiva* a la descripción cuántica: el *estado cuántico*.

Existe una ecuación muy precisa, la *ecuación de Schrödinger*, que proporciona una evolución temporal completamente determinista para este estado. Pero hay algo muy singular acerca de la relación entre el estado cuántico en su evolución temporal y el comportamiento real del mundo físico que se observa tiene lugar. De vez en cuando —cuandoquiera que consideramos que ha tenido lugar una «medida»— debemos descartar el estado cuántico que laboriosamente habíamos estado haciendo evolucionar y usarlo sólo para calcular las diversas probabilidades de que el estado «salte» a uno u otro de un conjunto de *nuevos* estados posibles. Además de la extrañeza de este «salto cuántico», está el problema de decidir cuál pueda ser la configuración física que decreta que realmente se ha hecho una «medida». Después de todo, el propio aparato de medida está construido presumiblemente a partir de constituyentes cuánticos, y también debería evolucionar de acuerdo con la ecuación determinista de Schrödinger. ¿Es necesaria la presencia de un ser consciente para que *realmente* tenga lugar una «medida»? Pienso que sólo una pequeña minoría de los físicos cuánticos sostendría esta opinión. ¿Presumiblemente los observadores humanos están ellos mismos también contruidos a partir de minúsculos constituyentes cuánticos!

Más adelante en este capítulo reexaminaremos algunas de las extrañas consecuencias de este «salto» del estado cuántico —por ejemplo, cómo una «medida» en un lugar ¡puede aparentemente provocar

que ocurra un «salto» en una región lejana! Antes de eso, encontraremos otro fenómeno extraño: a veces dos rutas alternativas, que un objeto puede recorrer sin problemas si cualquiera de ellas se atraviesa por separado, se cancelan completamente entre sí cuando ambas están simultáneamente permitidas, de modo que en este último caso ¡ninguna de ellas puede ser atravesada! Examinaremos también, con algún detalle, cómo se describen realmente los estados cuánticos. Veremos cómo estas descripciones difieren mucho de las correspondientes clásicas. Por ejemplo, ¡las partículas pueden parecer estar en dos lugares a la vez! También nos haremos una idea incipiente de cómo se complican las descripciones cuánticas cuando hay que considerar varias partículas juntas. Resulta que las partículas no tienen descripciones individuales para cada una de ellas, sino que deben considerarse como complicadas superposiciones de configuraciones alternativas de todas ellas en conjunto. Veremos cómo distintas partículas del mismo tipo no pueden tener identidades separadas una de otra. Examinaremos en detalle la extraña (y fundamentalmente mecano-cuántica) propiedad de *spin*. Consideraremos los importantes temas planteados por el experimento mental paradójico del «gato de Schrödinger» y las diversas actitudes diferentes que han expresado los teóricos como intentos, en parte, de resolver este enigma fundamental.

Quizá parte del material de este capítulo no sea tan fácil de comprender como el de los capítulos precedentes (o los siguientes), y es a veces algo técnica. He tratado de no hacer trampa en mis descripciones y tendremos que trabajar un poco más duro que en los otros casos. Esto es así si queremos alcanzar una auténtica comprensión del mundo cuántico. Allí donde un argumento quede poco claro le aconsejo que siga adelante y trate de hacerse una idea de la estructura global. Pero no desespere si una comprensión completa se le muestra esquiva. ¡Está en la propia naturaleza del tema!

Problemas con la teoría clásica

¿Cómo sabemos que la física clásica no nos muestra la auténtica realidad de nuestro mundo? Las principales razones son experimentales. La teoría cuántica no era algo que desearan los teóricos; la mayoría de ellos se encontraron conducidos a su pesar a esta extraña y, en muchos aspectos, filosóficamente insatisfactoria visión del mundo. Pero la teoría clásica, pese a su soberbia grandeza, tiene algunas dificultades profundas. La causa radical de éstas es el hecho de que

deben coexistir dos tipos de objetos físicos: las *partículas*, cada una de ellas descrita mediante un pequeño número *finito* (seis) de parámetros (tres posiciones y tres momentos); y los *campos*, que requieren un número *infinito* de parámetros. Esta dicotomía no es físicamente consistente. Para que un sistema con partículas y campos esté en equilibrio (esto es, «completamente asentado») toda la energía de las partículas debe cederse a los campos. Ésta es una consecuencia del fenómeno llamado «equipartición de la energía»: en el equilibrio la energía se reparte por igual entre todos los grados de libertad del sistema. Puesto que los campos tienen infinitos grados de libertad ¡a las pobres partículas no les queda nada en absoluto!

En particular, los átomos clásicos no serían estables pues todo el movimiento de las partículas se transferiría a los modos ondulatorios de los campos. Recordemos la imagen del átomo como un «sistema solar» introducida por el gran físico experimental anglo-neozelandés Ernest Rutherford en 1911. En lugar del Sol estaría el núcleo central, y en lugar de los planetas estarían los electrones en órbita —en una escala pequeñísima— mantenidos por el electromagnetismo en lugar de la gravitación. Un problema fundamental, y aparentemente insuperable, es que cuando un electrón orbital se mueve alrededor del núcleo debería emitir, de acuerdo con las ecuaciones de Maxwell, ondas electromagnéticas de una intensidad que crece rápidamente hasta infinito en una pequeñísima fracción de segundo, ¡al tiempo que describe una espiral que se cierra y se hunde en el núcleo! Sin embargo, no se observaba nada de esto. Lo que se *observa*, de hecho, es bastante inexplicable sobre la base de la teoría clásica. Los átomos pueden emitir ondas electromagnéticas (luz) pero sólo en destellos de frecuencias discretas específicas: las agudas *líneas espectrales* observadas (Fig. 6.1). Además, estas frecuencias satisfacen reglas «locas»² que no tienen ninguna base desde el punto de vista de la teoría clásica.

Otra manifestación de la inestabilidad de la coexistencia de campos y partículas es el fenómeno conocido como «radiación del cuerpo negro». Imaginemos un objeto a alguna temperatura definida, con la radiación electromagnética en equilibrio con las partículas. En 1900, Rayleigh y Jeans habían calculado que toda la energía sería absorbida por el campo —¡sin límite! Hay un absurdo filosófico implícito en esto (la «catástrofe ultravioleta»: la energía sigue fluyendo sin cesar hacia el campo, con frecuencias cada vez mayores) y la propia Naturaleza se comporta de forma más prudente. Para frecuencias *bajas* de las oscilaciones del campo la energía es como habían predicho Rayleigh y Jeans, pero en el extremo de frecuencias *altas*, donde ellos habían

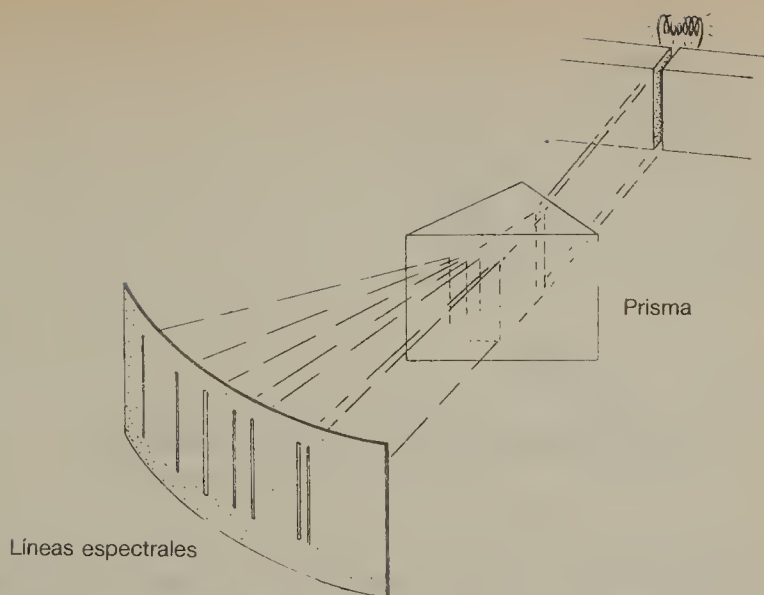


Fig. 6.1. Los átomos en un material calentado emiten luz que a menudo resulta contener sólo frecuencias muy específicas. Las diferentes frecuencias pueden ser separadas mediante el uso de un prisma, y proporcionan las líneas espectrales características del átomo.

predicho la catástrofe, las observaciones reales demuestran que la distribución de energía *no* crece sin límite sino que, por el contrario, cae a cero conforme crece la frecuencia. El mayor valor de la energía tiene lugar a una frecuencia (es decir, un color) muy concreta para una temperatura dada; ver Figura 6.2. (El rojo vivo de un atizador, o el color blanco-amarillento vivo del Sol son, en efecto, dos ejemplos familiares de esto.)

Los comienzos de la teoría cuántica

¿Cómo iban a resolverse estos enigmas? El esquema de partículas original de Newton *necesita* ciertamente ser complementado con el campo de Maxwell. ¿Podemos ir al otro extremo y suponer que *todo* es un campo, siendo las partículas pequeños «nudos» de tamaño finito de algún tipo de campo? Esto también tiene sus dificultades, pues entonces las partículas podrían cambiar continuamente sus formas, agitándose y oscilando de infinitos tipos de modos distintos. Pero no

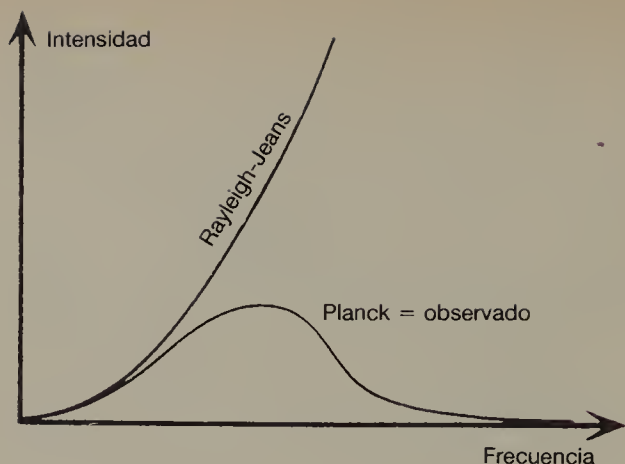


Fig. 6.2. Las discrepancias entre la intensidad de la radiación calculada clásicamente (Rayleigh-Jeans) y observada para un cuerpo caliente («cuerpo negro») llevaron a Planck a los inicios de la teoría cuántica.

es esto lo que vemos; en el mundo físico todas las partículas de la misma clase parecen ser *idénticas*. Por ejemplo, dos electrones cualesquiera son exactamente iguales entre sí. Incluso los átomos o las moléculas sólo pueden adoptar distintas configuraciones de tipo discreto³. Si las partículas *fueran* campos entonces se necesitaría algún nuevo ingrediente para hacer posible que los *campos* adopten características discretas.

En 1900, el brillante aunque conservador y cauto físico alemán Max Planck propuso una idea revolucionaria para eliminar los modos de alta frecuencia del «cuerpo negro»: que las oscilaciones electromagnéticas sólo ocurren en «cuantos» cuya energía E mantiene una relación definida con la frecuencia ν , dada por

$$E = h\nu$$

siendo h una nueva constante fundamental de la Naturaleza, ahora conocida como *constante de Planck*. Con este ingrediente extravagante, Planck pudo obtener un sorprendente acuerdo teórico con la dependencia experimentalmente observada de la intensidad con la frecuencia, la ahora llamada *ley de radiación de Planck*. (La constante de Planck es pequeñísima para los niveles cotidianos, alrededor de $6,6 \times 10^{-34}$ julios por segundo.) Con este golpe capital Planck anunciaba los primeros albores de la teoría cuántica venidera, aunque re-

cibió escasa atención hasta que Einstein hizo una propuesta insólita: ¡que el campo electromagnético sólo puede *existir* en estas unidades discretas! Recordemos que Maxwell y Hertz habían demostrado que la *luz* consiste en oscilaciones del campo electromagnético. Por lo tanto, según Einstein —y como Newton había insistido dos siglos antes— ¡la propia luz debe estar compuesta, después de todo, de *partículas*! (A comienzos del siglo XIX, el brillante teórico y experimentador inglés Thomas Young había establecido aparentemente que la luz consiste en ondas.)

¿Cómo es posible que la luz pueda consistir en partículas y en oscilaciones del campo al mismo tiempo? Estas dos concepciones parecen irrevocablemente opuestas, pero algunos hechos experimentales indicaban claramente que la luz es partículas y otros indicaban que es ondas. En 1923 el aristócrata francés y perspicaz físico príncipe Louis de Broglie llevó esta confusión partícula-onda un paso más lejos cuando, en su disertación doctoral (¡para la que se buscó la aprobación de Einstein!) propuso que ¡las propias partículas de *materia* se comportarían a veces como ondas! La frecuencia de la onda de de Broglie ν , para una partícula de masa m , satisface de nuevo la relación de Planck. Combinada con la relación de Einstein $E=mc^2$, esto nos dice que ν está relacionada con m mediante:

$$h\nu = E = mc^2$$

¡Así, según la propuesta de de Broglie, la dicotomía entre partículas y campos, que había sido una característica de la teoría clásica, *no* se respeta en la Naturaleza! En realidad, cualquier cosa que oscile con alguna frecuencia ν puede ocurrir *sólo* en unidades discretas de masa $h\nu/c^2$. La naturaleza consigue de algún modo construir un mundo consistente en el que ¡*partículas y oscilaciones-de-campo son la misma cosa*! O, más bien, su mundo está constituido por algún ingrediente más sutil, siendo las palabras «partícula» y «onda» imágenes sugerentes pero sólo parcialmente apropiadas.

Otra brillante utilización de la relación de Planck fue dada (en 1913) por Niels Bohr, físico danés y figura sobresaliente del pensamiento científico del siglo XX. Las reglas de Bohr exigían que el *momento angular* (ver p. 216) de los electrones en órbita en torno al núcleo puede darse sólo en múltiplos enteros de $h/2\pi$, para lo que Dirac introdujo más tarde el símbolo conveniente $\hbar$:

$$\hbar = h/2\pi$$

Por lo tanto, los únicos valores permitidos del momento angular (respecto a cualquier eje) son

$$0, \hbar, 2\hbar, 3\hbar, 4\hbar, \dots$$

Con este *nuevo* ingrediente el modelo de «Sistema Solar» para el átomo daba ahora, con una precisión considerable, muchos de los niveles de energía estables discretos ¡así como las reglas «locas» para las frecuencias espectrales que la Naturaleza obedece *realmente*!

Aunque sorprendentemente acertada, la brillante propuesta de Bohr proporcionaba un esquema de «piezas sueltas» algo provisional, conocido como la «antigua teoría cuántica». La teoría cuántica como la conocemos hoy surgió de dos esquemas posteriores independientes, que fueron iniciados por un par de físicos extraordinarios: un alemán, Werner Heisenberg, y un austriaco, Erwin Schrödinger. Al principio estos dos esquemas (la «mecánica matricial» en 1925 y la «mecánica ondulatoria» en 1926, respectivamente) parecían bastante diferentes, pero pronto se demostró que eran equivalentes y fueron englobados poco después en un marco más comprensivo y general, principalmente por obra del gran físico teórico británico Paul Adrien Maurice Dirac. En las secciones que siguen nos haremos una idea de esta teoría y sus extraordinarias consecuencias.

El experimento de la doble-rendija

Consideremos el experimento mecano-cuántico «arquetípico», según el cual un haz de electrones, o de luz, o algún otro tipo de «onda-partícula» se lanza a través de un par de estrechas rendijas hacia una pantalla que hay detrás (Fig. 6.3). Para ser precisos, con-

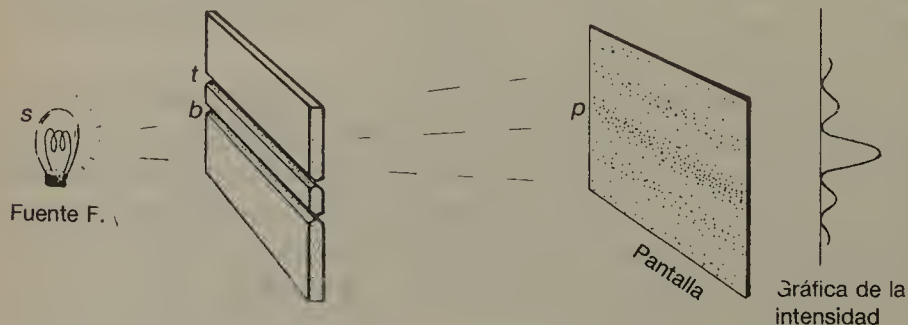


Fig. 6.3. El experimento de la doble-rendija con luz monocromática.

sideraremos *luz* y nos referiremos a los cuantos de luz como «fotones», de acuerdo con la terminología usual. La manifestación más evidente de la luz como *partículas* (es decir, como fotones) tiene lugar en la pantalla. La luz llega allí en unidades de energía discretas y localizadas, estando esta energía invariablemente relacionada con la frecuencia de la luz de acuerdo con la fórmula de Planck: $E = h\nu$. Nunca se recibe la energía de sólo «medio» fotón (o cualquier otra fracción). La recepción de la luz es un fenómeno de todo-o-nada en unidades de fotones. Solamente se han visto números enteros de fotones.

Sin embargo, parece surgir un comportamiento de *tipo ondulatorio* cuando los fotones atraviesan las rendijas. Supongamos, primero, que sólo está abierta una rendija (mientras que la otra está bloqueada). Después de atravesarla, la luz se dispersará por el fenómeno llamado *difracción*, una característica de la propagación de ondas. Pero aún podemos mantener una imagen de partícula e imaginar que la proximidad de los bordes de la rendija ejerce alguna influencia que provoca que los fotones sean desviados a uno u otro lado en una cantidad aleatoria. Cuando la luz que atraviesa la rendija tiene una intensidad razonable, es decir, un gran número de fotones, la iluminación en la pantalla aparece muy uniforme. Pero si se reduce la intensidad, entonces podemos concebir que la distribución de la iluminación está formada por manchas individuales —de acuerdo con la imagen de partículas— en los puntos en donde los fotones individuales inciden sobre la pantalla. La apariencia suave de la iluminación es un efecto estadístico debido al número muy grande de fotones involucrados (ver Fig. 6.4). Para que sirva de comparación, ¡una bombilla de 60 vatios emite alrededor de 100 000 000 000 000 000 000 000 fotones por segundo!

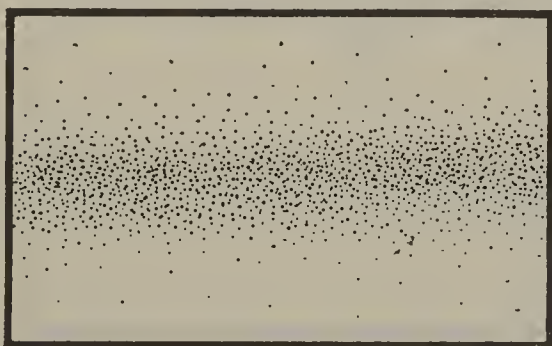


Fig. 6.4. Figura de intensidad en la pantalla cuando sólo está abierta una rendija: una distribución de minúsculos puntos discretos.

Los fotones parecen ser desviados efectivamente de una manera aleatoria cuando atraviesan la rendija, con probabilidades diferentes para diferentes ángulos de desviación, produciendo la distribución de iluminación observada.

Sin embargo, el problema clave para la imagen de partícula viene ¡cuando abrimos la otra rendija! Supongamos que la luz procede de una lámpara amarilla de sodio, de modo que esencialmente consta de un color puro sin mezcla; el término técnico es *monocromática*, es decir, de una longitud de onda o frecuencia definida, lo que significa, en la imagen de la partícula, que todos los fotones tienen la misma energía. En este caso, la longitud de onda es de unos 5×10^{-7} m. Tomemos las rendijas de aproximadamente 0,001 mm de anchura y separadas unos 0,15 mm, estando la pantalla a un metro de distancia. Para una intensidad de luz razonablemente alta, seguimos teniendo una figura de iluminación de apariencia regular, aunque ahora hay una *ondulabilidad* en ella, llamada *figura de interferencia*, con bandas cerca del centro de la pantalla de unos 3 mm de anchura (Fig. 6.5). Hubiéramos esperado que la apertura de la segunda rendija simplemente duplicaría la intensidad de iluminación en la pantalla. De hecho, esto es cierto si consideramos la iluminación *total* en conjunto. Pero ahora la *figura* detallada de intensidad se ve completamente distinta de la que era con una sola rendija. En algunos puntos de la pantalla —en donde la figura alcanza su mayor brillo— la intensidad de iluminación es el *cuádruple* de la que era antes, y no sólo el doble. En otros —donde la figura es más oscura— la intensidad se hace nula. Los puntos de intensidad nula plantean quizá el mayor enigma para la imagen en términos de partícula. Aquéllos eran puntos a los que podía llegar un fotón con bastante facilidad cuando sólo una de las

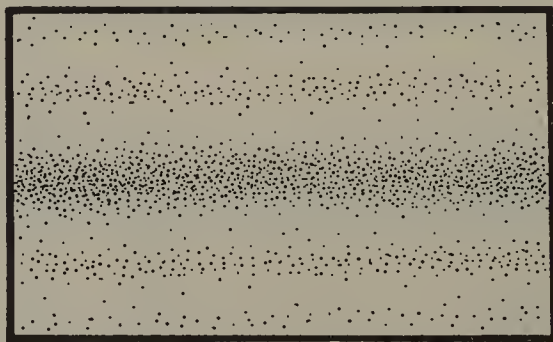


Fig. 6.5. Figura de intensidad cuando ambas rendijas están abiertas —una distribución ondulada de puntos discretos.

rendijas estaba abierta. Ahora que hemos abierto la otra, resulta repentinamente que el fotón está *imposibilitado* de alguna forma para hacer lo que podía hacer antes. ¿Cómo es posible que, al permitir una ruta *alternativa* al fotón, el efecto real sea que le hayamos *impedido* atravesar cualquiera de las rutas?

En la escala del fotón, si tomamos su longitud de onda como una medida de su «tamaño», la segunda rendija está separada unos 300 «tamaños de fotón» de la primera (teniendo cada rendija una anchura de un par de longitudes de onda) (ver Fig. 6.6), así que ¿cómo puede «saber» el fotón, cuando atraviesa una de las rendijas, si la otra rendija está o no abierta? De hecho, no hay límite en principio para la distancia a la que pueden estar las dos rendijas para que ocurra este fenómeno de «cancelación» o de «refuerzo».

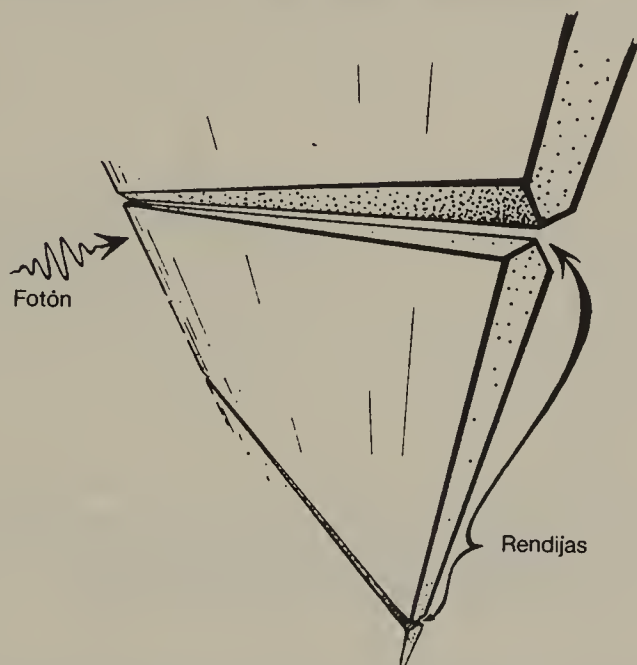


Fig. 6.6. Las rendijas ¡desde el punto de vista del fotón! ¿Cómo puede suponer una diferencia para él el que la segunda rendija, a una distancia de unos 300 «tamaños de fotón», esté abierta o cerrada?

¡Cuando la luz atraviesa la(s) rendija(s) parece estar comportándose como una *onda* y no como una *partícula*! Semejante cancelación —*interferencia destructiva*— es familiar como propiedad de las ondas

ordinarias. Si una onda puede viajar por una u otra de las dos rutas por separado, y si *ambas* le son accesibles, entonces es posible que ellas se cancelen mutuamente. En la Figura 6.7 he ilustrado cómo sucede esto. Cuando una porción de la onda que ha atravesado una de las rendijas encuentra a una porción que ha atravesado la otra, ambas se refuerzan mutuamente si están «en fase» (es decir, cuando coinciden las crestas de las dos porciones y también coinciden sus valles), pero se cancelan mutuamente si están exactamente «fuera de fase» (lo que significa que una tiene una cresta donde la otra tiene un valle). Con el experimento de la doble-rendija, los lugares brillantes de la pantalla se dan dondequiera que las distancias a las dos rendijas difieran en un número *entero* de longitudes de onda, de modo que coincidan las crestas y valles; y los lugares oscuros se dan cuando las diferencias entre las dos distancias están exactamente a medio camino entre estos valores, de modo que cresta coincide con valle y valle con cresta.

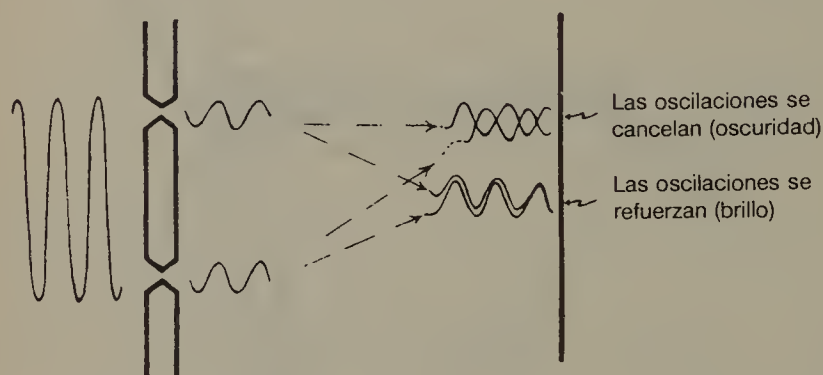


Fig. 6.7. Con una imagen puramente de ondas podemos entender la figura de bandas brillantes y oscuras en la pantalla (aunque no su carácter discreto) en términos de interferencia de ondas.

No hay nada enigmático en una onda clásica macroscópica ordinaria que viaje a través de las dos rendijas a la vez. Después de todo, una onda es sólo una «perturbación», bien de algún medio continuo (campo), o bien de alguna substancia compuesta de miríadas de minúsculas partículas puntuales. Una perturbación puede atravesar en parte una rendija y en parte la otra. Pero ahora las cosas son muy diferentes: ¡cada fotón individual se comporta por sí mismo enteramente como una onda! En cierto sentido, ¡cada partícula atraviesa

ambas rendijas a la vez e interfiere consigo misma! Reduciendo suficientemente la intensidad global de la luz podemos asegurar que no más de un fotón está cada vez en la vecindad de las rendijas. El fenómeno de la interferencia destructiva, en el que se realiza la posibilidad de que dos rutas alternativas para el fotón logran de algún modo cancelarse mutuamente, es algo que se aplica a fotones *individuales*. Si sólo una de las dos rutas está abierta al fotón, entonces el fotón puede recorrer la ruta. Si sólo la otra ruta está abierta, entonces el fotón puede recorrer esta otra ruta en lugar de la anterior. ¡Pero si *ambas* están abiertas las dos posibilidades se cancelan milagrosamente y el fotón parece incapaz de recorrer cualquiera de ellas!

El lector debería detenerse a considerar la importancia de este hecho extraordinario. No se trata en realidad de que la luz se comporta a veces como partículas y a veces como ondas; se trata de que *cada partícula individual* se comporta por sí misma de una manera ondulatoria; ¡y *diferentes posibilidades alternativas abiertas a una partícula pueden a veces cancelarse mutuamente!*

¿Realmente se desdobla el fotón de tal modo que parte de él atraviesa una rendija y parte de él atraviesa la otra? La mayoría de los físicos objetaría esta forma de expresar las cosas. Insistirán en que, si bien las dos rutas abiertas a la partícula deben contribuir al efecto final, éstas son precisamente rutas *alternativas*, y no debe pensarse que la partícula se desdobla en dos al atravesar las rendijas. En apoyo de esta opinión de que la partícula no atraviesa en parte una rendija y en parte la otra, puede considerarse la situación modificada en la que un *detector de partículas* se coloca en una u otra de las rendijas. Puesto que cuando se observa un fotón —o cualquier otra partícula— aparece siempre como un todo individual, y no como una fracción de un todo, nuestro detector debería o bien detectar un fotón entero o bien no detectar nada en absoluto. Sin embargo, cuando el detector está presente en una de las rendijas, de modo que el observador puede *decir* a través de qué rendija ha pasado el fotón, la figura de interferencia ondulatoria en la pantalla desaparece. Para que tenga lugar la interferencia debe haber aparentemente una «falta de conocimiento» acerca de qué rendija atravesó realmente la partícula.

Para obtener interferencia *ambas* alternativas deben contribuir, a veces «sumándose» —reforzándose mutuamente en una cantidad que es el doble de lo que cabría esperar— y a veces «restándose» de modo que las alternativas puedan «cancelarse» misteriosamente. ¡En realidad, según las reglas de la mecánica cuántica, lo que está sucediendo es aún más misterioso que eso! Las alternativas pueden efectivamente

sumarse (los puntos más brillantes de la pantalla), y pueden efectivamente restarse (los puntos oscuros); pero también deben poder combinarse en otras combinaciones extrañas, tales como:

$$\text{«alternativa A» más } i \times \text{«alternativa B»}$$

donde « i » es la «raíz cuadrada de menos uno» ($= \sqrt{-1}$) que encontramos en el Capítulo 3 (en los puntos de la pantalla de intensidad intermedia). De hecho, ¡cualquier número complejo puede jugar este papel en la «combinación de alternativas»!

El lector recordará mi advertencia en el Capítulo 3 acerca de que los números complejos son «absolutamente fundamentales para la estructura de la mecánica cuántica». Estos números no son solamente sutilezas matemáticas. Ellos mismos se abren paso hasta la atención de los físicos a través de persuasivos e inesperados hechos experimentales. Para comprender la mecánica cuántica debemos entender las ponderaciones o pesos estadísticos complejos. A continuación consideraremos lo que éstas implican.

Amplitudes de probabilidad

No hay nada especial en la utilización de fotones en las descripciones anteriores. Servirían igualmente los electrones o cualquier otro tipo de partículas, o incluso átomos enteros. Las reglas de la mecánica cuántica parecen incluso insistir en que las bolas de cricket * y los elefantes deberían comportarse de esta manera singular, según la cual diferentes posibilidades alternativas ¡pueden «sumarse» de algún modo en combinaciones de números complejos! Sin embargo, nunca vemos bolas de cricket o elefantes superpuestos de esta extraña manera. ¿Por qué no lo vemos? Éste es un tema difícil e incluso controvertido, y no quiero abordarlo todavía. Por el momento, como hipótesis de trabajo, supongamos simplemente que existen dos posibles niveles de descripción física que llamaremos el *nivel cuántico* y el *nivel clásico*. Sólo utilizaremos estas extrañas combinaciones de números complejos en el nivel cuántico. Las bolas de cricket y los elefantes son objetos del nivel clásico.

El nivel cuántico es el nivel de las moléculas, átomos, partículas subatómicas, etc. Normalmente se suele pensar que es un nivel de

* Los lectores poco familiarizados con el término «bola de cricket» deberían traducirlo como «primitiva versión inglesa del béisbol».

fenómenos de «muy pequeña» escala, pero esta «pequeñez» no se refiere en realidad al tamaño físico. Veremos que los efectos cuánticos pueden ocurrir sobre distancias de muchos metros, o incluso años-luz. Estaría mucho más cerca de la realidad pensar que algo está «en el nivel cuántico» si implica sólo diferencias muy pequeñas de energía. (Trataré de ser más preciso más adelante, especialmente en el Capítulo 8, p. 457). El nivel clásico es el nivel «macroscópico» del que somos más directamente conscientes. Es un nivel en el que nuestra imagen ordinaria de las «cosas que suceden» se mantiene válida, y en el que podemos utilizar nuestras ideas ordinarias de probabilidad. Veremos que los números complejos que debemos utilizar en el nivel cuántico están estrechamente relacionados con las probabilidades clásicas. No son exactamente lo mismo, pero para entender estos números complejos será útil recordar primero cómo se comportan las probabilidades clásicas.

Consideremos una situación clásica que es *incierto*, de modo que no sabemos cuál de las dos alternativas A o B tiene lugar. Semejante situación podría describirse mediante una combinación «ponderada» de estas alternativas:

$$p \times \text{«alternativa A»} \text{ más } q \times \text{«alternativa B»}$$

donde p es la *probabilidad* de que suceda A y q es la probabilidad de que suceda B. (Recuérdese que una probabilidad es un número real entre 0 y 1. Probabilidad 1 significa «certeza de que sucede» y probabilidad 0 significa «certeza de que no sucede». Probabilidad 1/2 significa «igualmente probable que suceda o que no suceda».) Si A y B son las *únicas* alternativas, entonces la suma de las dos probabilidades debe ser 1:

$$p + q = 1.$$

Sin embargo, si hay más alternativas esta suma debe ser menor que 1. Entonces, la razón $p:q$ da la *razón* de la probabilidad de que suceda A a la de que suceda B. Las probabilidades reales de que suceda A y de que suceda B, de entre estas dos únicas alternativas, serán $p/(p + q)$ y $q/(p + q)$ respectivamente. Podríamos utilizar también la misma interpretación en el caso de que $p + q$ sea mayor que 1. (Esto podría ser útil, por ejemplo, en el caso en que tengamos un experimento que se realiza muchas veces, siendo p el número de ocurrencias de «A» y q el número de ocurrencias de «B».) Diremos que p y q están *normalizadas* si $p + q = 1$, de modo que dan entonces las probabilidades reales, y no sólo las razones de las probabilidades.

En física cuántica tenemos algo que se *parece* mucho a esto, excepto que p y q son ahora números *complejos* —que preferiré denotar, en su lugar, por w y z respectivamente.

$$w \times \text{«alternativa A»} \text{ más } z \times \text{«alternativa B»}$$

¿Cómo tenemos que interpretar w y z ? Ciertamente no son probabilidades (o razones de probabilidades) puesto que cada una de ellas independientemente puede ser negativa o compleja, pero en muchos aspectos se comportan como probabilidades. Se las denomina (cuando están adecuadamente normalizadas —véase más adelante) *amplitudes de probabilidad*, o simplemente *amplitudes*. Además, se utiliza frecuentemente el tipo de terminología que sugieren las probabilidades tal como: «existe una amplitud w de que suceda A y una amplitud z de que suceda B». No son *realmente* probabilidades, pero por el momento supondremos que lo son —o mejor dicho, que son análogas de las probabilidades en el nivel cuántico.

¿Cómo funcionan las probabilidades *ordinarias*? Será útil que pensemos en un objeto macroscópico, digamos una bola, que es golpeada y lanzada a través de uno de entre dos agujeros hacia una pantalla que hay detrás —como el experimento de la doble-rendija descrito antes (cf. Fig. 6.3), pero en donde ahora una bola macroscópica clásica reemplaza al fotón de la discusión previa. Habrá alguna probabilidad $P(f,s)$ de que la bola llegue al agujero superior s *después* de ser golpeada en f , y alguna probabilidad $P(f,i)$ de que llegue al agujero inferior i . Además, si seleccionamos un punto particular p en la pantalla, habrá alguna probabilidad $P(s,p)$ de que cuando la bola *pasa* a través de s llegue al punto particular p de la pantalla, y alguna probabilidad $P(i,p)$ de que cuando pasa a través de i llegue a p . Si solamente está abierto el agujero superior s , entonces para obtener la probabilidad de que la bola llegue a p vía s después de ser golpeada multiplicamos la probabilidad de que vaya de f a s por la probabilidad de que vaya de s a p :

$$P(f,s) \times P(s,p)$$

Análogamente, si solamente está abierto el agujero inferior, entonces la probabilidad de que la bola vaya de f a p es:

$$P(f,i) \times P(i,p)$$

Si *ambos* agujeros están abiertos, entonces la probabilidad de que vaya de f a p *vía* s sigue siendo la primera expresión $P(f,s) \times P(s,p)$,

igual que si sólo s estuviera abierto, y la probabilidad de que vaya de f a p vía i sigue siendo $P(f,i) \times P(i,p)$, de modo que la probabilidad total $P(f,p)$ de que llegue a p partiendo de f es la *suma* de estas dos:

$$P(f,p) = P(f,s) \times P(s,p) + P(f,i) \times P(i,p)$$

En el nivel *cuántico* las reglas son exactamente las mismas, excepto que ahora son estas extrañas *amplitudes* complejas las que deben desempeñar el papel de las probabilidades que teníamos antes. Así, en el ejemplo de la doble-rendija antes considerado, tenemos una amplitud $A(f,s)$ de que un fotón llegue a la rendija superior s a partir de la fuente f , y una amplitud $A(s,p)$ de que llegue al punto p de la pantalla a partir de la rendija s , y multiplicamos estas dos para obtener la amplitud

$$A(f,s) \times A(s,p),$$

de que llegue a la pantalla en p vía s . Como sucede con las probabilidades, ésta es la amplitud correcta si suponemos que la rendija superior está abierta (esté abierta o no la rendija inferior i .) Análogamente, suponiendo que i está abierta, existe una amplitud

$$A(f,i) \times A(i,p)$$

de que el fotón llegue a p procedente de f vía i (esté o no s abierta). Si ambas rendijas están abiertas, entonces tenemos una amplitud total

$$A(f,p) = A(f,s) \times A(s,p) + A(f,i) \times A(i,p)$$

de que el fotón llegue a p procedente de f .

Todo esto está muy bien, pero no nos sirve de mucho hasta que sepamos cómo interpretar estas amplitudes cuando un efecto cuántico se amplifica hasta que alcanza el nivel clásico. Podríamos, por ejemplo, tener un detector de fotones, o *fotocélula*, colocada en p , que proporciona una manera de amplificar un suceso en el nivel cuántico —la llegada de un fotón a p — hasta un acontecer clásico discernible, digamos un «click» audible. (También valdría que la pantalla actuara como placa fotográfica de modo que el fotón deje una mancha visible, pero por claridad nos atendremos al uso de la fotocélula.) ¡Debe haber entonces una *probabilidad* real de que ocurra el «click», y no sólo una de estas misteriosas «amplitudes»! ¿Cómo tenemos que pasar de las amplitudes a las probabilidades cuando ascendemos desde el nivel cuántico al clásico? Resulta que hay una regla muy bella aunque misteriosa para esto.

La regla dice que debemos tomar el *módulo al cuadrado* de la amplitud compleja cuántica para obtener la probabilidad clásica. ¿Qué es un «módulo al cuadrado»? Recuérdese nuestra descripción de los números complejos en el plano de Argand (Capítulo 3, p. 126). El *módulo* $|z|$ de un número complejo z es simplemente la distancia desde el origen (esto es, desde el punto O) hasta el punto definido por z . El *módulo al cuadrado* $|z|^2$ es simplemente este módulo elevado al cuadrado. Así, si

$$z = x + iy,$$

donde x e y son números reales, entonces (por el teorema de Pitágoras, ya que la línea de O a z es la hipotenusa del triángulo rectángulo O, x, z) nuestro módulo al cuadrado requerido es

$$|z|^2 = x^2 + y^2$$

Nótese que para que esto sea una probabilidad real «normalizada», el valor de $|z|^2$ debe estar entre 0 y 1. Esto significa que, para una amplitud adecuadamente normalizada, el punto z en el plano de Argand debe estar en alguna parte dentro del *círculo unidad* (ver Fig. 6.8). A veces, sin embargo, queremos considerar combinaciones

$$w \times \text{alternativa A} + z \times \text{alternativa B},$$

donde w y z son simplemente *proporcionales* a las amplitudes de probabilidad, y entonces no necesitan estar dentro de este círculo. La condición para que estén *normalizadas* (y por consiguiente proporcionen las verdaderas amplitudes de probabilidad) es que la suma de los *módulos al cuadrado* sea la unidad:

$$|w|^2 + |z|^2 = 1.$$

Si no están normalizadas de esta forma, entonces las verdaderas amplitudes para A y B, respectivamente, son $w/\sqrt{|w|^2 + |z|^2}$ y $z/\sqrt{|w|^2 + |z|^2}$, que *están* dentro del círculo unidad.

Vemos ahora que una amplitud de probabilidad no es como una probabilidad, después de todo, sino más bien como una «raíz cuadrada compleja» de una probabilidad. ¿Cómo cambia esto las cosas cuando se amplifican los efectos del nivel cuántico hasta alcanzar el nivel clásico? Recordemos que al manipular las probabilidades y amplitudes necesitábamos a veces multiplicarlas y a veces necesitábamos sumarlas. El primer punto a señalar es que la operación de *multiplicación* no plantea ningún problema en nuestro paso del nivel cuántico al clásico. Esto se debe al notable hecho matemático de que el módulo

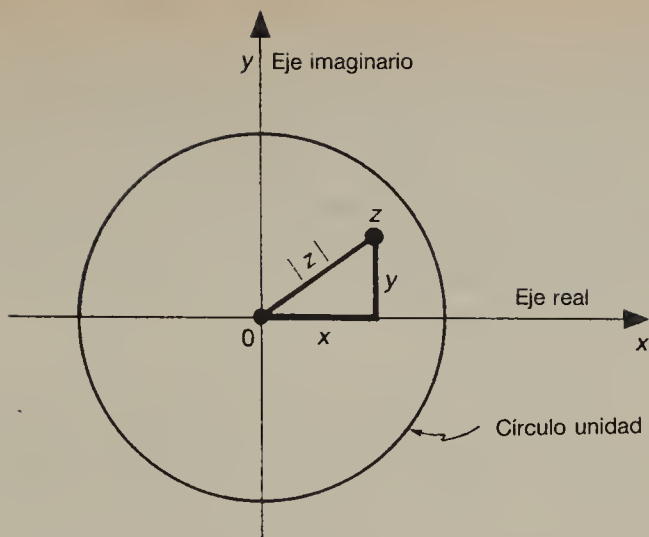


Fig. 6.8. Una amplitud de probabilidad representada como un punto z dentro del círculo unidad en el plano de Argand. El cuadrado de la distancia al centro $|z|^2$ puede transformarse en una probabilidad real cuando los efectos se amplifican hasta el nivel clásico.

al cuadrado del producto de dos números complejos es igual al producto de sus módulos al cuadrado individuales:

$$|zw|^2 = |z|^2 |w|^2$$

(Esta propiedad se sigue inmediatamente de la descripción geométrica del producto de un par de números complejos, dada en el Capítulo 3; pero en términos de las partes real e imaginaria $z = x + iy$, $w = u + iv$, es un pequeño milagro. ¡Inténtelo!)

Este hecho tiene como consecuencia que si existe solamente una ruta abierta a una partícula, v.g. cuando hay sólo una rendija abierta (digamos la s) en el experimento de las dos rendijas, entonces podemos razonar «clásicamente» y las probabilidades vienen a ser las mismas ya se haga o no una detección adicional de la partícula en un punto intermedio (en s) *. Podemos tomar el módulo al cuadrado en ambas etapas o sólo al final, v.g.

$$|A(f,s)|^2 \times |A(s,p)|^2 = |A(f,s) \times A(s,p)|^2,$$

* Esta detección debe hacerse de tal forma que no perturbe el paso de la partícula a través de s . Esto podría conseguirse colocando detectores en *otros* lugares que no sean s e *infiriendo* el paso de la partícula por s cuando estos otros detectores no hacen *click*.

y la respuesta para la probabilidad resultante viene a ser la misma de ambos modos.

Sin embargo, si hay más de una ruta disponible (v.g. si ambas rendijas están abiertas) entonces necesitamos formar una *suma* y es aquí donde comienzan a emerger las características distintivas de la mecánica cuántica. Cuando formamos el módulo al cuadrado de la suma $w+z$, de dos números complejos w y z , normalmente *no* obtenemos la suma de sus módulos al cuadrado por separado; hay un «término correctivo» adicional:

$$|w + z|^2 = |w|^2 + |z|^2 + 2 |w| |z| \cos\theta$$

Aquí, θ es el ángulo que el par de puntos z y w subtienden en el origen en el plano de Argand (ver Fig. 6.9). (Recuérdese que el coseno de un ángulo es la razón «cateto adyacente/hipotenusa» para un triángulo rectángulo.) El lector curioso que no esté familiarizado con la fórmula anterior puede entretenerse en derivarla directamente utilizando la geometría introducida en el Capítulo 3. ¡De hecho, esta fórmula no es otra que la familiar «regla del coseno» ligeramente modificada! Es este término correctivo $2 |w| |z| \cos\theta$ el que proporciona la interferencia cuántica entre alternativas mecano-cuánticas. El valor de $\cos\theta$ puede ir desde -1 a 1 . Cuando $\theta = 0^\circ$ tenemos $\cos\theta = 1$ y las dos alternativas se refuerzan mutuamente de modo que la probabilidad total es mayor que la suma de las probabilidades individuales. Cuando $\theta = 180^\circ$ tenemos $\cos\theta = -1$ y las dos alternativas tienden a cancelarse, dando una probabilidad total menor que la suma de las probabilidades individuales (interferencia destructiva). Cuando $\theta = 90^\circ$ tenemos $\cos\theta = 0$ y estamos en una situación intermedia en la que las dos probabilidades se suman. Para sistemas grandes o com-

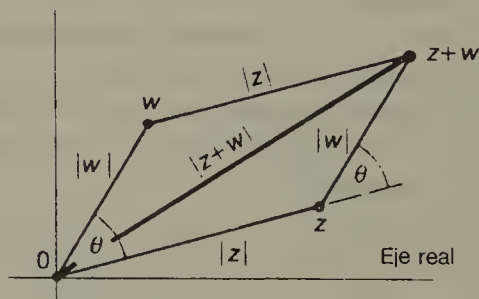


Fig. 6.9. Geometría relativa al término correctivo $2|w| |z| \cos \theta$ al módulo al cuadrado de la suma de dos amplitudes.

plicados los términos correctivos generalmente se «cancelan en promedio» —debido a que el valor «promedio» de $\cos\theta$ es cero— ¡y nos quedamos con las reglas ordinarias de la probabilidad clásica! No obstante, en el nivel cuántico estos términos dan efectos de interferencia importantes.

Consideremos el experimento de la doble-rendija cuando las dos rendijas están abiertas. La amplitud para que el fotón llegue a p es una *suma*, $w + z$, donde

$$w = A(f,s) \times A(s,p) \quad \text{y} \quad z = A(f,i) \times A(i,p).$$

En los puntos *más brillantes* de la pantalla tenemos $w = z$ (de modo que $\cos\theta = 1$) de donde

$$|w + z|^2 = |2w|^2 = 4|w|^2$$

que es el *cuádruple* de la probabilidad $|w|^2$ cuando sólo está abierta la rendija superior —y por consiguiente el *cuádruple* de la intensidad, cuando hay un gran número de fotones, en concordancia con nuestras observaciones. En los puntos *oscuros* de la pantalla tenemos $w = -z$ (de modo que $\cos\theta = -1$) de donde

$$|w + z|^2 = |w - w|^2 = 0,$$

es decir, es *cero* (¡interferencia destructiva!), otra vez en concordancia con la observación. En los puntos exactamente intermedios tenemos $w = iz$ o $w = -iz$ (de modo que $\cos\theta = 0$) de donde

$$|w + z|^2 = |w \pm iw|^2 = |w|^2 + |w|^2 = 2|w|^2$$

que da *doble* intensidad que para una sola rendija (que sería el caso para las partículas clásicas). Veremos el final de la próxima sección cómo calcular dónde están realmente las zonas brillantes, oscuras e intermedias.

Deberíamos señalar un punto final. Cuando ambas rendijas están abiertas la amplitud de que la partícula llegue a p vía s es en realidad $w = A(f,s) \times A(s,p)$, pero no podemos interpretar su módulo al cuadrado $|w|^2$ como la probabilidad de que la partícula atravesase «realmente» la rendija superior para llegar a p . Esto nos daría respuestas absurdas, especialmente si p está en una zona oscura de la pantalla. Pero si decidimos «detectar» la presencia del fotón en s , amplificando el efecto de su presencia (o ausencia) *allí* hasta el nivel clásico, entonces *podemos* utilizar $|A(f,s)|^2$ para la probabilidad de que el fotón esté realmente presente en s . Pero tal detección borraría entonces la

figura ondulatoria. Para que tenga lugar la interferencia debemos asegurar que el paso del fotón a través de las rendijas *quede en el nivel cuántico*, de modo que *ambas* rutas alternativas deben contribuir y pueden a veces cancelarse mutuamente. En el nivel cuántico las rutas alternativas individuales tienen sólo amplitudes, no probabilidades.

El estado cuántico de una partícula

¿Qué tipo de imagen de la «realidad física» se nos presenta en el nivel cuántico, en donde deben poder coexistir siempre las diferentes «posibilidades alternativas» abiertas a un sistema sumadas y ponderadas con estos extraños pesos estadísticos complejos? Muchos físicos no confían en encontrar tal imagen. En lugar de ello, aseguran contentarse con la idea de que la teoría cuántica proporciona simplemente un procedimiento de cálculo para obtener las probabilidades y no una imagen objetiva del mundo físico. Algunos, de hecho, aseguran que la teoría cuántica afirma que no es posible ninguna imagen objetiva —al menos ninguna que sea consistente con los hechos físicos. Por mi parte, considero bastante injustificado este pesimismo. En cualquier caso, sobre la base de lo que hemos discutido hasta ahora, sería prematuro adoptar esta postura. Más adelante tendremos que abordar algunas de las implicaciones más sorprendentemente enigmáticas de los efectos cuánticos, y quizá empezaremos a apreciar más completamente las razones de esta desesperación. Pero por ahora procedamos de forma más optimista y estudiemos la imagen con la que la teoría cuántica parece decirnos que debemos enfrentarnos.

Esta imagen es la que presenta un *estado cuántico*. Tratemos de pensar en una sola partícula cuántica. Clásicamente, una partícula está determinada por su posición en el espacio, y también necesitamos conocer su velocidad (o, de forma equivalente, su momento) para saber qué va a hacer a continuación. En mecánica cuántica *cada posición simple* que la partícula pudiera tener es una «alternativa» disponible para ella. Hemos visto que todas las alternativas deben combinarse de algún modo, con pesos estadísticos complejos. Esta colección de pesos estadísticos complejos describe el estado cuántico de la partícula. Es una práctica estándar, en teoría cuántica, utilizar la letra griega ψ (pronúnciese «psi») para esta colección de pesos estadísticos considerada como una función compleja de la posición —llamada la *función de onda* de la partícula. Para cada posición x , esta función de onda tiene un valor específico, denotado por $\psi(x)$, que es la amplitud

de que la partícula esté en x . Podemos utilizar simplemente la letra ψ para etiquetar el estado cuántico como un todo. Estoy asumiendo la idea de que la *realidad física* de la localización de la partícula es, en realidad, su estado cuántico ψ .

¿Cómo representar la función compleja ψ ? Esto es un poco difícil para todo el espacio tri-dimensional a la vez, de modo que simplifiquemos un poco y supongamos que la partícula está limitada a estar en una línea uni-dimensional —a lo largo del eje- x , pongamos por caso, de un sistema de coordenadas estándar (cartesiano). Si ψ hubiera sido una función real, entonces podríamos haber imaginado un «eje- y » perpendicular al eje- x y representar la *gráfica* de ψ (Fig. 6.10a). Sin embargo, aquí necesitamos un «eje- y complejo» —que estaría en un plano de Argand— para describir el valor de la función *compleja* ψ . Para ello, podemos utilizar, en nuestra imaginación, dos dimensiones espaciales más: por ejemplo, la dirección- y en el espacio como el eje *real* de este plano de Argand y la dirección- z como el eje *imaginario*. Para una imagen precisa de la función de onda representamos $\psi(x)$ como un punto en este plano de Argand (esto es, en el plano (y,z) a través de cada posición en el eje- x). A medida que x varía, este punto varía también, y su curso describe una curva en el espacio que se enrosca en la vecindad del eje- x (ver Fig. 6.10b). Llamemos a esta curva la *curva- ψ* de la partícula. La probabilidad de encontrar la partícula en un punto concreto x , si hubiera un detector de partículas situado en ese punto, se obtiene tomando el módulo al cuadrado de la amplitud $\psi(x)$,

$$|\psi(x)|^2$$

que es el cuadrado de la distancia de la curva- ψ al eje- x *.

Para formar una imagen completa de este tipo para la función de onda en todo el espacio físico tri-dimensional, serían necesarias *cinco* dimensiones: tres dimensiones para el espacio físico más otras dos para el plano de Argand en cada punto en el que representar $\psi(x)$. Sin embargo, nuestra imagen simplificada es aún útil. Si decidimos examinar el comportamiento de la función de onda a lo largo de

* Aquí surge una dificultad técnica debido a que la probabilidad real de encontrar una partícula en un punto *preciso* sería nula. En lugar de ello, consideraremos que $|\psi(x)|^2$ define una *densidad de probabilidad*, lo que significa que lo que está definido es la probabilidad de encontrar una partícula dentro de algún intervalo de tamaño finito centrado en el punto en cuestión. Por lo tanto $\psi(x)$ define una *densidad de amplitud*, y no ya una amplitud.

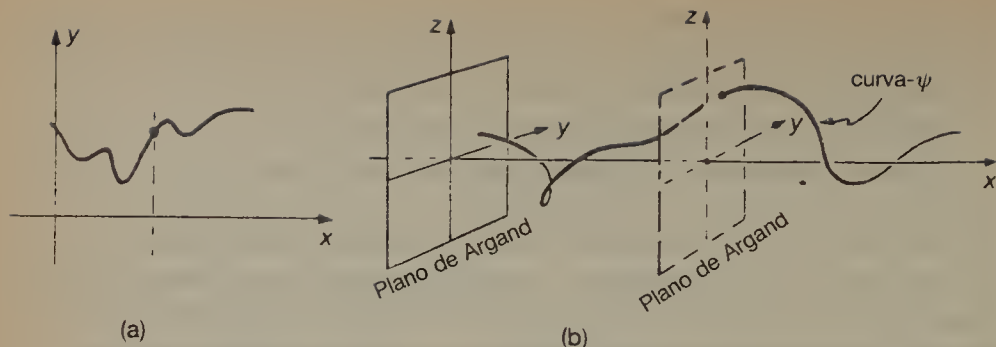


Fig. 6.10. a) Gráfica de una función real de una variable real x . b) Gráfica de una función compleja ψ de una variable real x .

cualquier línea particular en el espacio físico podemos hacerlo simplemente tomando nuestro eje- x a lo largo de esta línea y utilizando provisionalmente las otras dos direcciones espaciales para proporcionar los planos de Argand requeridos. Esto se mostrará útil para nuestra comprensión del experimento de la doble-rendija.

Como mencioné antes, en física clásica necesitamos conocer la velocidad (o el momento) de una partícula para determinar lo que va a hacer a continuación. Ahora, la mecánica cuántica nos permite una notable economía. ¡La función de onda ψ contiene ya las diversas amplitudes de los diferentes momentos posibles! (Algunos lectores contrariados pueden estar pensando que «ya es hora» de un poco de economía, considerando lo mucho que hemos tenido que complicar la simple imagen clásica de una partícula puntual. Aunque tengo mucha simpatía por esos lectores, les advierto que se tomen ávidamente los bocados que se les ofrecen ¡pues lo peor está aún por llegar!) ¿Cómo es que las amplitudes de las velocidades están determinadas por ψ ? En realidad, es mejor pensar en las amplitudes de los momentos. (Recordemos que el momento es la velocidad multiplicada por la masa de la partícula; cf. pp. 215-216.) Lo que se hace es aplicar el llamado *análisis armónico* a la función. Estaría fuera de lugar aquí explicar esto en detalle, pero está estrechamente relacionado con lo que se hace con los sonidos musicales. Una onda de forma cualquiera puede descomponerse como una suma de diferentes «armónicos» (de ahí el término «análisis armónico») que son los tonos puros de las distintas notas (es decir, distintas frecuencias puras). En el caso de la función de onda ψ , los «tonos puros» corresponden a los diferentes valores posibles del momento que pudiera tener la partícula, y el

tamaño de la contribución a ψ de cada «tono puro» proporciona la amplitud de ese valor del momento. Los mismos «tonos puros» se denominan *estados de momento*.

¿Qué apariencia tiene un estado de momento como curva- ψ ? Tiene el aspecto de un *sacacorchos*, para el que el nombre matemático oficial es una *hélice* (Fig. 6.11) *. Aquellos sacacorchos que están enroscados apretadamente corresponden a momentos grandes, y los que apenas se enroscan dan momentos muy pequeños. Hay un caso límite en el que no se enroscan en absoluto y la curva- ψ es una línea recta: el caso de momento cero. La famosa *relación de Planck* está implícita en esto. Un paso de rosca pequeño significa corta longitud de onda y alta *frecuencia* y, por lo tanto, alto momento y alta *energía*; y un paso de rosca grande significa baja frecuencia y baja energía, siendo la energía E siempre proporcional a la frecuencia ν ($E = h\nu$). Si los planos de Argand están orientados de la forma normal (es decir, con la descripción x,y,z dada arriba, con los usuales ejes dextrógiros), entonces los momentos que apuntan en la dirección positiva del eje- x corresponden a los sacacorchos destrógiros (que son los sacacorchos de tipo usual).

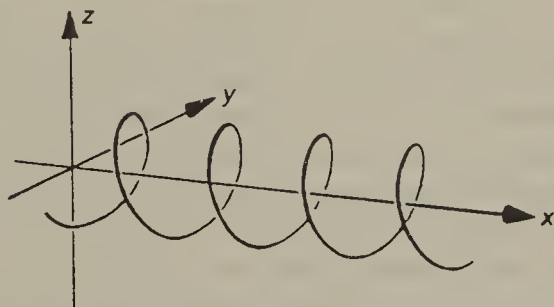


Fig. 6.11. Un estado de momento tiene una curva- ψ que es un sacacorchos.

A veces es útil describir los estados cuánticos no en términos de funciones de onda ordinarias, como se hizo arriba, sino en términos de funciones de onda del *momento*. Esto equivale a considerar la descomposición de ψ en términos de los diversos estados de momento

* En términos de una descripción analítica más estándar, cada uno de nuestros sacacorchos (es decir, estados de momento), vendría dado por una expresión $\psi = e^{ipx/\hbar} = \cos (px/\hbar) + i \sin (px/\hbar)$ (ver Capítulo 3, p. 125, donde p es el valor del momento en cuestión).

y construir una nueva función $\tilde{\psi}$, que esta vez es una función del momento p en lugar de la posición x , y cuyo valor $\tilde{\psi}(p)$, para cada p , da el tamaño de la contribución del estado de momento p a ψ . (El espacio de los p se llama *espacio de momentos*.) La interpretación de $\tilde{\psi}$ es que, para cada particular elección de p , el número complejo $\tilde{\psi}(p)$ da la *amplitud de que la partícula tenga un momento p* .

Hay un nombre matemático para la relación entre las funciones ψ y $\tilde{\psi}$. Estas funciones se llaman *transformadas de Fourier* una de otra —por el ingeniero/matemático francés Joseph Fourier (1768-1830). Sólo haré aquí unos pocos comentarios sobre esta relación. El primer punto es que existe una notable simetría entre ψ y $\tilde{\psi}$. Para regresar a ψ desde $\tilde{\psi}$, aplicamos efectivamente el mismo procedimiento que aplicamos para obtener $\tilde{\psi}$ a partir de ψ . Ahora es $\tilde{\psi}$ la que se somete a análisis armónico, los «tonos puros» (es decir, los sacacorchos en la imagen del espacio de momentos) se llaman ahora *estados de posición*. Cada posición x determina dicho «tono puro» en el espacio de momentos, y el tamaño de la contribución a $\tilde{\psi}$ de este «tono puro» da el valor $\psi(x)$.

Un estado de posición corresponde, en la imagen ordinaria del espacio de posiciones, a una función ψ que tiene un pico muy agudo en el valor de x en cuestión, siendo nulas todas las demás amplitudes para otros valores diferentes del x dado. Semejante función se denomina una *función delta* (de Dirac) —aunque técnicamente no es exactamente una «función» en el sentido ordinario ya que toma un valor infinito en x . Análogamente, los estados de momento (sacacorchos en la imagen del espacio de posiciones) dan funciones delta en la imagen del espacio de momentos (ver Fig. 6.12) ¡Vemos así que la transformada de Fourier de un sacacorchos es una función delta, y *viceversa*!

La descripción en el espacio de posiciones es útil cuando intentamos realizar una medida de la posición de la partícula, lo que equivale a hacer algo que amplifique los efectos de las diferentes posiciones posibles de la partícula hasta el nivel clásico. (Hablando en términos generales, las fotocélulas y las placas fotográficas realizan medidas de posición para los fotones.) La descripción en el espacio de momentos es útil cuando nos proponemos medir el momento de la partícula, es decir, amplificar los efectos de los diferentes momentos posibles hasta el nivel clásico. (Pueden utilizarse los efectos de retroceso o de difracción en cristales para medidas del momento.) En cada caso, el módulo al cuadrado de la función de onda correspondiente (ψ o $\tilde{\psi}$) da la probabilidad requerida para el resultado de la medida requerida.

Terminaremos esta sección volviendo una vez más al experimento

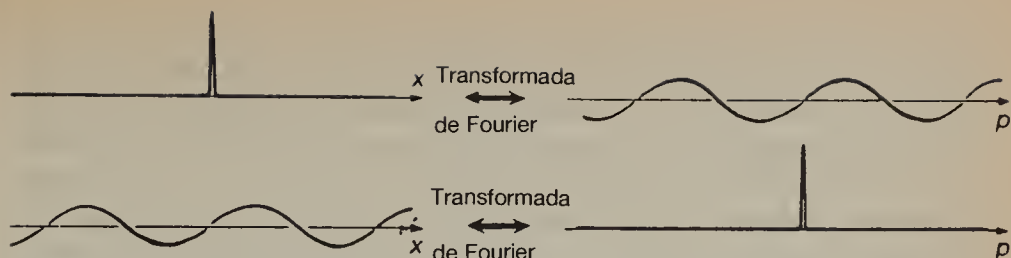


Fig. 6.12. Funciones delta en el espacio de posiciones corresponden a sacacorchos en el espacio de momentos y *viceversa*.

de la doble-rendija. Hemos aprendido que, según la mecánica cuántica, incluso una simple partícula debe comportarse por sí misma como toda una onda. Esta onda viene descrita por la función de onda ψ . Las ondas más uniformemente «onduladas» son los estados de momento. En el experimento de la doble rendija suponíamos fotones de una frecuencia definida; de este modo la función de onda del fotón está compuesta de estados de momento en diferentes direcciones, en los que el paso de rosca es el mismo para todos los sacacorchos y esta distancia es la *longitud de onda*. (La longitud de onda está fijada por la frecuencia.)

Cada longitud de onda de un fotón se extiende inicialmente desde la fuente f y (si no se hace ninguna detección en las rendijas) atraviesa ambas rendijas en su camino hasta la pantalla. Sin embargo, sólo una pequeña parte de esta función de onda emerge de las rendijas, y consideramos que cada rendija actúa como una nueva fuente a partir de la que se extienden funciones de onda independientes. Estas dos porciones de la función de onda interfieren entre sí de modo que, cuando alcanzan la pantalla, hay lugares en donde las dos porciones se suman y lugares en donde se cancelan. Para descubrir dónde se suman las ondas y dónde se cancelan, tomemos algún punto p de la pantalla y examinemos las líneas rectas que van a p desde cada una de las dos rendijas s e i . A lo largo de la línea sp tenemos un sacacorchos, y a lo largo de la línea ip tenemos otro sacacorchos. (Tenemos también sacacorchos a lo largo de las líneas fs y fi , pero si suponemos que la fuente está a la misma distancia de las dos rendijas, entonces *en* las rendijas los sacacorchos habrán girado la misma cantidad.) Ahora las cantidades que han girado los sacacorchos cuando alcanzan la pantalla en p dependerán de las longitudes de las líneas sp e ip . Cuando estas longitudes difieran en un número entero de longitudes de onda, entonces, en p , ambos sacacorchos tendrán el

mismo desplazamiento respecto a sus ejes (esto es, $\theta = 0$, donde el ángulo θ es el mismo que en la sección previa), de modo que las amplitudes respectivas se sumarán y tendremos un punto *brillante*. Cuando estas longitudes difieran en un número entero de longitudes de onda más media de longitud de onda, entonces, en p , los desplazamientos de ambos sacacorchos respecto a sus ejes serán *opuestos* ($\theta = 180^\circ$), de modo que las amplitudes respectivas se cancelarán y obtenemos un punto *oscuro*. En todos los demás casos habrá algún ángulo entre los desplazamientos de los sacacorchos cuando alcancen p , de modo que las amplitudes se suman de una forma intermedia y tenemos una región de intensidad intermedia (ver Fig. 6.13).

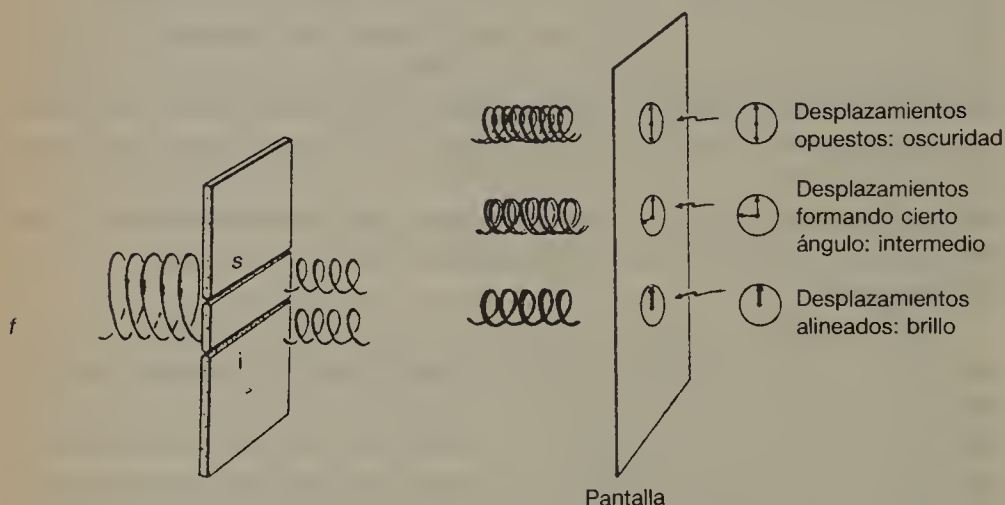


Fig. 6.13. El experimento de la doble rendija analizado en términos de la descripción de los estados de momento del fotón como sacacorchos.

El principio de incertidumbre

La mayoría de los lectores habrá oído hablar del *principio de incertidumbre* de Heisenberg. Según este principio no es posible medir (es decir, amplificar hasta el nivel clásico) con precisión la posición y el momento de una partícula al mismo tiempo. Peor aún: existe un *límite absoluto* para el producto de estas precisiones, llamémoslas Δx y Δp respectivamente, que viene dado por la relación

$$\Delta x \times \Delta p \geq \hbar.$$

Esta fórmula nos dice que cuanto más precisa sea la medida de la posición x , con menor precisión se puede determinar el momento p , y *viceversa*. Si la posición fuera medida con precisión *infinita*, entonces el momento quedaría *totalmente* indeterminado; por el contrario, si el momento se mide exactamente, entonces la localización de la partícula queda totalmente indeterminada. Para hacernos una idea del tamaño del límite dado por la relación de Heisenberg, supongamos que la posición de un electrón se ha medido con la precisión del nanometro (10^{-9} m); entonces el momento se haría tan indeterminado que no podríamos esperar que, un segundo después, el electrón estuviera ¡a menos de 100 kilómetros de distancia!

En algunas descripciones, se nos lleva a creer que esto se debe simplemente a algún tipo de torpeza inherente al proceso de medida. Consiguientemente, en el caso del electrón recién considerado, el intento de localizarlo le da inevitablemente, según este modo de verlo, un «empujón» aleatorio de tal intensidad que el electrón sale despedido probablemente a una gran velocidad del orden de magnitud indicado por el principio de Heisenberg. En otras descripciones se nos dice que la incertidumbre es una propiedad de la propia partícula, y su movimiento tiene una aleatoriedad inherente, lo que significa que su comportamiento es intrínsecamente impredecible en el nivel cuántico. Todavía hay otras exposiciones en las que se nos informa que una partícula cuántica es algo incomprensible para lo que los mismos conceptos de posición clásica y momento clásico son inaplicables. Yo no me siento a gusto con ninguna de estas. La primera es algo confusa; la segunda es ciertamente errónea; y la tercera, indebidamente pesimista.

¿Qué nos dice realmente la descripción de la función de onda? Recordemos primero nuestra descripción de un estado de momento. Éste es el caso en el que el momento está especificado exactamente. La curva- ψ es un sacacorchos que permanece enroscado siempre a la misma distancia del eje. Por consiguiente, las amplitudes para los diferentes valores de la posición tienen todos los mismos módulos al cuadrado. Así, si se realiza una medida de la posición la probabilidad de encontrar la partícula en un punto es la misma que la de encontrarla en otro cualquiera. ¡Realmente la posición de la partícula está completamente indeterminada! ¿Qué hay sobre el estado de posición? En este caso, la curva- ψ es una función delta. La partícula está localizada exactamente —en la posición del pico de la función delta— siendo nulas las amplitudes para las otras posiciones. Las amplitudes de momento se obtienen mejor considerando la descripción en el es-

pacio de momentos, donde es ahora la curva- $\tilde{\psi}$ la que es un sacacorchos y ahora son las diferentes amplitudes de momento las que tienen igual módulo al cuadrado. ¡Al realizar una medida del momento de la partícula el resultado estaría completamente indeterminado!

Es interesante examinar un caso intermedio en el que las posiciones y los momentos están ambos sólo parcialmente limitados, aunque necesariamente sólo hasta un grado consistente con la relación de Heisenberg. La curva- ψ y la correspondiente curva- $\tilde{\psi}$ para este caso (transformadas de Fourier una de otra) se ilustran en la Figura 6.14. Nótese que la distancia de cada curva al eje es apreciable sólo en una región bastante pequeña; lejos de ésta, la curva se ciñe muy estrechamente al eje. Esto significa que los módulos al cuadrado son de tamaño apreciable sólo en una región muy limitada, tanto en el espacio de posiciones como en el espacio de momentos. Por lo tanto, la partícula puede estar algo localizada en el espacio pero hay una cierta dispersión; análogamente el momento está bastante definido, de modo que la partícula se mueve con una velocidad bastante definida y la dispersión de las posibles posiciones no crece demasiado rápidamente con el tiempo. Un estado cuántico de este tipo se conoce como *paquete de ondas*; frecuentemente se toma como la mejor aproximación en teoría cuántica a una partícula clásica. Sin embargo, la dispersión en los valores del momento (es decir, de la velocidad) implica que un paquete de ondas se *dispersará* con el tiempo. Cuanto más localizada esté la posición de partida, más rápidamente se ensanchará.



Fig. 6.14. Paquetes de onda. Éstos están localizados tanto en el espacio de posiciones como en el de momentos.

Los procedimientos de evolución de U y R

En esta descripción del desarrollo temporal de un paquete de ondas está implícita la *ecuación de Schrödinger* que nos dice cómo evo-

luciona realmente la función de onda con el tiempo. En efecto, lo que dice la ecuación de Schrödinger es que si descomponemos ψ en estados de momento («tonos puros») entonces cada una de las componentes individuales se moverá con una velocidad que es c^2 dividida por la velocidad de una partícula clásica que tuviera el momento en cuestión. De hecho, la ecuación matemática de Schrödinger se escribe de forma más concisa que esto. Veremos su forma exacta más adelante. Se asemeja algo a las ecuaciones de Hamilton o de Maxwell (teniendo estrechas relaciones con ambas) y, como aquellas ecuaciones, ¡da una evolución *completamente determinista* de la función de onda una vez que esta función de onda se especifica para un instante cualquiera! (Ver p. 363).

Considerando que ψ describe la «realidad» del mundo, en tanto que está gobernada por la ecuación determinista de Schrödinger no tenemos nada de este indeterminismo que se supone es una característica inherente a la teoría cuántica. Nos referiremos a este proceso como el proceso de evolución **U**. Sin embargo, cuandoquiera que «hacemos una medida», amplificando los efectos cuánticos hasta el nivel clásico, cambiamos las reglas. Ahora *no* utilizamos **U** sino que en su lugar adoptamos un procedimiento completamente diferente, que llamaré **R**, que consiste en formar los cuadrados de los módulos de las amplitudes cuánticas ¡para obtener probabilidades clásicas! ⁴ Es el procedimiento **R**, y *sólo* **R**, el que introduce incertidumbres y probabilidades en la teoría cuántica.

El proceso determinista **U** parece ser la parte de la teoría cuántica de mayor interés para el trabajo de los físicos; no obstante, los filósofos están más intrigados por la *reducción **R** del vector de estado* (o, como a veces se le describe gráficamente: *colapso de la función de onda*) no determinista. Ya consideremos **R** simplemente como un cambio en el «conocimiento» disponible del sistema, o ya lo tomemos (como yo lo hago) como algo «real», disponemos realmente de dos modos matemáticos completamente *distintos* para describir cómo cambia con el tiempo el vector de estado de un sistema físico. En efecto, **U** es completamente determinista, mientras que **R** es una ley probabilista; **U** mantiene la superposición compleja cuántica, pero **R** la viola totalmente; **U** actúa de una forma continua, pero **R** es descaradamente discontinuo. Según los procedimientos estándar de la mecánica cuántica no hay ninguna posibilidad de «deducir» **R** como un ejemplo complicado de **U**. Es sencillamente un procedimiento *diferente* de **U**, que proporciona la otra «mitad» de la interpretación del formalismo cuántico. Todo el no-determinismo de la teoría procede de **R** y no de

U. Ambos U y R son necesarios para todos los maravillosos acuerdos que hay entre la teoría cuántica y los hechos observacionales.

Volvamos a nuestra función de onda ψ . Supongamos que es un estado de momento. Seguirá siendo ese mismo estado de momento durante el resto del tiempo mientras la partícula no interaccione con algo. (Esto es lo que nos dice la ecuación de Schrödinger.) En cualquier instante que decidamos «medir su momento» seguiremos teniendo la misma respuesta definida. No hay probabilidades aquí. La predecibilidad es tan estricta como en la teoría clásica. Sin embargo, supongamos que en alguna etapa decidimos medir (es decir, amplificar hasta el nivel clásico) la posición de la partícula. Ahora encontramos que se nos presenta una colección de amplitudes de probabilidad cuyos módulos debemos elevar al cuadrado. En este punto dominan las probabilidades, y hay una completa incertidumbre sobre el resultado que va a dar la medida. Esta incertidumbre está en acuerdo con el principio de Heisenberg.

Supongamos, por el contrario, que ψ parte de un estado de posición (o muy próxima a un estado de posición). Ahora, la ecuación de Schrödinger nos dice que *no* permanecerá en un estado de posición, sino que se dispersará rápidamente. De todas formas, el *modo* en que se dispersa está completamente fijado por esta ecuación. No hay nada indeterminado o probabilístico en este comportamiento. En principio existirían experimentos que podrían verificar este hecho. (Más tarde veremos más sobre esto.) Pero si imprudentemente decidimos medir el momento, entonces encontramos amplitudes para todos los diferentes valores posibles del momento que tienen iguales módulos al cuadrado, y hay incertidumbre completa sobre el resultado del experimento —de nuevo en acuerdo con el principio de Heisenberg.

Análogamente, si ψ parte como un paquete de ondas, su evolución futura está completamente fijada por la ecuación de Schrödinger y, en principio, se podrían idear experimentos para seguir la pista de este hecho. Pero en cuanto decidamos medir la partícula de una forma *diferente* de ésta —por ejemplo, medir su posición o momento— entonces encontramos que se introducen esas incertidumbres, de nuevo en acuerdo con el principio de Heisenberg, con probabilidades dadas por los cuadrados de los módulos de las amplitudes.

Indudablemente esto es muy extraño y misterioso; pero no es una imagen incomprensible del mundo. Hay mucho en esta imagen que está gobernado por leyes muy claras y precisas. No hay, sin embargo, una regla clara, por ahora, sobre cuándo deberíamos invocar la regla probabilística R en lugar de la determinista U. ¿En qué consiste el

«hacer una medida»? ¿Por qué (y cuándo) los cuadrados de los módulos de las amplitudes «se hacen probabilidades»? ¿Puede entenderse el «nivel clásico» mecano-cuánticamente? Éstas son preguntas profundas y enigmáticas que serán abordadas más adelante en este capítulo.

¿Partículas en dos lugares a la vez?

En las descripciones anteriores he adoptado una visión bastante más «realista» de la función de onda de lo que es quizá usual entre los físicos cuánticos. He adoptado el punto de vista de que el estado «objetivamente real» de una partícula viene realmente descrito por su función de onda ψ . Parece que mucha gente encuentra que ésta es una postura a la que es difícil adherirse seriamente. Un motivo para ello parece ser el que esta postura implica que consideremos que las partículas individuales tienen una extensión espacial en lugar de estar siempre concentradas en puntos simples. Para un estado de momento, esta dispersión llega a su punto más extremo puesto que ψ está uniformemente distribuida sobre la totalidad del espacio. Más que pensar que la propia partícula se extiende en el espacio la gente prefiere pensar que su posición está «completamente indeterminada», de modo que todo lo que se puede decir sobre su posición es que es tan probable que la partícula esté en un lugar como que esté en cualquier otro. Sin embargo, hemos visto que la función de onda no proporciona simplemente una distribución de probabilidad para las distintas posiciones: proporciona una distribución de *amplitudes* para las diferentes posiciones. Si conocemos esta distribución de amplitudes (es decir, la función ψ), entonces conocemos —a partir de la ecuación de Schrödinger— la forma precisa en que evolucionará el estado de la partícula de un instante a otro. Necesitamos esta idea de partícula «extendida» para que su «movimiento» (esto es, la evolución de ψ en el tiempo) esté determinado; y si *aceptamos* esta idea, vemos que el movimiento de la partícula *está* determinado de forma precisa. La «concepción probabilista» con respecto a $\psi(x)$ se haría apropiada si realizáramos una medida de la posición de la partícula y $\psi(x)$ se utilizara *entonces* sólo en su forma como un módulo al cuadrado: $|\psi(x)|^2$.

Parece que tenemos que aceptar esta imagen de una partícula que puede extenderse sobre grandes regiones del espacio y que es probable que permanezca extendida hasta que se lleve a cabo la siguiente medida de posición. Incluso cuando está localizada como un estado

de posición, una partícula empieza a dispersarse en el instante siguiente. Un estado de momento puede parecer difícil de aceptar como una imagen de la «realidad» de la existencia de la partícula, pero es quizá aún más difícil aceptar como «real» el estado con *dos picos* que ocurre cuando la partícula emerge inmediatamente después de atravesar un par de rendijas (Fig. 6.15). En la dirección vertical la forma de la función de onda ψ tendría un pico agudo en cada una de las rendijas, siendo la suma * de una función de onda ψ_s , que tiene un pico en la rendija superior, y ψ_i , que tiene un pico en la rendija inferior:

$$\psi(x) = \psi_s(x) + \psi_i(x)$$

Si tomamos ψ representando la «realidad» del estado de la partícula, entonces ¡debemos aceptar que la partícula «está» de hecho en *dos* lugares a la vez! Según esta concepción la partícula *ha atravesado realmente ambas rendijas a la vez*.

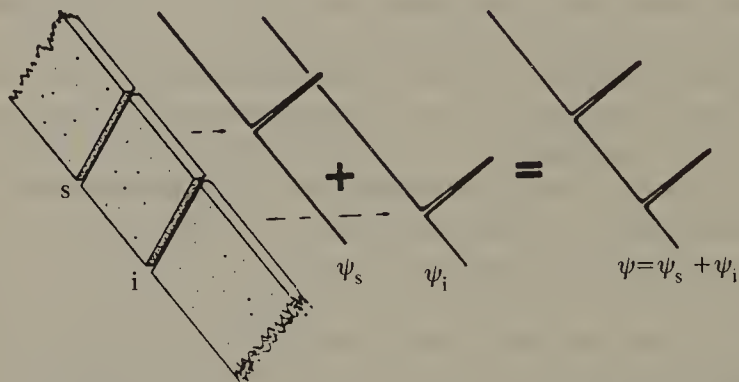


Fig. 6.15. La función de onda del fotón tiene picos en dos lugares a la vez cuando emerge del par de rendijas.

Recuérdese la objeción estándar a la idea de que la partícula «atravesara ambas rendijas a la vez»: si realizamos una medida *en* las rendijas para determinar cuál de las rendijas ha atravesado, encontraremos siempre que la partícula *entera* está en una u otra de las rendijas. Pero esto aparece así debido a que estamos realizando una *medida de posición* sobre la partícula y, por lo tanto, ψ proporciona *ahora* sim-

* La descripción mecano-cuántica más usual dividiría esta suma por un factor de normalización —en este caso $\sqrt{2}$, para dar $(\psi_s + \psi_i)/\sqrt{2}$ — pero no hay necesidad por ahora de complicar la descripción.

plemente una distribución de probabilidades $|\psi|^2$ para la posición de la partícula de acuerdo con el procedimiento del módulo-cuadrado, y encontramos que está sólo en un lugar u otro. Pero existen varios tipos de medida que *podríamos* realizar en las rendijas, además de las medidas de posición. Para aquéllas, necesitaríamos conocer la función de onda ψ con sus dos picos, y no sólo $|\psi|^2$, para diferentes posiciones x . Semejante medida podría distinguir el estado con dos picos

$$\psi = \psi_s + \psi_i,$$

dado arriba, de otros estados de dos picos, tales como

$$\psi_s - \psi_i$$

o

$$\psi_s + i\psi_i.$$

(Ver Fig. 6.16, para las curvas- ψ en cada uno de estos diferentes casos.) ¡Puesto que existen medidas que distinguen estas diversas posibilidades, todas estas deben ser *diferentes* modos «reales» posibles en que puede existir la partícula!

Las rendijas no tienen que estar próximas para que un fotón las atravesase «ambas a la vez». Para ver que una partícula cuántica puede estar en «dos lugares a la vez», no importa cuál sea la distancia entre estos, consideramos un montaje experimental un poco diferente del experimento de la doble-rendija. Como antes, tenemos una lámpara que emite luz monocromática de fotón en fotón; pero en lugar de dejar que la luz atravesase un par de rendijas la reflejamos en un espejo semireflectante que forma un ángulo de 45° con la dirección del haz luminoso. (Un espejo semi-reflectante es un espejo que refleja

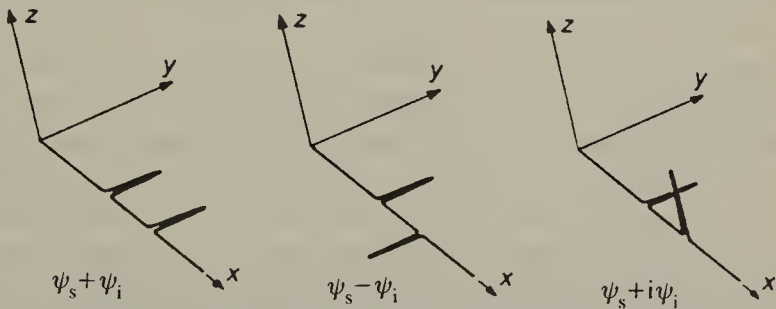


Fig. 6.16. Tres formas diferentes en las que una función de onda de un fotón puede tener un doble pico.

exactamente la mitad de la luz que incide sobre él, mientras que la mitad restante se transmite directamente a través del espejo.) Después de su encuentro con el espejo la función de onda del fotón se descompone en dos, con una parte reflejada lateralmente y la otra que continúa en la misma dirección de partida del fotón. La función de onda tiene otra vez dos picos, pero ahora los dos picos están mucho más separados, un pico que describe el fotón reflejado y el otro pico que describe el fotón transmitido. (Ver Fig. 6.17). Además, a medida que pasa el tiempo la separación entre los picos se hace cada vez mayor y crece sin límite. Imaginemos que las dos porciones de la función de onda escapan al espacio y que esperamos durante todo un año. Entonces los dos picos de la función de onda del fotón estarán separados más de un año-luz. De algún modo, ¡el fotón se ha encontrado estando en dos lugares a la vez, a más de un año-luz de distancia!

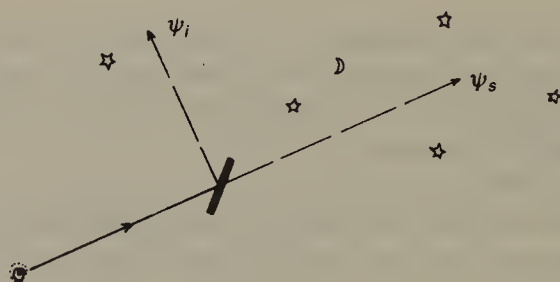


Fig. 6.17. Los dos picos de una función de onda con doble pico podrían estar a años-luz de distancia. Esto puede conseguirse por medio de un espejo semi-reflectante.

¿Existe alguna razón para tomar esta imagen en serio? ¿No podemos considerar simplemente que el fotón tiene un 50 % de probabilidades de estar en uno de los lugares y un 50 % de probabilidades de estar en el otro? ¡No, no podemos! Por mucho que haya viajado el fotón, siempre existe la posibilidad de que las dos partes del haz de fotones puedan ser reflejadas hacia atrás de modo que vuelvan a encontrarse y produzcan efectos de interferencia que no podrían resultar de una ponderación probabilística de las dos alternativas. Supongamos que cada parte del haz encuentra un espejo totalmente reflectante, colocado con el ángulo apropiado para que los rayos se junten, y en el punto de encuentro se coloca otro espejo semi-reflectante, con el mismo ángulo que el primero. En las direcciones finales

de los dos haces se colocan dos fotocélulas (ver Fig. 6.18) ¿Qué encontramos? Si fuera simplemente el caso de que hubiera un 50 % de probabilidades de que el fotón siguiera una ruta y un 50 % de probabilidades de que siguiera la otra, entonces deberíamos encontrar un 50 % de probabilidades de que uno de los detectores registre el fotón y un 50 % de probabilidades de que lo haga el otro detector. Sin embargo, no es esto lo que sucede. Si las dos posibles rutas tienen exactamente la misma longitud entonces resulta que hay un 100 % de probabilidades de que el fotón llegue al detector A, colocado en la dirección del movimiento inicial del fotón, y un 0 % de probabilidades de que llegue al otro detector B; ¡el fotón incidirá con *certeza* en el detector A! (Podemos verlo utilizando la descripción del sacacorchos dada antes, como en el caso del experimento de la doble-rendija.)

Por supuesto, nunca se ha llevado a cabo semejante experimento con caminos de una longitud del orden de un año-luz, pero nadie (entre los físicos cuánticos convencionales) pone en duda el resultado mencionado. En realidad se han llevado a cabo experimentos de este tipo con caminos de una longitud de varios metros, y los resultados están en completo acuerdo con las predicciones mecano-cuánticas (cf. Wheeler, 1983). ¿Qué nos dice esto acerca de la *realidad* del estado de existencia del fotón entre su primer y su último encuentro con el espejo semi-reflectante? ¡Parece inevitable que, en algún sentido, el fotón debe haber *recorrido* ambas rutas al mismo tiempo! En efecto,

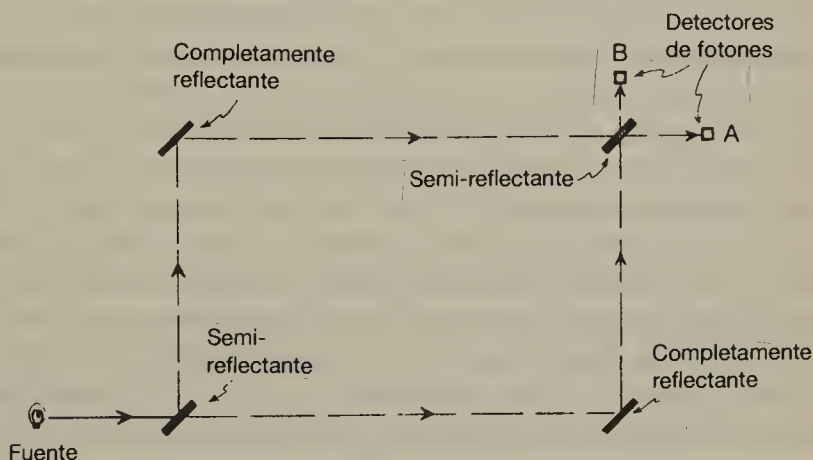


Fig. 6.18. Los dos picos de una función con doble pico no pueden ser considerados simplemente como probabilidades ponderadas de que el fotón tenga una localización u otra. Es posible hacer que las dos rutas que toma el fotón interfieran mutuamente.

si se colocara una pantalla absorbente en cualquiera de las dos rutas entonces se haría igualmente probable que se alcance A o B; pero cuando ambas rutas están abiertas (y tienen la misma longitud) sólo puede alcanzarse A. ¡El bloqueo de una de las rutas *hace posible* realmente que se alcance B! Con ambas rutas abiertas, el fotón «sabe» de alguna manera que *no* está permitido llegar a B, así que efectivamente debe tener alguna idea de ambas rutas.

La concepción de Niels Bohr de que no puede asociarse ningún «significado» objetivo a la existencia del fotón entre los instantes de las medidas me parece una idea demasiado pesimista para adoptar con respecto a la realidad del estado del fotón. La mecánica cuántica nos da una *función de onda* para describir la «realidad» de la posición del fotón, y la función de onda del fotón entre los espejos semireflectantes es simplemente un estado con dos picos cuya distancia puede ser a veces muy considerable.

Notamos, también, que el simple «estar en dos lugares concretos a la vez» no es una descripción completa del estado del fotón: necesitamos poder distinguir el estado $\psi_s + \psi_i$ del estado $\psi_s - \psi_i$, pongamos por caso, (o de $\psi_s + i\psi_i$, pongamos por caso), donde ψ_s y ψ_i se refieren ahora a las posiciones del fotón en cada una de las dos rutas (¡ahora «transmitida» y «reflejada» respectivamente!). Es este tipo de distinción el que determina si el fotón, cuando llega al último espejo semi-reflectante, alcanzará A con certeza o alcanzará B con certeza (o alcanzará A o B con una probabilidad intermedia).

Esta característica enigmática de la realidad cuántica —es decir, que debemos tomar en serio que, de varias formas (¡diferentes!) una partícula pueda «estar en dos lugares a la vez»— surge del hecho de que debemos poder sumar estados cuánticos, utilizando pesos probabilísticos complejos, para obtener otros estados cuánticos. Este tipo de superposición de estados es una característica general —e importante— de la mecánica cuántica, conocida como *superposición lineal cuántica*; es la que nos permite componer estados de momento a partir de estados de posición, o estados de posición a partir de estados de momento. En estos casos la superposición lineal se aplica a una colección *infinita* de estados diferentes, es decir, a todos los diferentes estados de posición, o a todos los diferentes estados de momento. Pero, como veremos, la superposición lineal cuántica resulta bastante enigmática cuando se aplica a sólo un *par* de estados. Las reglas son que dos estados *cualesquiera*, independientemente de lo distintos que puedan ser, pueden coexistir en cualquier superposición lineal compleja. En realidad, cualquier objeto físico, constituido él mismo por

partículas individuales, debería poder existir en semejante superposición de estados ampliamente separados en el espacio y, por lo tanto, ¡«estar en dos lugares a la vez»! El formalismo de la mecánica cuántica no hace distinción, a este respecto, entre partículas simples y complicados sistemas de muchas partículas. ¿Por qué, entonces, no tenemos experiencia de que cuerpos macroscópicos, digamos bolas de cricket, o incluso personas, tengan dos localizaciones completamente diferentes al mismo tiempo? Ésta es una cuestión profunda, y la teoría cuántica actual no nos proporciona una respuesta realmente satisfactoria. Para un objeto tan substancial como una bola de cricket debemos considerar que el sistema está «en el nivel clásico» —o, como se suele expresar normalmente, habrá que hacer una «observación» o «medida» sobre la bola de cricket— y, por tanto, las amplitudes de probabilidad complejas que ponderan nuestra superposición lineal deben ahora tener sus módulos elevados al cuadrado y ser tratadas como probabilidades que describen las alternativas reales. ¡Sin embargo, esta realidad supone una petición de principio del *por qué* podemos cambiar así nuestras reglas cuánticas de $\mathbf{U}$ a $\mathbf{R}$! Volveré más tarde a este punto.

Espacio de Hilbert

Recuérdese que en el Capítulo 5 se introdujo el concepto de *espacio de fases* para la descripción de un sistema físico. Un punto simple del espacio de fases representaba el estado (clásico) de un sistema físico entero. En la teoría cuántica el concepto análogo apropiado es el de un *espacio de Hilbert*. * Ahora, un simple punto del espacio de Hilbert representa el *estado cuántico* de un sistema entero. Necesitaremos hacernos una idea de la estructura matemática de un espacio de Hilbert. Espero que el lector no se desanime por esto. No hay nada que sea muy complicado matemáticamente en lo que voy a decir, aunque algunas de las ideas pueden resultar poco familiares.

La propiedad más fundamental de un espacio de Hilbert es que constituye lo que se llama un *espacio vectorial* —en realidad, un espacio vectorial *complejo*. Esto significa que podemos *sumar* dos elementos cualesquiera del espacio y obtener otro elemento del espacio;

* David Hilbert, a quien ya hemos encontrado en capítulos anteriores, introdujo este importante concepto —para el caso de infinitas dimensiones— mucho antes del descubrimiento de la mecánica cuántica, ¡y para un propósito completamente diferente!

y también podemos realizar estas sumas con pesos estadísticos complejos. Debemos poder hacerlo así puesto que éstas son las operaciones de la *superposición lineal cuántica* que acabamos de considerar, a saber, las operaciones que nos dan $\psi_s + \psi_i$, $\psi_s - \psi_i$, $\psi_s + i\psi_i$, etc., para el fotón anterior. En esencia, todo lo que queremos decir al utilizar la expresión «espacio vectorial complejo» es que podemos formar sumas ponderadas de este tipo ⁵.

Será conveniente adoptar una notación (debida esencialmente a Dirac) según la cual los elementos del espacio de Hilbert —llamados *vectores de estado*— se representan mediante algún símbolo en un paréntesis angular, tales como $|\psi\rangle$, $|\chi\rangle$, $|\varrho\rangle$, $|1\rangle$, $|2\rangle$, $|3\rangle$, $|n\rangle$, $|\uparrow\rangle$, $|\downarrow\rangle$, $|\rightarrow\rangle$, $|\nearrow\rangle$, etc. Así, estos símbolos representan ahora estados cuánticos. Para la operación de suma de dos vectores de estado escribimos

$$|\psi\rangle + |\chi\rangle$$

y ponderada con números complejos w y z :

$$w|\psi\rangle + z|\chi\rangle$$

(donde $w|\psi\rangle$ significa $w \times |\psi\rangle$, etc.). En consecuencia, las combinaciones anteriores $\psi_s + \psi_i$, $\psi_s - \psi_i$, $\psi_s + i\psi_i$ se escriben ahora como $|\psi_s\rangle + |\psi_i\rangle$, $|\psi_s\rangle - |\psi_i\rangle$, $|\psi_s\rangle + i|\psi_i\rangle$, respectivamente. Podemos también multiplicar sencillamente un *simple* estado $|\psi\rangle$ por un número complejo w para obtener:

$$w|\psi\rangle$$

(Éste es en realidad un caso muy particular de lo anterior, cuando $z = 0$.)

Recuérdese que admitimos considerar combinaciones con pesos estadísticos complejos en las que w y z no necesitan ser las amplitudes de probabilidad reales sino que son simplemente *proporcionales* a dichas amplitudes. En consecuencia, adoptamos la regla de que podemos multiplicar globalmente un vector de estado por un número complejo distinto de cero y el estado físico sigue siendo el mismo. (Los verdaderos valores de w y z cambiarían pero la razón $w:z$ permanecería constante.) Cada uno de los vectores

$$|\psi\rangle, 2|\psi\rangle, -|\psi\rangle, i|\psi\rangle, \sqrt{2}|\psi\rangle, \pi|\psi\rangle, (1 - 3i)|\psi\rangle, \text{etc.}$$

representa el mismo estado físico —como lo hace cualquier $z|\psi\rangle$, con $z \neq 0$. El único elemento del espacio de Hilbert que *no* tiene inter-

pretación como un sistema físico es el vector *nulo* $\mathbf{0}$ (o el *origen* del espacio de Hilbert).

Para tener algún tipo de representación geométrica de todo esto, consideremos primero el concepto más usual de un vector «real». Normalmente visualizamos este vector simplemente como una *flecha* dibujada en un plano o en un espacio tri-dimensional. La suma de dos de estas flechas se obtiene mediante la ley del paralelogramo (ver Fig. 6.19). La operación de multiplicar un vector por un número (real) se obtiene, en términos de la representación de «flechas», multiplicando simplemente la longitud de la flecha por el número en cuestión y manteniendo constante la dirección de la flecha. Si el número por el que multiplicamos es negativo, entonces se invierte la dirección de la flecha; o si el número es cero, obtenemos el vector $\mathbf{0}$ que no tiene dirección. (El vector $\mathbf{0}$ se representa por la «flecha nula» de longitud cero.) Un ejemplo de cantidad vectorial es la fuerza que actúa sobre una partícula. Otros ejemplos son las velocidades, aceleraciones y momentos clásicos. También están los cuadri-vectores momento que consideramos al final del último capítulo. Éstos eran vectores en *cuatro* dimensiones en lugar de dos o tres. Sin embargo, para un espacio de Hilbert necesitamos vectores en dimensiones aún mucho mayores (a menudo infinitas, de hecho, pero ésta no va a ser una consideración importante aquí). Recuerdese que también se utilizaban flechas para representar vectores en el espacio de fases clásico —que ciertamente podía ser de dimensión muy alta. Las «dimensiones» en un espacio

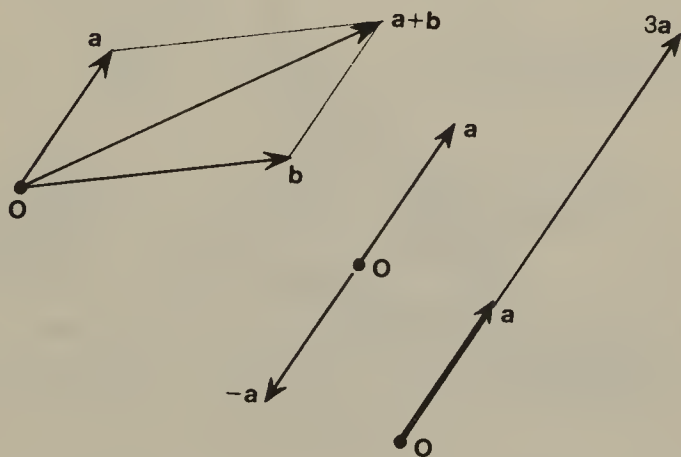


Fig. 6.19. La suma de vectores del espacio de Hilbert y la multiplicación por escalares pueden visualizarse del modo normal como en el caso de los vectores en el espacio ordinario.

de fases no representan direcciones espaciales ordinarias, y tampoco lo hacen las «dimensiones» de un espacio de Hilbert. En lugar de ello, cada dimensión de un espacio de Hilbert corresponde a uno de los diferentes estados físicos independientes de un sistema cuántico.

Debido a la equivalencia entre $|\psi\rangle$ y $z|\psi\rangle$, un estado físico corresponde realmente a toda una *línea a través del origen 0*, o *rayo*, en el espacio de Hilbert (descrita por todos los múltiplos de algún vector), y no simplemente a un vector particular en dicha línea. El rayo consiste en todos los múltiplos posibles de un vector de estado particular. (¡Téngase en cuenta que son múltiplos *complejos*, de modo que la línea es en realidad una línea *compleja*, pero es mejor no preocuparnos ahora por eso!) (Ver Fig. 6.20). Dentro de poco encon-

traremos una elegante representación de este espacio de rayos para el caso de un espacio de Hilbert *bi*-dimensional. Un espacio de Hilbert de infinitas dimensiones aparece incluso en el caso sencillo de la localización de una sola partícula. ¡Existe entonces una dimensión por cada una de las posibles posiciones que pudiera tener la partícula! Cada posición de la partícula define todo un «eje de coordenadas» en el espacio de Hilbert, de modo que con infinitas posiciones individuales diferentes para la partícula tenemos infinitas direcciones independientes diferentes (o «dimensiones») en el espacio de Hilbert. Los estados de momento se representarán también en este *mismo* espacio de Hilbert. Los estados de momento son expresables como combina-

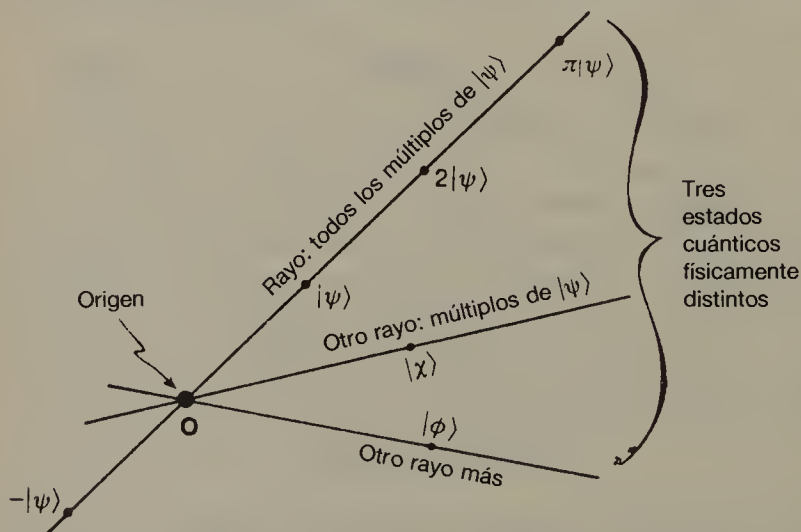


Fig. 6.20. Cada estado cuántico físico está representado por un rayo completo en el espacio de Hilbert.

ciones de estados de posición, de manera que cada estado de momento corresponde a un eje «en diagonal», que está inclinado respecto a los ejes del espacio de posición. El conjunto de todos los estados del momento proporciona un nuevo conjunto de ejes, y el paso de los ejes del espacio de posición al de los ejes del espacio de momento implica una *rotación* en el espacio de Hilbert.

No necesitamos intentar visualizar esto de una manera precisa. ¡Eso no sería razonable! Sin embargo, algunas ideas tomadas de la geometría euclídea ordinaria son muy útiles para nosotros. En particular, los ejes que hemos considerado (*o bien* todos los ejes en el espacio de posición *o bien* todos los ejes en el espacio de momentos) deben considerarse como mutuamente *ortogonales*, es decir, que forman «ángulos rectos» entre sí. La «ortogonalidad» entre rayos es un concepto importante para la mecánica cuántica: rayos ortogonales se refieren a estados que son *independientes* uno de otro. Los diferentes estados de posición posibles de una partícula son todos mutuamente ortogonales, como lo son todos los diferentes estados de momento posibles. Pero los estados de posición no son ortogonales a los estados de momento. La situación se ilustra, muy esquemáticamente, en la Fig. 6.21.

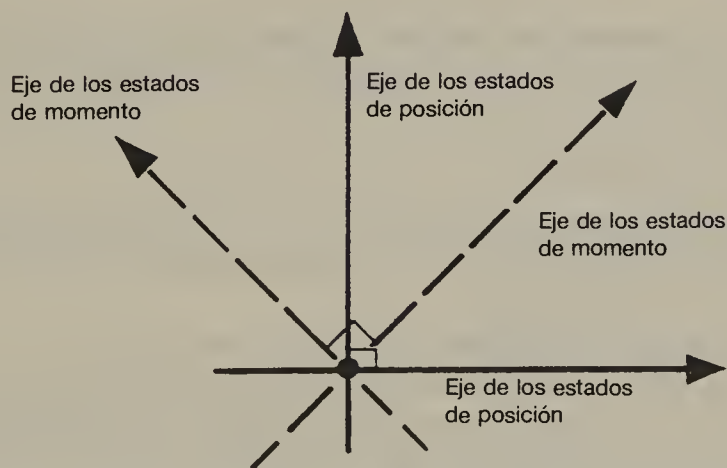


Fig. 6.21. Estados de posición y estados de momento proporcionan diferentes elecciones de ejes ortogonales en el mismo espacio de Hilbert.

Medidas

La regla general **R** para una *medida* (u observación) requiere que los diferentes aspectos de un sistema cuántico que pueden ser ampli-

ficados simultáneamente hasta el nivel clásico —y entre los que el sistema debe entonces elegir— deben ser siempre *ortogonales*. Para una medida *completa*, el conjunto seleccionado de alternativas constituye un conjunto de vectores *base* ortogonales, lo que significa que cualquier vector del espacio de Hilbert puede ser (unívocamente) expresado linealmente en términos de ellos. Para una medida de *posición* —en un sistema que conste de una sola partícula— estos vectores base definirán los ejes de posición que acabamos de considerar. Para el *momento* habrá un conjunto diferente que defina los ejes del momento, y para cada tipo diferente de medida completa habrá otro conjunto más. Tras la medida, el estado del sistema *salta* a uno de los ejes del conjunto determinado por la medida. No hay ley dinámica que nos diga cuál de entre los estados seleccionados elegirá la Naturaleza. Su elección es aleatoria, estando dados los valores de la probabilidad por los cuadrados de los módulos de las amplitudes de probabilidad.

Supongamos que se hace una medida completa sobre un sistema cuyo estado es $|\psi\rangle$, siendo los vectores base para la medida seleccionada

$$|0\rangle, |1\rangle, |2\rangle, |3\rangle, \dots$$

Puesto que éstos forman un conjunto completo, cualquier vector de estado, y en particular $|\psi\rangle$, puede representarse linealmente * en términos de ellos:

$$|\psi\rangle = z_0|0\rangle + z_1|1\rangle + z_2|2\rangle + z_3|3\rangle + \dots$$

Geométricamente, las componentes $z_0, z_1, z_2, \dots$ miden los *tamaños de las proyecciones ortogonales* del vector $|\psi\rangle$ sobre los diversos ejes $|0\rangle, |1\rangle, |2\rangle, \dots$ (ver Fig. 6.22).

Nos gustaría poder interpretar los números complejos $z_0, z_1, z_2, \dots$ como las amplitudes de probabilidad que buscamos, de modo que los cuadrados de sus módulos proporcionen las diversas probabilidades de que el sistema se encuentre, tras la medida, en los respectivos estados $|0\rangle, |1\rangle, |2\rangle, \dots$. Sin embargo, esto no basta tal como están las cosas ya que no hemos fijado las «escalas» de los diversos vectores base $|0\rangle, |1\rangle, |2\rangle, \dots$. Para ello, debemos especificar que son, en cierto sentido, *vectores unitarios* (esto es, vectores de «longitud» unidad), y

* Debe entenderse que está permitida una suma *infinita* de vectores. La definición *completa* de un espacio de Hilbert (que es demasiado técnica para entrar en ella) incluye las reglas concernientes a tales sumas infinitas.

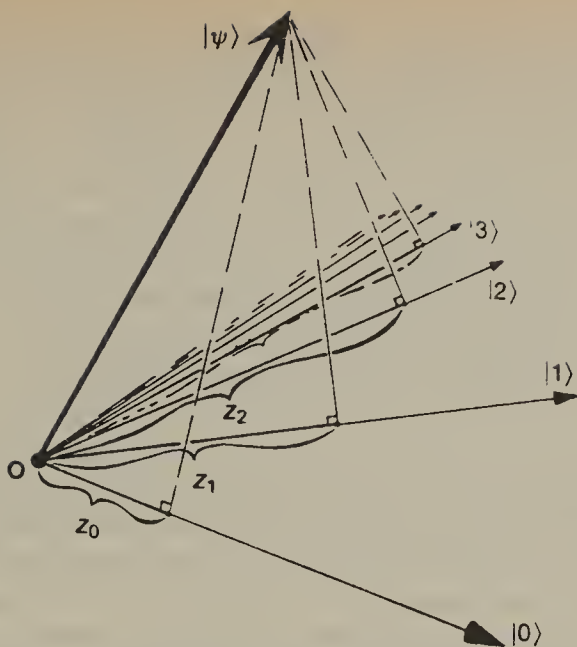


Fig. 6.22. Los tamaños de las proyecciones ortogonales del estado $|\psi\rangle$ sobre los ejes $|0\rangle, |1\rangle, |2\rangle, \dots$ proporcionan las amplitudes requeridas $z_0, z_1, z_2, \dots$

por lo tanto, en terminología matemática, forman lo que se llama una *base ortonormal* (de vectores mutuamente *ortogonales* y *normalizados* a la unidad) ⁶. Si $|\psi\rangle$ está también normalizado para ser un vector unitario, entonces las amplitudes requeridas serán ahora las componentes $z_0, z_1, z_2, \dots$ de $|\psi\rangle$, y las requeridas probabilidades respectivas serán $|z_0|^2, |z_1|^2, |z_2|^2, \dots$. Si $|\psi\rangle$ no es un vector unitario, entonces estos números serán *proporcionales* a las requeridas amplitudes y probabilidades, respectivamente. Las verdaderas amplitudes serán

$$\frac{z_0}{|\psi|}, \frac{z_1}{|\psi|}, \frac{z_2}{|\psi|}, \text{ etc.}$$

y las verdaderas probabilidades

$$\frac{|z_0|^2}{|\psi|^2}, \frac{|z_1|^2}{|\psi|^2}, \frac{|z_2|^2}{|\psi|^2}, \text{ etc.}$$

donde $|\psi|$ es la «longitud» del vector de estado $|\psi\rangle$. Esta «longitud» es un número real positivo definido para cada vector de estado (0 tiene longitud cero), y $|\psi| = 1$ si $|\psi\rangle$ es un vector unitario.

Una medida completa es un tipo de medida muy idealizado. La medida completa de la posición de una partícula, por ejemplo, ¡exi-

giría que fuéramos capaces de localizar la partícula con precisión infinita, en cualquier parte en el Universo! Un tipo de medida más elemental es uno en el que simplemente planteamos una pregunta *sí/no* tal como: «¿está la partícula a la izquierda o a la derecha de cierta línea?» o «¿está el momento de la partícula dentro de cierto intervalo?», etc. Las medidas *sí/no* son en realidad el tipo más fundamental de medida. (Podemos, por ejemplo, estrechar tanto como queramos el margen de la posición de una partícula o del momento utilizando solamente medidas *sí/no*.) Supongamos que el resultado de una medida *sí/no* resulta ser **SÍ**. Entonces el vector de estado debe encontrarse en la región «**SÍ**» del espacio de Hilbert, que llamaré **S**. Si, por el contrario, el resultado de la medida es **NO**, entonces el vector de estado se encuentra en la región «**NO**» del espacio de Hilbert, que llamaré **N**. Las regiones **S** y **N** son completamente ortogonales entre sí, en el sentido de que cualquier vector de estado perteneciente a **S** debe ser ortogonal a cualquier vector de estado perteneciente a **N** (y *viceversa*). Además, cualquier vector de estado $|\psi\rangle$ puede ser expresado (de forma única) como una suma de vectores, uno de cada una de las **S** y **N**. En terminología matemática decimos que **S** y **N** son *complementos ortogonales* uno de otro. Así, $|\psi\rangle$ se expresa unívocamente como

$$|\psi\rangle = |\psi_S\rangle + |\psi_N\rangle$$

donde $|\psi_S\rangle$ pertenece a **S** y $|\psi_N\rangle$ pertenece a **N**. Aquí $|\psi_S\rangle$ es la *proyección ortogonal* del estado $|\psi\rangle$ sobre **S** y, análogamente, $|\psi_N\rangle$ es la proyección ortogonal de $|\psi\rangle$, sobre **N** (ver Fig. 6.23).

Durante la medida, el estado $|\psi\rangle$ *salta* —y se convierte en (proporcional a) $|\psi_S\rangle$ o en $|\psi_N\rangle$. Si el resultado es **SÍ**, entonces salta a $|\psi_S\rangle$, y si es **NO**, salta a $|\psi_N\rangle$. Si $|\psi\rangle$ está normalizado, las probabilidades respectivas de estas dos ocurrencias son las *longitudes al cuadrado*

$$|\psi_S|^2, \quad |\psi_N|^2$$

de los estados proyectados. Si $|\psi\rangle$ no está normalizada, debemos dividir cada una de estas expresiones por $|\psi|^2$. (El «teorema de Pitágoras», $|\psi|^2 = |\psi_S|^2 + |\psi_N|^2$, asegura que la suma de estas probabilidades es la unidad, ¡como debería ser!) Nótese que la probabilidad de que $|\psi\rangle$ salte a $|\psi_S\rangle$ viene dada por la razón en que se reduce el cuadrado de su longitud en esta proyección.

Debe hacerse una puntualización final concerniente a estos «actos de medida» que pueden realizarse en un sistema cuántico. Una con-

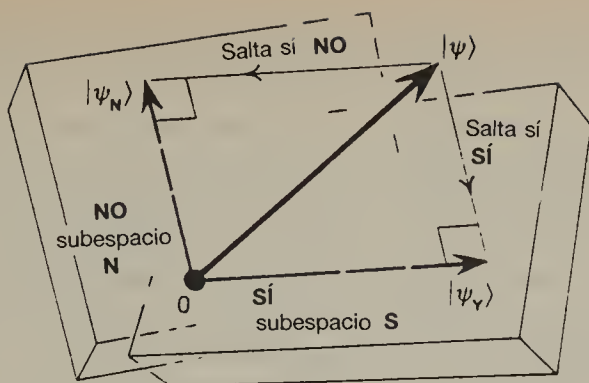


Fig. 6.23. Reducción del vector de estado. Una medida si/no puede describirse en términos de un par de subespacios S y N que son complementos ortogonales entre sí. En la medida, el estado $|\psi\rangle$ salta a su proyección sobre uno u otro de estos subespacios, con probabilidades dadas por el factor en que decrece el cuadrado de la longitud del vector de estado en la proyección.

secuencia de los principios de la teoría es que para *un estado cualquiera* —digamos el estado $|\chi\rangle$ — puede realizarse *en principio* una medida *sí/no*⁷ para la que la respuesta es **SÍ** si el estado medido es (proporcional a) $|\chi\rangle$, y **NO** si es ortogonal a $|\chi\rangle$. Así, la región S anterior podría consistir en todos los múltiplos de un estado escogido $|\chi\rangle$. Esto parece tener la fuerte implicación de que los vectores de estado deben ser *objetivamente reales*. Cualquiera que sea el estado del sistema físico —y llamaremos $|\chi\rangle$ a dicho estado— existe una medida que puede ser realizada *en principio* para la que $|\chi\rangle$ es el *único* estado (salvo proporcionalidad) para el que la medida da el resultado **SÍ** con *certeza*. Para algunos estados $|\psi\rangle$, esta medida sería extremadamente difícil de realizar —quizá «imposible» en la práctica— pero el hecho de que esta medida se podría realizar *en principio*, según la teoría, tendrá algunas consecuencias sorprendentes para nuestra discusión posterior en este capítulo.

El spin y la esfera de estados de Riemann

La cantidad conocida en mecánica cuántica como «*spin*» se considera a veces como la más «mecano-cuántica» de todas las cantidades físicas, así que será prudente que le prestemos cierta atención. ¿Qué es el *spin*? En esencia, es una medida de la rotación de una partícula.

El término «spin» sugiere algo como el giro de una bola de cricket o de béisbol. Recordemos el concepto de *momento angular* que, como la energía y el momento, es una magnitud que se *conserva* (ver Capítulo 5, p. 216, y también p. 293). El momento angular de un cuerpo persiste mientras el cuerpo no sea perturbado por fuerzas de rozamiento o de otro tipo. En esto consiste, en realidad, el spin mecano-cuántico pero ahora es el «giro» de una sola partícula el que nos interesa, no el movimiento orbital de miríadas de partículas individuales en torno a su centro de masas común (que sería el caso de una bola de cricket). Es un hecho físico notable el que la mayoría de las partículas encontradas en la Naturaleza realmente «giran» en este sentido, con una cantidad específica para cada tipo de partícula⁸. Veremos, sin embargo, que el spin de una simple partícula mecano-cuántica tiene algunas propiedades muy peculiares que no son ni mucho menos las que esperaríamos a partir de nuestras experiencias con bolas de cricket que giran y cosas similares.

En primer lugar, la *magnitud* del spin de una partícula es siempre la *misma* para este tipo concreto de partícula. Es solamente la dirección del eje de giro la que puede variar (de una manera muy extraña, como vamos a ver). ¡Esto está en absoluto contraste con una bola de cricket, que puede girar de cualquier manera y en diferentes cantidades dependiendo de cómo sea lanzada! Para un electrón, protón o neutrón la cantidad de spin es siempre $\hbar/2$, que es sólo la *mitad* del menor valor positivo que admitía originalmente Bohr para sus momentos angulares cuantizados de los átomos. (Recordemos que estos valores eran $0, \hbar, 2\hbar, 3\hbar, \dots$) Aquí requerimos la mitad de la unidad básica $\hbar$ —y, en cierto sentido, $\hbar/2$ es ella misma la unidad básica más fundamental. Esta cantidad de momento angular no estaría permitida para un objeto compuesto únicamente de un cierto número de partículas orbitando sin que ninguna de ellas estuviese girando sobre sí misma; sólo puede aparecer debido a que el spin es una propiedad *intrínseca* de la propia partícula (es decir, que no surge del movimiento orbital de sus «partes» en torno a su centro).

Una partícula cuyo spin es un múltiplo *impar* de $\hbar/2$ (es decir, $\hbar/2, 3\hbar/2$, ó $5\hbar/2$, etc.) se llama *fermión*, y exhibe una curiosa rareza de la descripción mecano-cuántica: ¡una rotación completa de 360° transforma su vector de estado no en sí mismo sino en *menos* sí mismo! La mayoría de las partículas de la Naturaleza son realmente fermiones, y más adelante oiremos más sobre ellas y sus singulares maneras —tan vitales para nuestra experiencia. Las partículas restantes, para las que el spin es un múltiplo *par* de $\hbar/2$, es decir, un múltiplo entero

de $\hbar$ (a saber, $0, \hbar, 2\hbar, 3\hbar\dots$), se llaman bosones. Bajo una rotación de 360° el vector de estado de un bosón vuelve a sí mismo, y no a su negativo.

Consideremos una partícula de *spin* $1/2$, esto es, con un valor $\hbar/2$ para el spin. Para ser más concreto, me referiré a la partícula como un *electrón*, pero un protón o neutrón, o incluso un tipo adecuado de átomo, servirían igualmente. (Se admite que una «partícula» puede poseer partes individuales con tal de que pueda ser tratada mecano-cuánticamente como un todo simple, con un momento angular total bien definido.) Tomemos el electrón en reposo y consideremos sólo su estado de spin. El espacio de estados mecano-cuánticos (espacio de Hilbert) resulta ser ahora *bi*-dimensional, de modo que podemos tomar una base de sólo *dos* estados. Representaré estos estados como $|\uparrow\rangle$ y $|\downarrow\rangle$, para indicar que para $|\uparrow\rangle$ el spin gira a derechas alrededor de la dirección vertical hacia *arriba*, mientras que para $|\downarrow\rangle$, gira a derechas alrededor de la dirección hacia *abajo* (Fig. 6.24). Los estados $|\uparrow\rangle$ y $|\downarrow\rangle$ son mutuamente ortogonales y los suponemos normalizados ($|\uparrow|^2 = |\downarrow|^2 = 1$). Cualquier posible estado de spin del electrón es una superposición lineal, digamos $w|\uparrow\rangle + z|\downarrow\rangle$, de sólo los *dos* estados ortonormales $|\uparrow\rangle$ y $|\downarrow\rangle$, es decir, de *arriba* y *abajo*.



Fig. 6.24. Una base para los estados de spin de un electrón consta de sólo dos estados. Estos pueden tomarse como *spin arriba* y *spin abajo*.

Ahora bien, no hay nada especial en las direcciones «arriba» y «abajo». Exactamente igual podríamos haber decidido describir el spin girando en (es decir, a derechas alrededor de) cualquier otra dirección, pongamos por caso *derecha* $|\rightarrow\rangle$ como opuesto a *izquierda* $|\leftarrow\rangle$. Entonces, (para una elección adecuada de la escala compleja para $|\uparrow\rangle$ y $|\downarrow\rangle$) encontramos *

$$|\rightarrow\rangle = |\uparrow\rangle + |\downarrow\rangle \quad \text{y} \quad |\leftarrow\rangle = |\uparrow\rangle - |\downarrow\rangle$$

* Como ya he hecho antes, prefiero no complicar las descripciones con factores como $1/\sqrt{2}$ que aparecerían si exigiésemos que $|\rightarrow\rangle$ y $|\leftarrow\rangle$ estén normalizados.

Esto nos da una nueva visión: cualquier estado de spin electrónico es una superposición lineal de los dos estados ortogonales $|\rightarrow\rangle$ y $|\leftarrow\rangle$, es decir, de *derecha* e *izquierda*. Podríamos escoger, en su lugar, alguna dirección completamente arbitraria, por ejemplo la dada por el vector de estado $|\nearrow\rangle$. Ésta es otra vez una cierta combinación lineal compleja de $|\uparrow\rangle$ y $|\downarrow\rangle$, digamos que

$$|\nearrow\rangle = w |\uparrow\rangle + z |\downarrow\rangle$$

y todo estado de spin será una superposición lineal de este estado y el estado ortogonal $|\nwarrow\rangle$, que apunta en dirección opuesta ⁹ a $|\nearrow\rangle$. (Nótese que el concepto de «ortogonal» en el espacio de Hilbert no corresponde necesariamente a «formando ángulos rectos» en el espacio ordinario. Los vectores ortogonales del espacio de Hilbert corresponden en este caso a direcciones diametralmente opuestas en el espacio, más que en direcciones que forman ángulos rectos.)

¿Cuál es la relación geométrica entre la dirección que determina $|\nearrow\rangle$ en el espacio y los números complejos w y z ? Puesto que el estado físico dado por $|\nearrow\rangle$ queda inalterado si multiplicamos $|\nearrow\rangle$ por un número complejo distinto de cero, será únicamente la *razón* de z a w la realmente significativa. Escribamos

$$q = z/w$$

para esta razón. Entonces q es simplemente algún número complejo, excepto que el valor « $q = \infty$ » también está permitido, para poder cubrir la situación $w = 0$, es decir, cuando la dirección del spin es la vertical hacia abajo. A menos que $q = \infty$, podemos representar q como un punto en el plano de Argand, igual que hicimos en el Capítulo 2. Imaginamos que este plano de Argand está situado horizontalmente en el espacio, con el eje real en la dirección hacia la «derecha» en la descripción anterior (esto es, en la dirección del estado de spin $|\rightarrow\rangle$). Imaginemos una esfera de radio unidad cuyo centro está en el origen de este plano de Argand, de modo que los puntos 1, i , -1 , $-i$ están todos en el ecuador de la esfera. Consideremos el punto en el polo Sur, que designaremos ∞ , y proyectemos entonces desde este punto de modo que el plano de Argand entero se aplica sobre la esfera. Así, cualquier punto q en el plano de Argand corresponde a un único punto q en la esfera, obtenido alineando los dos puntos con el polo Sur (Fig. 6.25). Esta correspondencia se llama proyección *estereográfica* y tiene muchas hermosas propiedades geométricas (v.g. conserva los ángulos y aplica círculos en círculos). La proyección nos da una caracterización de los puntos de la esfera mediante números

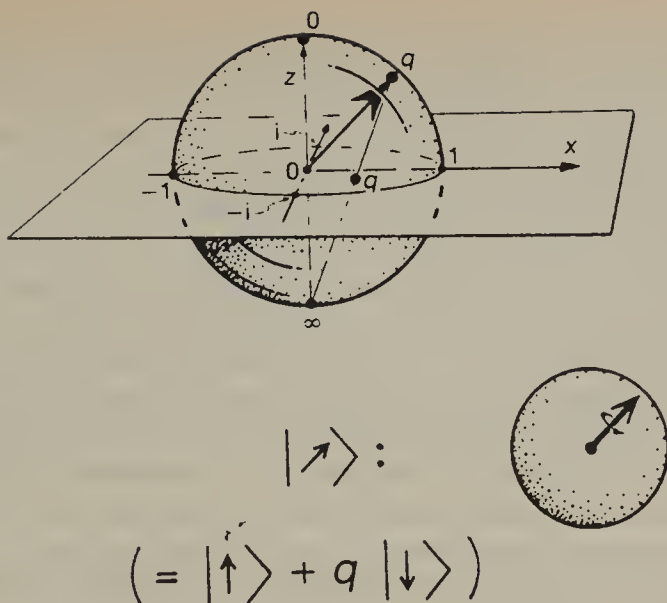


Fig. 6.25. La esfera de Riemann, representada aquí como el espacio de estados, físicamente distintos, de una partícula de spin $\frac{1}{2}$. La esfera se proyecta estereográficamente desde su polo Sur (∞) en el plano de Argand que pasa por su ecuador.

complejos junto con ∞ , es decir, mediante el conjunto de razones complejas q posibles. Una esfera caracterizada de esta forma particular se llama *esfera de Riemann*. La importancia de la esfera de Riemann, para los estados de spin del electrón, reside en que la dirección del spin definido por $|\nearrow\rangle = w|\uparrow\rangle + z|\downarrow\rangle$ viene dada por la dirección real desde el centro al punto $q = z/w$ marcado en la esfera de Riemann. Notemos que el polo Norte corresponde al estado $|\uparrow\rangle$, que está dado por $z = 0$, es decir por $q = 0$, y el polo Sur a $|\downarrow\rangle$, dado por $w = 0$, es decir por $q = \infty$. El punto más extremo hacia la derecha está caracterizado por $q = 1$, que proporciona el estado $|\rightarrow\rangle = |\uparrow\rangle + |\downarrow\rangle$, y el punto más extremo a la izquierda por $q = -1$, que proporciona el estado $|\leftarrow\rangle = |\uparrow\rangle - |\downarrow\rangle$. El punto más alejado hacia el fondo de la esfera está caracterizado por $q = i$, correspondiente al estado $|\uparrow\rangle + i|\downarrow\rangle$ en el que el spin apunta alejándose de nosotros, y el punto más próximo, $q = -i$, corresponde a $|\uparrow\rangle - i|\downarrow\rangle$, en el que el spin apunta directamente hacia nosotros. El punto más general, caracterizado por q , corresponde a $|\uparrow\rangle + q|\downarrow\rangle$.

¿Cómo enlaza todo esto con las medidas que pudiéramos realizar sobre el spin del electrón? ¹⁰ Seleccionemos alguna dirección en el espacio; llamemos a esta dirección α . Si medimos el spin del electrón en esta dirección, la respuesta **SÍ** dice que el electrón está (ahora) girando realmente según (esto es, a derechas alrededor de) α , mientras que **NO** dice que gira en dirección opuesta a α .

Supongamos que la respuesta es **SÍ**; entonces caracterizamos el estado resultante $|\alpha\rangle$. Si repetimos simplemente la medida, utilizando exactamente la misma dirección α que antes, entonces encontraremos que la respuesta debe ser de nuevo **SÍ** con un 100 % de probabilidad. Pero si para la segunda medida cambiamos la dirección a una nueva dirección β , entonces encontramos que hay una probabilidad más pequeña para la respuesta **SÍ**, saltando el estado ahora a $|\beta\rangle$, y hay alguna posibilidad de que la respuesta a la segunda medida pudiera ser **NO**, saltando el estado ahora a la dirección opuesta a β . ¿Cómo calculamos esta probabilidad? La respuesta está contenida en la receta dada al final de la sección previa. La probabilidad de **SÍ**, resulta ser para la segunda medida

$$1/2(1 + \cos\theta),$$

donde θ es el ángulo entre las direcciones ¹¹ de α y β . La probabilidad de **NO** para la segunda medida es, en consecuencia

$$1/2 (1 - \cos\theta)$$

Podemos ver a partir de aquí que si la segunda medida se realiza a ángulos rectos con la primera entonces la probabilidad de cualquiera de los resultados es del 50 % ($\cos 90^\circ = 0$): ¡el resultado de la segunda medida es completamente aleatorio! Si el ángulo entre las dos medidas es agudo, entonces la respuesta **SÍ** es más probable que **NO**. Si es obtuso, entonces **NO** es más probable que **SÍ**. En el caso extremo en que β es opuesta a α , las probabilidades se hacen de 0 % para **SÍ** y 100 % para **NO**; es decir, es seguro que el resultado de la segunda medida será el contrario del de la primera. (Ver Feynman *et al.*, 1965 para más información sobre el spin.)

La esfera de Riemann juega un papel fundamental (pero no siempre reconocido) en *cualquier* sistema cuántico de dos estados, describiendo el conjunto de estados cuánticos posibles (salvo proporcionalidad). Para una partícula de spin 1/2, su papel geométrico es particularmente evidente puesto que los puntos de la esfera corresponden a las posibles direcciones espaciales para el eje de giro. En muchas otras situaciones es más difícil de ver el papel de la esfera de Rie-

mann. Consideremos un fotón que acaba de atravesar un par de rendijas, o que ha sido reflejado por un espejo semi-reflectante. El estado del fotón es alguna combinación lineal tal como $|\psi_s\rangle + |\psi_i\rangle$, $|\psi_s\rangle - |\psi_i\rangle$, o $|\psi_s\rangle + i|\psi_i\rangle$, de dos estados $|\psi_s\rangle$ y $|\psi_i\rangle$ que describen dos localizaciones muy distintas. La esfera de Riemann describe todavía el conjunto de posibilidades físicamente distintas, pero ahora sólo de manera *abstracta*. El estado $|\psi_s\rangle$ está representado por el polo Norte («cima») y $|\psi_i\rangle$ por el polo Sur («fondo»). Entonces $|\psi_s\rangle + |\psi_i\rangle$, $|\psi_s\rangle - |\psi_i\rangle$ y $|\psi_s\rangle + i|\psi_i\rangle$ están representados por los diversos puntos en el ecuador, y en general $w|\psi_s\rangle + z|\psi_i\rangle$ está representado por el punto dado por $q = z/w$. En muchos casos, tal como éste, el «valor de la esfera de Riemann» de posibilidades está bastante oculto, sin ninguna relación clara con la geometría espacial.

Objetividad y medibilidad de los estados cuánticos

Pese al hecho de que normalmente sólo disponemos de probabilidades para el resultado de un experimento, parece haber algo *objetivo* en un estado mecano-cuántico. Se afirma a menudo que el vector de estado es simplemente una descripción conveniente de «nuestro conocimiento» acerca de un sistema físico —o, tal vez, que el vector de estado no describe en realidad un sistema simple sino que meramente proporciona información probabilística sobre un «conjunto» de un gran número de sistemas preparados de forma similar. Tales opiniones me resultan irrazonablemente tímidas en relación con lo que la mecánica cuántica tiene que decirnos sobre la *realidad* del mundo físico.

Parte de esta cautela, o duda, respecto a la «realidad física» de los vectores de estado parece surgir del hecho de que lo que es físicamente medible está limitado de forma estricta, según la teoría. Consideramos un estado de spin de un electrón, como los descritos antes. Supongamos que el estado de spin sea $|\alpha\rangle$, pero que nosotros no lo sabemos; es decir, nosotros no conocemos la *dirección* α del eje alrededor del que se supone que está girando el electrón. ¿Podemos determinar esta dirección mediante una medida? No, no podemos. Lo más que podemos hacer es extraer «un poco» de información —es decir, la respuesta a una simple pregunta sí/no. Podemos seleccionar alguna dirección β en el espacio y medir el spin del electrón en dicha dirección. Obtendremos la respuesta **SÍ** o la respuesta **NO**; pero, desde ese instante, habremos perdido la información sobre la dirección original del spin. Con una respuesta **SÍ** sabemos que el estado es

ahora proporcional a $|\beta\rangle$, y con una respuesta **NO** sabemos que el estado está *ahora* en la dirección opuesta a β . En ninguno de los dos casos nos dice esto cuál es la dirección α del estado *antes* de la medida, sino que simplemente da alguna información probabilística sobre α .

Por otra parte, parecería haber algo completamente *objetivo* sobre la propia dirección α en la que «resultaba estar girando» el electrón antes de que se hiciese la medida *. En efecto, *podríamos* haber decidido medir el spin del electrón en la dirección α —¡y el electrón tendría que estar preparado para dar la respuesta **SÍ**, con *certeza*, si nuestra conjetura hubiera estado en el camino correcto! De algún modo, la «información» de que el electrón va a dar realmente esta respuesta está almacenada en el estado de spin del electrón.

Creo que, al discutir la cuestión de la realidad física, debemos hacer una distinción entre lo que es «objetivo» y lo que es «medible» según la mecánica cuántica. En efecto, el vector de estado de un sistema es *no medible*, en el sentido de que no podemos verificar exactamente (salvo proporcionalidad), mediante experimentos realizados sobre el sistema, cuál es este estado; no obstante, el vector de estado *parece* ser (de nuevo salvo proporcionalidad) una propiedad totalmente *objetiva* del sistema, estando caracterizado completamente por los resultados que debe dar en los experimentos que *pudieran* realizarse. En el caso de una simple partícula de spin $1/2$, como un electrón, esta objetividad no es irrazonable ya que simplemente afirma que hay *alguna* dirección en la que el spin del electrón está exactamente definido, incluso aunque no podamos saber cuál es esta dirección. (Sin embargo, veremos más adelante que esta imagen «objetiva» es mucho más extraña con sistemas más complicados —incluso para un sistema que conste simplemente de un *par* de partículas de spin $1/2$.)

Pero, ¿es necesario que el spin del electrón tenga *en absoluto* un estado físicamente definido antes de ser medido? En muchos casos *no* lo tendrá, puesto que no puede considerarse como un sistema cuántico por sí mismo; en lugar de ello, el estado cuántico debe tomarse generalmente como la descripción de un electrón inextricablemente mezclado con un gran número de otras partículas. En circunstancias

* Esta objetividad es una característica de nuestra seria aceptación del formalismo estándar de la mecánica cuántica. En un punto de vista *no-estándar* el sistema podría «saber» realmente, antes de tiempo, el resultado que iba a dar en cualquier medida. Esto podría darnos una imagen *diferente*, aparentemente objetiva, de la realidad física.

particulares, no obstante, el electrón (al menos en lo que respecta a su spin) *puede* ser considerado independientemente. En tales circunstancias, como cuando su spin ha sido medido previamente en alguna dirección (quizá desconocida) y luego ha permanecido sin perturbar durante algún tiempo, el electrón *tiene* una dirección de spin perfecta y objetivamente definida según la teoría cuántica estándar.

Copia de un estado cuántico

La objetividad pero no-medibilidad de un estado de spin del electrón ilustra otro hecho importante: *¡es imposible copiar un estado cuántico y dejar intacto el estado original!* En efecto, supongamos que pudiéramos hacer tal copia de un estado de spin del electrón $|\alpha\rangle$. Si pudiéramos hacerlo una vez, entonces podríamos hacerlo otra vez, y otra y otra. El sistema resultante podría tener un enorme momento angular con una dirección muy bien definida. Esta dirección, a saber α , podría ser determinada mediante una medida macroscópica. Esto violaría la *no*-medibilidad fundamental del estado de spin $|\alpha\rangle$.

Sin embargo, *es* posible copiar un estado cuántico si estamos dispuestos a destruir el estado del original. Por ejemplo, podríamos tener un electrón en un estado de spin $|\alpha\rangle$ desconocido y un neutrón, por ejemplo, en otro estado de spin $|\gamma\rangle$. Es totalmente legítimo intercambiarlos, de modo que el estado de spin del neutrón es ahora $|\alpha\rangle$ y el del electrón es $|\gamma\rangle$. Lo que no podemos hacer es *duplicar* $|\alpha\rangle$ (¡a menos que ya *supiéramos* cuál es $|\alpha\rangle$ realmente!) (Cf. también Wootters y Zurek, 1982.)

Recordemos la «máquina teleportadora» discutida en el Capítulo 1 (p. 53). Ésta depende de que sea posible, en principio, reconstruir una copia completa del cuerpo y el cerebro de una persona en un planeta distante. Resulta intrigante especular que la «consciencia» de una persona pueda depender de algún aspecto de un estado cuántico. Si es así, la teoría cuántica nos prohibiría hacer una copia de esta «consciencia» sin destruir el estado del original —y, de esta forma, la «paradoja» de la teleportación podría resolverse. La posible relevancia de los efectos cuánticos para la función cerebral se considerará en los dos últimos capítulos.

El spin del fotón

Consideraremos a continuación el «spin» de un fotón y su relación con la esfera de Riemann. Los fotones *poseen* spin pero, debido a que siempre viajan a la velocidad de la luz, no podemos considerar el spin como si estuviera en un punto fijo; en lugar de ello, el spin del fotón está siempre en la dirección del movimiento. El spin del fotón se llama *polarización*, que es el fenómeno del que depende el comportamiento de las gafas de sol «polaroid». Tomen dos piezas de polaroide puestas una sobre otra y miren a través de ellas. Verán que en general dejan pasar una cierta cantidad de luz. Ahora giren una de las dos piezas manteniendo la otra fija. La cantidad de luz que dejan pasar variará. En una cierta orientación, para la que la luz transmitida presenta un máximo, el segundo polaroide no resta prácticamente nada de la luz entrante; mientras que para una orientación que forma un ángulo recto con la anterior, el segundo polaroide reduce la luz prácticamente a cero.

Lo que sucede se puede entender mejor en términos de la imagen ondulatoria de la luz. En este caso, necesitamos la descripción de Maxwell en términos de campos eléctricos y magnéticos oscilantes. En la Figura 6.26 se ilustra la luz *plano-polarizada*. El campo eléctrico oscila de un lado a otro en un plano —llamado *plano de polarización*— y el campo magnético oscila al unísono pero en un plano que forma un ángulo recto con el del campo eléctrico. Cada pieza de polaroide deja pasar la luz cuyo plano de polarización está alineado con la estructura del polaroide. Cuando el segundo polaroide tiene su estructura orientada de la misma forma que el primero, entonces toda la luz que atraviesa el primer polaroide atravesará también el segundo; pero cuando los dos polaroides tienen sus estructuras formando ángulos rectos, el segundo polaroide bloquea toda la luz que dejaba pasar el primero. Si los dos polaroides están orientados formando un ángulo φ entre ellos, entonces el segundo polaroide deja pasar una fracción

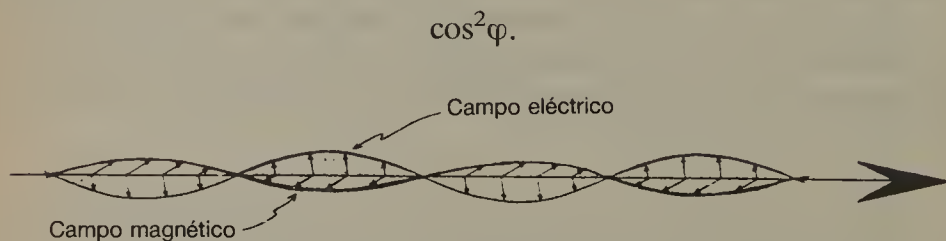


Fig. 6.26. Una onda electromagnética plano-polarizada.

En la imagen de partículas debemos pensar que *cada fotón individual* posee polarización. El primer polaroide actúa como un medidor de polarización que da la respuesta **SÍ** si el fotón está realmente polarizado en la dirección apropiada, y el fotón puede pasar. Si el fotón está polarizado en la dirección ortogonal, entonces la respuesta es **NO** y el fotón es absorbido. (¡Aquí, «ortogonal» en el sentido del espacio de Hilbert *corresponde* a «a un ángulo recto» en el espacio ordinario!) Suponiendo que el fotón atraviesa el primer polaroide, entonces el segundo polaroide plantea la pregunta correspondiente pero para alguna otra dirección. Puesto que el ángulo entre estas dos direcciones es φ , como antes, tenemos ahora una *probabilidad* $\cos^2\varphi$ de que el fotón atraviese el segundo polaroide una vez que atravesó el primero.

¿Dónde interviene la esfera de Riemann? Para obtener la colección completa de estados de polarización en forma compleja debemos considerar polarizaciones *circular* y *elíptica*. Para una onda clásica éstas quedan ilustradas en la Figura 6.27. Con polarización circular el campo eléctrico *rota*, en lugar de oscilar, y el campo magnético rota al unísono siempre formando un ángulo recto con el campo eléctrico. Para la polarización elíptica existe una combinación de movimientos oscilatorio y rotacional, y el vector que describe el campo eléctrico describe una *elipse* en el espacio. En la descripción cuántica cada fotón *individual* presenta estas diferentes alternativas de estar polarizado —los estados de *spin del fotón*.

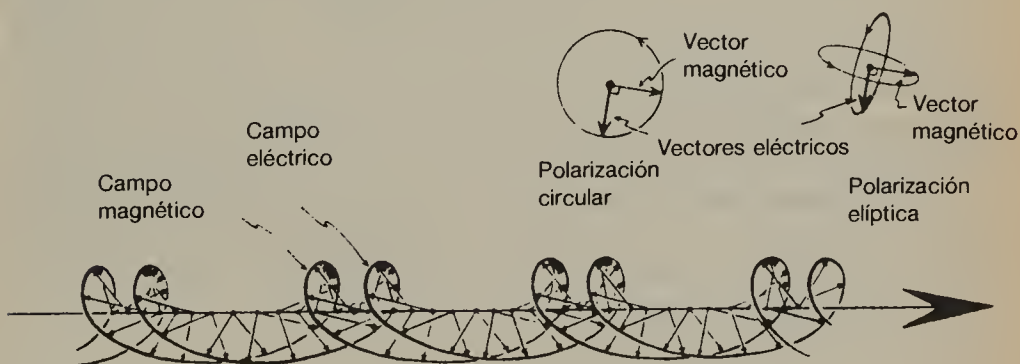


Fig. 6.27. Una onda electromagnética circularmente polarizada. (La polarización elíptica es intermedia entre las Figs. 6.26 y 6.27.)

Para ver cómo la colección de posibilidades es otra vez la esfera de Riemann imaginemos que el fotón está viajando verticalmente hacia arriba. El polo Norte representa ahora el estado D de spin a

derechas, lo que significa que el vector campo eléctrico rota alrededor de la vertical en sentido contrario a las agujas del reloj (tal como se vería desde arriba) a medida que pasa el fotón. El polo Sur representa el estado I de spin a *izquierdas*. (Podemos imaginar el fotón girando como las balas de rifle, o bien a derechas o bien a izquierdas.) El estado de spin general $|D\rangle + q|I\rangle$ es una combinación lineal compleja de los dos y corresponde a un punto, caracterizado por q , en la esfera de Riemann. Para encontrar la relación entre q y la elipse de polarización tomemos primero la *raíz cuadrada* de q para obtener otro número complejo p :

$$p = \sqrt{q}$$

A continuación marquemos p , en lugar de q , en la esfera de Riemann y consideremos el plano que pasa por el centro de la esfera y es perpendicular a la línea recta que une el centro con el punto q . Este plano intersecta a la esfera en un círculo y proyectamos este círculo verticalmente hacia abajo para obtener la elipse de polarización (Fig. 6.28) *. La esfera de Riemann de los q describe aún la totalidad de los espacios de polarización del fotón, pero la raíz cuadrada p de q proporciona su realización espacial.

Para calcular probabilidades podemos utilizar la misma fórmula $1/2(1+\cos\theta)$ que utilizamos para el electrón, siempre que la apliquemos a q y no a p . Consideremos polarización *plana*. Midamos primero la polarización del fotón en una dirección y luego en otra a un ángulo φ , con la primera, direcciones que corresponden a dos valores de p en el ecuador de la esfera que subtienden φ en el centro. Puesto que las p son las raíces cuadradas de las q , el ángulo φ subtendido en el centro por los puntos q es *doble* del que subtienden los puntos p : $\theta = 2\varphi$. Así, la probabilidad de **SÍ** para la segunda medida, una vez que se obtuvo **SÍ** para la primera (es decir, que el fotón atravesase el segundo polaroide una vez que ha atravesado el primero) es $1/2(1+\cos 2\varphi)$, que (por simple trigonometría) es lo mismo que el $\cos^2\varphi$ que resultaba antes.

* El número complejo $-p$ sería tan válido como p para la raíz cuadrada de q , y da la misma elipse de polarización. La raíz cuadrada tiene que ver con el hecho de que el fotón es una partícula sin masa y de *spin uno*, es decir, el *doble* de la unidad fundamental $\hbar/2$. Para un *gravitón* —el aún no detectado cuanto de gravitación— el spin será *dos*, es decir, *cuatro* veces la unidad fundamental, y tendríamos que tomar la *raíz cuarta* de q en la descripción anterior.

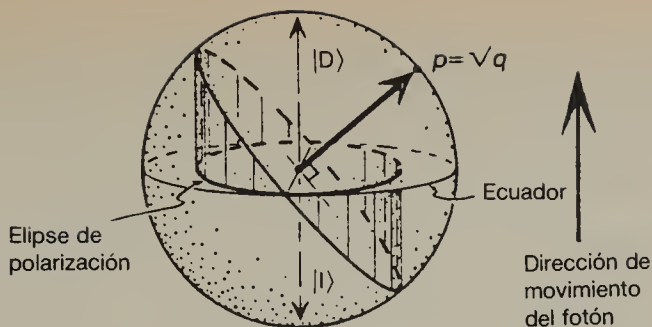


Fig. 6.28. La esfera de Riemann (pero ahora de $\sqrt{q}$) describe también estados de polarización de un fotón. (El vector que apunta hacia $\sqrt{q}$ se llama *vector de Stokes*.)

Objetos con gran spin

Para un sistema cuántico en el que el número de estados base es mayor que dos, el espacio de estados físicamente distinguibles es más complicado que la esfera de Riemann. Sin embargo, en el caso del spin la *propia* esfera de Riemann tiene siempre un papel geométrico directo que jugar. Consideremos una partícula o átomo en reposo *con masa* y de spin $n \times \hbar/2$. El spin define entonces un sistema cuántico de $(n + 1)$ estados. (Para una partícula *sin masa* que gira, es decir, una que viaje a la velocidad de la luz, tal como un fotón, el spin es siempre un sistema de *dos* estados como el descrito antes; pero para una partícula con masa el número de estados crece con el spin.) Si decidimos medir este spin en alguna dirección encontramos que existen $n + 1$ diferentes resultados posibles, dependiendo de qué cantidad del spin resulte estar orientada a lo largo de dicha dirección. En términos de la unidad fundamental $\hbar/2$, los posibles resultados para el valor del spin en dicha dirección son $n, n-2, n-4, \dots, 2-n$, ó $-n$. Así, para $n=2$ los valores son 2, 0 ó -2 ; para $n=3$, los valores son 3, 1, -1 , ó -3 ; etc. Los valores *negativos* corresponden al spin que apunta principalmente en la dirección *opuesta* a la que se está midiendo. En el caso de spin $1/2$, es decir $n = 1$, el valor 1 corresponde a **SÍ** y el valor -1 corresponde a **NO** en la descripción dada arriba.

Ahora bien, resulta, aunque no intentaré explicar las razones (Majorana, 1932, Penrose 1987a), que *todo estado de spin* (salvo proporcionalidad) para spin $\hbar n/2$ está caracterizado unívocamente por un conjunto (no ordenado) de n puntos en la esfera de Riemann —es

decir, por n direcciones (normalmente distintas) que salen del centro hacia afuera (ver Fig. 6.29). (Estas direcciones están caracterizadas por las medidas que podríamos realizar en el sistema: si medimos el spin en una de ellas es seguro que el resultado no estará completamente en la dirección opuesta, esto es, dará uno de los valores n , $n-2$, $n-4$, ..., $2-n$, pero *no* $-n$.) En el caso particular $n = 1$, como el electrón anterior, tenemos *un* punto en la esfera de Riemann, y éste es simplemente el punto caracterizado por q en las descripciones anteriores. Pero para valores mayores del spin la imagen es más elaborada y es como la acabo de describir —aunque, por alguna razón, no es muy familiar a los físicos.

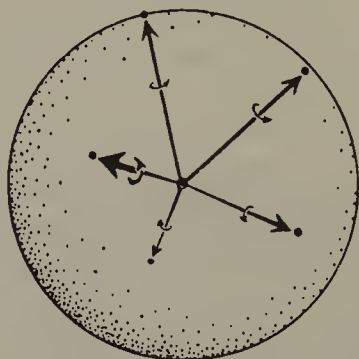


Fig. 6.29. Un estado general de spin superior, para una partícula con masa, puede describirse como una colección de estados de spin $\frac{1}{2}$ apuntando en direcciones arbitrarias.

Hay algo más bien curioso y enigmático acerca de esta descripción. Frecuentemente uno tiende a pensar que, en un sentido apropiado de límite, las descripciones cuánticas de los átomos (o las partículas elementales o las moléculas) coincidirán aproximadamente con las newtonianas clásicas cuando el sistema sea grande y complicado. Sin embargo, dicho así, *esto es sencillamente falso* pues, como hemos visto, los estados de spin de un objeto de gran momento angular corresponderán a un gran número de puntos salpicados por toda la esfera de Riemann *. Podemos considerar el spin del objeto como compuesto

* Más exactamente, el momento angular se describe mediante una combinación lineal compleja de tales colecciones con diferentes números de puntos, ya que, en el caso de un sistema complicado, puede haber superpuestos varios valores diferentes del spin total. ¡Esto sólo hace que la imagen total sea aún *menos* parecida a la de un momento angular clásico!

de un lote completo de spines $1/2$ apuntando en todas las diferentes direcciones que determinan estos puntos. Sólo unos pocos de estos estados combinados —a saber aquellos en los que la mayoría de los puntos se concentran en una pequeña región de la esfera (es decir, en los que la mayoría de los spines $1/2$ apuntan aproximadamente en la misma dirección)— corresponderán a los verdaderos estados de momento angular que encontramos normalmente con los objetos clásicos como las bolas de cricket. Hubiéramos esperado que si escogemos un estado de spin para el que la medida total del spin es un número muy grande (en términos de $\hbar/2$), aunque al «azar» en todo lo demás, entonces empezará a emerger algo similar al spin clásico. Pero no es así ni mucho menos como funcionan las cosas. ¡En general, los estados de spin cuánticos de un spin total grande no se parecen en nada a los clásicos!

¿Cómo debe hacerse, entonces, la correspondencia con el momento angular de la física clásica? Aunque la mayoría de los estados cuánticos de spin grande *no* se parecen a los clásicos, ellos son combinaciones lineales de estados (ortogonales) *cada uno* de los cuales *se parece* a uno clásico. De algún modo se realiza automáticamente una «medida» en el sistema y el estado «salta» (con cierta probabilidad) a uno u otro de estos estados similares a los clásicos. La situación es análoga con otras propiedades del sistema medibles clásicamente, y no sólo el momento angular. Es este aspecto de la mecánica cuántica el que debe entrar en juego cuandoquiera que un sistema «alcanza el nivel clásico». Diré más sobre esto más adelante, pero antes de que podamos discutir estos sistemas cuánticos «grandes» o «complicados» tendremos que tener alguna noción de la manera singular en que trata realmente la mecánica cuántica los sistemas que incluyen más de una partícula.

Sistemas de muchas partículas

Las descripciones mecano-cuánticas de estados de muchas partículas son, por desgracia, bastante complicadas. De hecho, pueden llegar a ser *extremadamente* complicadas. ¡Debemos pensar en términos de superposiciones de *todas* las diferentes localizaciones posibles de todas las partículas por separado! Esto da un vasto espacio de estados posibles —mucho más que en el caso de un *campo* en la teoría clásica. Hemos visto que incluso el estado cuántico para una *simple* partícula, a saber, una función de onda, tiene el mismo tipo de complicación

que tiene un campo clásico completo. Esta imagen (que requiere un número *infinito* de parámetros para especificarlo) es ya mucho más complicada que la imagen clásica de una partícula (que sólo necesita unos pocos números para especificar su estado —en realidad, seis, si no tiene grados de libertad internos, tales como spin; cf. Capítulo 5, p. 229). Esto puede parecer ya suficientemente grave, y podría pensarse que para describir un estado cuántico para dos partículas serían necesarios *dos* campos, uno para describir el estado de cada una de las partículas. ¡Nada de eso! Con dos o más partículas la descripción del estado es mucho más elaborada que esto, como veremos.

El estado cuántico de una *simple* partícula (sin spin) está definido por un número (amplitud) complejo para cada posible posición que la partícula pudiera ocupar. La partícula tiene una amplitud de estar en un punto *A* y una amplitud de estar en un punto *B* y una amplitud de estar en un punto *C*, etc. Ahora pensemos en *dos* partículas. La primera partícula podría estar en *A* y la segunda partícula en *B*, por ejemplo. Tendría que haber una amplitud para esta posibilidad. A la inversa, la primera partícula podría estar en *B* y la segunda en *A*, y esto también necesitaría una amplitud; o la primera podría estar en *B* y la segunda en *C*; o quizá ambas partículas podrían estar en *A*. Cada una de estas alternativas necesitará una amplitud. Así, la función de onda no es sólo un par de funciones de posición (es decir, un par de campos): es ¡una función de *dos* posiciones!

Para tener una idea de lo mucho más complicado que es especificar una función de dos posiciones que dos funciones de una posición imaginemos una situación en la que sólo hay disponible un conjunto finito de posiciones posibles. Supongamos que sólo hay diez posiciones permitidas, dadas por los estados (ortonormales)

$$|0\rangle, |1\rangle, |2\rangle, |3\rangle, |4\rangle, |5\rangle, |6\rangle, |7\rangle, |8\rangle, |9\rangle.$$

El estado $|\psi\rangle$ de una simple partícula será alguna combinación

$$|\psi\rangle = z_0|0\rangle + z_1|1\rangle + z_2|2\rangle + z_3|3\rangle + \dots + z_9|9\rangle$$

en la que las diversas componentes $z_0, z_1, z_2, z_3, \dots, z_9$ proporcionan las respectivas amplitudes de que la partícula esté en cada punto. Diez números complejos especifican el estado de la partícula. Para un estado de dos partículas necesitamos una amplitud para cada *par de posiciones*. Existen

$$10^2 = 100$$

diferentes pares (ordenados) de posiciones, de modo que necesitamos ¡cien números complejos! Si simplemente tuviéramos dos estados de una partícula (es decir, «dos funciones de posición» en lugar de «una función de dos posiciones» como la anterior) hubiéramos necesitado sólo *veinte* números complejos.

Podemos caracterizar estos cien números complejos como

$$z_{00}, z_{01}, z_{02}, \dots, z_{09}, z_{10}, z_{11}, z_{12}, \dots, z_{99}$$

y los correspondientes vectores base (ortonormales) como ¹²

$$|0\rangle|0\rangle, |0\rangle|1\rangle, |0\rangle|2\rangle, \dots, |0\rangle|9\rangle, |1\rangle|0\rangle, \dots, |9\rangle|9\rangle.$$

Entonces el estado general de dos-partículas $|\psi\rangle$, tendría la forma

$$|\psi\rangle = z_{00}|0\rangle|0\rangle + z_{01}|0\rangle|1\rangle + \dots + z_{99}|9\rangle|9\rangle.$$

Esta notación de «producto» de estados tiene el significado siguiente: si $|\alpha\rangle$ es un posible estado para la primera partícula (no necesariamente un estado de posición), y si $|\beta\rangle$ es un estado posible para la segunda partícula, entonces el estado que afirma que el estado de la primera partícula es $|\alpha\rangle$, y el de la segunda es $|\beta\rangle$ se escribirá

$$|\alpha\rangle|\beta\rangle$$

También pueden tomarse «productos» entre cualquier otro par de estados cuánticos, no necesariamente estados de partículas individuales. Así, interpretamos siempre el estado del producto $|\alpha\rangle|\beta\rangle$ (no necesariamente estados de partículas individuales) como describiendo la conjunción:

«el primer sistema tiene estado $|\alpha\rangle$ y el segundo sistema tiene estado $|\beta\rangle$ »

(Una interpretación similar sería válida para $|\alpha\rangle|\beta\rangle|\gamma\rangle$, etc.; ver más abajo.) Sin embargo, el estado *general* de dos-partículas no tiene en realidad esta forma de «producto». Por ejemplo, podría ser

$$|\alpha\rangle|\beta\rangle + |\varrho\rangle|\sigma\rangle$$

donde $|\varrho\rangle$ es otro estado posible para el primer sistema, y $|\sigma\rangle$ otro estado posible para el segundo. Este estado es una *superposición lineal*; a saber, la primera conjunción ($|\alpha\rangle$ y $|\beta\rangle$) *más* la segunda conjunción ($|\varrho\rangle$ y $|\sigma\rangle$), y no puede re-expresarse como un simple producto (es decir, como una conjunción de dos estados). Como un ejemplo más, el estado $|\alpha\rangle|\beta\rangle - i|\varrho\rangle|\sigma\rangle$ describiría una superposición lineal diferente. Nótese que la mecánica cuántica exige mantener una

clara distinción entre los significados de las palabras «y» y «más». Existe una desafortunada tendencia en el habla moderna —como en los folletos de seguros— a usar erróneamente «más» en el sentido de «y». ¡Aquí debemos ser mucho más cuidadosos!

La situación con tres partículas es muy similar. Para especificar un estado general de tres-partículas, en el caso anterior en el que sólo había disponibles diez alternativas para la posición, necesitamos ahora *¡mil números complejos!* La base completa para los estados de tres-partículas será

$$|0\rangle|0\rangle|0\rangle, |0\rangle|0\rangle|1\rangle, |0\rangle|0\rangle|2\rangle, \dots |9\rangle|9\rangle|9\rangle$$

Los estados particulares de tres-partículas tienen la forma

$$|\alpha\rangle|\beta\rangle|\gamma\rangle$$

(donde $|\alpha\rangle$, $|\beta\rangle$, y $|\gamma\rangle$ no necesitan ser estados de posición), pero para el estado general de tres-partículas tendríamos que superponer muchos estados de este tipo simple de «producto». La pauta correspondiente para cuatro o más partículas deberá estar ahora clara.

Hasta ahora se ha hecho la discusión para el caso de partículas *distinguibles*, en el que consideramos que la «primera partícula», la «segunda partícula», y la «tercera partícula», etc., son todas de diferentes *tipos*. Sin embargo, la mecánica cuántica tiene la característica sorprendente de que para partículas *idénticas* las reglas son diferentes. De hecho, las reglas son tales que, en un sentido preciso, las partículas de un tipo específico tienen que ser *exactamente* idénticas y ya no sólo, por ejemplo, muy aproximadamente idénticas. Esto se aplica a todos los electrones y se aplica a todos los fotones. Pero, como debe ser, ¡todos los electrones son idénticos entre sí de un modo *diferente* de como son idénticos todos los fotones! La diferencia estriba en el hecho de que los electrones son fermiones mientras que los fotones son bosones. Estos dos tipos generales de partículas tienen que ser tratados de forma bastante diferente.

Antes de que confunda completamente al lector con tales insuficiencias verbales explicaré cómo deben ser caracterizados en realidad los estados fermiónicos y bosónicos. La regla es como sigue. Si $|\psi\rangle$ es un estado que incluye cierto número de fermiones de un tipo particular, entonces al intercambiar dos cualesquiera de los fermiones, $|\psi\rangle$ debe sufrir una transformación

$$|\psi\rangle \rightarrow -|\psi\rangle$$

Si $|\psi\rangle$ incluye un número de bosones de un tipo particular, entonces al intercambiar dos cualesquiera de los bosones, $|\psi\rangle$ debe sufrir una transformación

$$|\psi\rangle \rightarrow |\psi\rangle$$

Una consecuencia de esto es que *dos fermiones no pueden estar en el mismo estado*; en efecto, si lo estuvieran su intercambio no afectaría en absoluto al sistema total de modo que deberíamos tener: $-|\psi\rangle = |\psi\rangle$, es decir, $|\psi\rangle = 0$, lo que no está permitido para un estado cuántico. Esta propiedad se conoce como *principio de exclusión de Pauli* ¹³, y tiene consecuencias fundamentales para la estructura de la materia. Todos los constituyentes principales de la materia son efectivamente fermiones: electrones, protones y neutrones. ¡Sin el principio de exclusión la materia colapsaría sobre sí misma!

Examinemos de nuevo nuestras diez posiciones y supongamos ahora que tenemos un estado consistente en dos fermiones idénticos. El estado $|0\rangle|0\rangle$ está excluido por el principio de Pauli (si intercambiamos el primer factor con el segundo el estado vuelve otra vez a sí mismo en lugar de a su negativo). Además, $|0\rangle|1\rangle$ no puede existir en esta forma puesto que no se transforma en su negativo bajo el intercambio; pero esto se remedia fácilmente reemplazándolo por

$$|0\rangle|1\rangle - |1\rangle|0\rangle.$$

(Podría incluirse, si quisieramos, un factor global $1/\sqrt{2}$ con fines de normalización.) Este estado cambia correctamente de signo bajo un intercambio de la primera partícula con la segunda, pero ahora no tenemos $|0\rangle|1\rangle$ y $|1\rangle|0\rangle$ como estados independientes. ¡En lugar de aquellos *dos* estados ahora sólo se permite *un* estado! En total hay

$$1/2 (10 \times 9) = 45$$

estados de este tipo, uno por cada par no ordenado de diferentes estados a partir de $|0\rangle, |1\rangle, \dots, |9\rangle$. Por lo tanto, se necesitan 45 números complejos para especificar un estado de dos fermiones en nuestro sistema. Para tres fermiones necesitamos tres posiciones distintas y los estados base son de la forma

$$|0\rangle|1\rangle|2\rangle + |1\rangle|2\rangle|0\rangle + |2\rangle|0\rangle|1\rangle + |0\rangle|2\rangle|1\rangle + |2\rangle|1\rangle|0\rangle + |1\rangle|0\rangle|2\rangle$$

habiendo $(10 \times 9 \times 8)/6 = 120$ de estos estados en total; de modo que se necesitan 120 números complejos para especificar un estado de tres fermiones. La situación es similar para números más altos de fermiones.

Para un par de bosones idénticos, los estados base independientes son de dos tipos, a saber, estados como

$$|0\rangle|1\rangle + |1\rangle|0\rangle$$

y estados como

$$|0\rangle|0\rangle$$

(que ahora están permitidos), dando $10 \times 11/2 = 55$ en total. Por lo tanto, se necesitan 55 números complejos para nuestros estados de dos bosones. Para tres bosones hay estados base de tres tipos diferentes y se necesitan $(10 \times 11 \times 12)/6 = 220$ números complejos; y así sucesivamente.

Por supuesto que he estado considerando aquí una situación simplificada para mostrar las ideas principales. Una descripción más realista requeriría un continuo de estados de posición, pero las ideas esenciales son las mismas. Otra pequeña complicación es la presencia de *spin*. Para una partícula de spin $1/2$ (necesariamente un fermión) habrá dos estados posibles para cada posición. Caractericémoslos por « $\uparrow$ » (spin «arriba») y « $\downarrow$ » (spin «abajo»). Entonces, para una simple partícula tenemos, en nuestra situación simplificada, veinte estados básicos en lugar de diez:

$$|0\uparrow\rangle, |0\downarrow\rangle, |1\uparrow\rangle, |1\downarrow\rangle, |2\uparrow\rangle, |2\downarrow\rangle, \dots, |9\uparrow\rangle, |9\downarrow\rangle.$$

pero, aparte de esto, la discusión procede igual que antes (de modo que para dos de tales fermiones necesitamos $(20 \times 19)/2 = 190$ números; para tres necesitamos $(20 \times 19 \times 18)/6 = 1140$; etc.).

En el Capítulo 1 me referí al hecho de que, según la teoría moderna, si se intercambia una partícula del cuerpo de una persona con una partícula similar de uno de los ladrillos de su casa, entonces no habrá sucedido nada en absoluto. Si esta partícula fuera un bosón, entonces, como hemos visto, el estado $|\psi\rangle$ quedaría totalmente infectado. Si la partícula fuera un fermión, entonces el estado $|\psi\rangle$ quedaría reemplazado por $-|\psi\rangle$, que es físicamente idéntico a $|\psi\rangle$. (Podemos remediar este cambio de signo, si lo creemos necesario, sin más que tomar la precaución de dar un giro completo de 360° a una de las dos partículas al hacer el intercambio. ¡Recuérdese que los fermiones cambian de signo bajo una rotación semejante mientras que los estados bosónicos permanecen inalterados!) La teoría moderna (de alrededor de 1926) nos dice realmente algo profundo sobre la cuestión de la identidad física de fragmentos de material físico. Estrictamente ha-

blando, no podemos referirnos a «este particular electrón» o «ese fotón individual». Afirmar que «el primer electrón está aquí y el segundo allí» es afirmar que el estado tiene la forma $|0\rangle|1\rangle$ que, como hemos visto, ¡no está permitida para un estado fermiónico! Sin embargo sí que es permisible afirmar «hay un par de electrones, uno aquí y uno allí». Es legítimo referirse al conglomerado de todos los electrones o de todos los protones o de todos los fotones (incluso aunque esto ignora las *interacciones* entre diferentes tipos de partícula). Los electrones individuales proporcionan una aproximación a esta imagen global, como lo hacen los protones individuales o los fotones individuales. Esta aproximación funciona para la mayoría de los propósitos, pero existen varias circunstancias para las que no lo hace, siendo algunos contraejemplos dignos de resaltar la superconductividad, la superfluidez y el comportamiento del láser.

La imagen del mundo físico que nos ha presentado la mecánica cuántica no es en absoluto la imagen a la que nos había acostumbrado la física clásica. Pero ¡sujétense el sombrero: todavía hay cosas más extrañas en el mundo cuántico!

La «paradoja» de Einstein, Podolsky y Rosen

Como se mencionó al principio de este capítulo, algunas de las ideas de Albert Einstein fueron fundamentales para el desarrollo de la teoría cuántica. Recuérdese que fue él quien propuso por primera vez el concepto del «fotón» —el cuanto de campo electromagnético— ya en 1905, a partir del cual se desarrolló la idea de la dualidad onda-corpúsculo. (El concepto de «bosón» fue también en parte suyo, como lo fueron muchas otras ideas centrales en la teoría.) Pese a todo, Einstein nunca pudo aceptar que la teoría que más tarde se desarrolló a partir de estas ideas pudiera ser algo más que una descripción provisional del mundo físico. Su aversión frente al aspecto probabilístico de la teoría es bien conocida y queda recogida en su contestación a una de las cartas de Max Born en 1926 (citada en Pais 1982, p. 443):

La mecánica cuántica es algo muy serio. Pero una voz interior me dice que, de todos modos, no es ése el camino. La teoría dice mucho, pero en realidad no nos acerca demasiado al secreto del Viejo. En todo caso estoy convencido de que Él no juega a los dados.

Parece, sin embargo, que lo que más molestaba a Einstein, más incluso que este indeterminismo físico, era una aparente *falta de objetividad* en la manera en que parecía que debía describirse la teoría cuántica. En mi exposición de la teoría cuántica he procurado enfatizar que la descripción del mundo que proporciona la teoría es en realidad bastante objetiva, aunque con frecuencia muy extraña y contraintuitiva. Por el contrario, Bohr parece haber considerado el estado cuántico de un sistema (entre medidas) como sin genuina realidad física, actuando simplemente como un resumen de «nuestro conocimiento» acerca de dicho sistema. Pero ¿no podrían diferentes observadores tener un conocimiento diferente de un sistema, de tal modo que la función de onda parecería ser algo esencialmente *subjetivo* —o «toda en la mente de los físicos»? No deberíamos permitir que nuestra maravillosamente precisa imagen física del mundo, desarrollada durante muchos siglos, se evaporara por completo; para ello Bohr necesitaba considerar que el mundo en el *nivel clásico* tenía una realidad objetiva. Pero no habría «realidad» en los estados de nivel *cuántico* que parecen subyacer en todo.

Semejante imagen era inadmisible para Einstein, quien creía que debe haber un mundo físico objetivo, incluso en las escalas minúsculas de los fenómenos cuánticos. En sus numerosas discusiones con Bohr intentó (aunque sin éxito) demostrar que había contradicciones inherentes a la imagen cuántica de las cosas, y que debe haber una estructura aún más profunda debajo de la teoría cuántica, probablemente más afín a la imagen que nos había presentado la física clásica. Tal vez subyaciendo a los comportamientos probabilistas de la mecánica cuántica estuviera la acción estadística de ingredientes o «partes» más pequeñas del sistema de las que no tenemos conocimiento directo. Los seguidores de Einstein, en particular David Bohm, desarrollaron el punto de vista de las «variables ocultas», según el cual habría alguna realidad definida pero los parámetros que definen exactamente el sistema no nos serían directamente accesibles, y las probabilidades cuánticas surgen debido a que no se pueden conocer los valores de estos parámetros antes de la medida.

¿Puede una teoría de variables-ocultas ser consistente con todos los hechos observacionales de la física cuántica? Parece que la respuesta es sí, pero sólo si la teoría es esencialmente *no-local* en el sentido de que los parámetros ocultos ¡deben poder afectar instantáneamente a regiones del sistema arbitrariamente lejanas! *Esto* no le hubiera gustado a Einstein, debido particularmente a las dificultades que surgen para la relatividad especial. Las consideraré más adelante.

La teoría de variables-ocultas más afortunada es la que se conoce como modelo de de Broglie-Bohm (de Broglie 1956, Bohm 1952). No discutiré aquí estos modelos puesto que mi propósito en este capítulo es dar sólo una visión general de la teoría cuántica estándar, y no de las diversas propuestas rivales. Si queremos objetividad física, pero estamos preparados para prescindir del determinismo, la teoría estándar será suficiente. Consideramos simplemente que el vector de estado proporciona «realidad» —evolucionando normalmente según el procedimiento determinista continuo U , pero «saltando» de cuando en cuando de forma extraña según R , cuandoquiera que un efecto se amplifica hasta el nivel clásico. Sin embargo, persiste el problema de la no-localidad y las dificultades aparentes con la relatividad. Echemos una mirada a algunas de éstas.

Supongamos que tenemos un sistema físico que consta de dos subsistemas A y B . Por ejemplo, consideremos que A y B son dos partículas diferentes. Supongamos que las dos alternativas (ortonormales) para el estado A son $|\alpha\rangle$ y $|\varrho\rangle$, mientras que el estado de B podría ser $|\beta\rangle$ o $|\sigma\rangle$. Como hemos visto antes, el estado general combinado no será simplemente un producto («y») de un estado de A por un estado de B , sino una superposición («más») de tales productos. (Diremos entonces que A y B están *correlacionados*.) Consideremos que el estado del sistema es

$$|\alpha\rangle|\beta\rangle + |\varrho\rangle|\sigma\rangle$$

Realicemos ahora una medida sí/no sobre A que distinga $|\alpha\rangle$ (SÍ) de $|\varrho\rangle$ (NO). ¿Qué pasar con B ? Si la medida da SÍ, entonces el estado resultante debe ser

$$|\alpha\rangle|\beta\rangle$$

mientras que si da NO, entonces es

$$|\varrho\rangle|\sigma\rangle$$

Por consiguiente, nuestra medida de A provoca que el estado de B salte: a $|\beta\rangle$ en el supuesto de una respuesta SÍ; y a $|\sigma\rangle$, en el supuesto de una respuesta NO. La partícula B no necesita estar localizada en alguna parte cerca de A ; podría estar a años-luz de distancia. ¡Pero B salta simultáneamente con la medida de A !

Pero ¡esperen! —puede estar diciendo el lector. ¿Qué es este pretendido «salto»? ¿Por qué no son las cosas de esta otra forma? Imaginemos una caja de la que sabemos que contiene una bola blanca y

una bola negra. Supongamos que se sacan las bolas y, sin mirarlas, se llevan a dos rincones opuestos de la habitación. Entonces, si se examina una bola y resulta ser blanca (el análogo de « $|\alpha\rangle$ » arriba) —¡ya está!— la otra resulta ser negra (el análogo de « $|\beta\rangle$ »)! Si, por el contrario, la primera resulta ser negra (« $|\alpha\rangle$ ») entonces, en un abrir y cerrar de ojos, el estado incierto de la segunda bola salta a «blanco, con certeza» (« $|\beta\rangle$ »). Nadie en su sano juicio, insistirá el lector, atribuirá el cambio repentino del estado «incierto» de la segunda bola hasta ser «negra, con certeza» o «blanca, con certeza» a alguna misteriosa «influencia» no-local que se propaga instantáneamente desde la primera bola en el instante en que esta bola es examinada.

Pero la Naturaleza es en realidad mucho más extraordinaria que esto. En lo que antecede podríamos imaginar que el *sistema* «sabía» ya, por ejemplo, que el estado de B era $|\beta\rangle$ y el de A era $|\alpha\rangle$ (o que el de B era $|\sigma\rangle$ y el de A era $|\rho\rangle$) antes de que se realizase la medida en A; y era simplemente el *experimentador* el que no lo sabía. Al encontrar que A está en el estado $|\alpha\rangle$, éste simplemente *infiere* que B está en $|\beta\rangle$. Ese sería un punto de vista clásico —semejante al de una teoría local de variables ocultas— y no habría lugar a ningún «salto» *físico*. (¡Todo estaría en la mente del experimentador!) Según esta idea, cada parte del sistema «sabe» por adelantado los resultados de cualquier experimento que se pudiera realizar sobre ella. Las probabilidades aparecen debido únicamente a la falta de conocimiento por parte del experimentador. ¡Curiosamente resulta que este punto de vista *no funcionará* como explicación de las enigmáticas probabilidades aparentemente no-locales que aparecen en la teoría cuántica!

Para ver esto consideraremos una situación como la anterior pero en la que la *elección de la medida* en el sistema A no se decide hasta que A y B estén muy separados. ¡El comportamiento de B parece entonces estar influido instantáneamente por esta misma elección! Este tipo «EPR» de «experimento mental» aparentemente paradójico se debe a Albert Einstein, Boris Pokolsky y Nathan Rosen (1935). Daré una variante propuesta por David Bohm (1951). El hecho de que ninguna descripción local 'realista' (v.g. de variables ocultas, o de «tipo clásico») pueda dar las probabilidades cuánticas correctas se sigue de un famoso teorema debido a Bell. (Ver Bell 1987, Rae 1986, Squires, 1986).).

Supongamos que se producen dos partículas de spin $\frac{1}{2}$ —que llamaré un *electrón* y un *positrón* (esto es, un *anti-electrón*)— en la desintegración de una simple partícula de spin cero en algún punto central, y que las dos se alejan en direcciones exactamente opuestas

(Fig. 6.30). Por la conservación del momento angular los spines del electrón y el positrón deben sumar cero puesto que éste era el momento angular de la partícula inicial. Esto tiene como consecuencia que, cuando medimos el spin del electrón en alguna dirección, cualquiera que sea la dirección que escojamos, ¡el positrón tiene un spin en la dirección *opuesta*! Las dos partículas podrán estar a kilómetros o incluso años-luz de distancia, pero esa misma *elección* de la medida en una partícula ¡parece haber fijado *instantáneamente* el eje de giro de la otra!

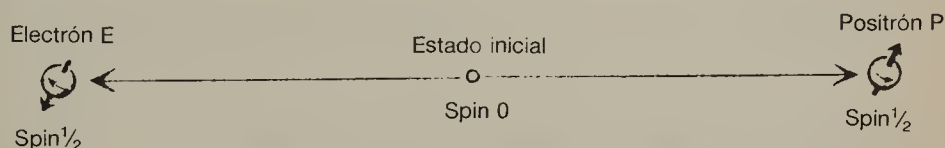


Fig. 6.30. Una partícula de spin cero se desintegra en dos partículas de spin $\frac{1}{2}$, un electrón E y un positrón P. Aparentemente, la medida del spin de una de las partículas de spin $\frac{1}{2}$ fija *instantáneamente* el estado de spin de la otra.

Veamos cómo el formalismo cuántico conduce a esta conclusión. Representamos el estado combinado de momento-angular-cero de las dos partículas mediante el vector de estado $|Q\rangle$, y encontraremos una relación como

$$|Q\rangle = |E \uparrow\rangle |P \downarrow\rangle - |E \downarrow\rangle |P \uparrow\rangle$$

en donde E designa al electrón y P al positrón. Aquí se han descrito las cosas en términos de las direcciones arriba/abajo del spin. Encontramos que el estado total es una superposición lineal del electrón con spin hacia arriba y el positrón con spin hacia abajo, y del electrón con spin abajo y el positrón con spin arriba. Ahora, si medimos el spin del electrón en la dirección arriba/abajo y encontramos que en realidad es arriba, entonces debemos saltar al estado $|E \uparrow\rangle |P \downarrow\rangle$, de modo que el estado de spin del positrón debe ser abajo. Si, por el contrario, encontramos que el spin del electrón es abajo, entonces el estado salta a $|E \downarrow\rangle |P \uparrow\rangle$ de modo que el spin del positrón es arriba.

Supongamos ahora que hemos elegido algún otro par de direcciones opuestas, por ejemplo derecha e izquierda, en donde

$$|E \rightarrow\rangle = |E \uparrow\rangle + |E \downarrow\rangle, \quad |P \rightarrow\rangle = |P \uparrow\rangle + |P \downarrow\rangle$$

y

$$|E \leftarrow\rangle = |E \uparrow\rangle - |E \downarrow\rangle, \quad |P \leftarrow\rangle = |P \uparrow\rangle - |P \downarrow\rangle;$$

entonces encontramos (¡comprueben el álgebra si quieren!):

$$\begin{aligned}
 & |E \rightarrow \rangle |P \leftarrow \rangle - |E \leftarrow \rangle |P \rightarrow \rangle \\
 &= (|E \uparrow \rangle + |E \downarrow \rangle) (|P \uparrow \rangle - |P \downarrow \rangle) - (|E \uparrow \rangle - |E \downarrow \rangle) (|P \uparrow \rangle + |P \downarrow \rangle) \\
 &= |E \uparrow \rangle |P \uparrow \rangle + |E \downarrow \rangle |P \uparrow \rangle - |E \uparrow \rangle |P \downarrow \rangle - |E \downarrow \rangle |P \downarrow \rangle \\
 &\quad - |E \uparrow \rangle |P \uparrow \rangle + |E \downarrow \rangle |P \uparrow \rangle - |E \uparrow \rangle |P \downarrow \rangle + |E \downarrow \rangle |P \downarrow \rangle \\
 &= -2(|E \uparrow \rangle |P \downarrow \rangle - |E \downarrow \rangle |P \uparrow \rangle) \\
 &= -2|Q \rangle
 \end{aligned}$$

que (aparte del factor -2 que no es importante) es el mismo estado que el de partida. ¡Por lo tanto, nuestro estado original puede considerarse también como una superposición lineal del electrón con spin hacia la derecha y el positrón hacia la izquierda, y el del electrón con spin hacia la izquierda y el positrón hacia la derecha! Esta expresión es útil si decidimos medir el spin del electrón en una dirección derecha/izquierda en lugar de arriba/abajo. Si encontramos que el spin está realmente hacia la derecha, entonces el estado salta a $|E \rightarrow \rangle |P \leftarrow \rangle$, de modo que el positrón tiene spin hacia la izquierda. Si, por el contrario, encontramos que el electrón tiene spin hacia la izquierda, entonces el estado salta a $|E \leftarrow \rangle |P \rightarrow \rangle$ de modo que el positrón tiene spin hacia la derecha. Si hubiéramos decidido medir el spin del electrón en cualquier otra dirección, la historia sería exactamente análoga: el estado de spin del positrón saltaría instantáneamente a estar en esa dirección o en la opuesta dependiendo del resultado de la medida sobre el electrón.

¿Por qué no podemos modelizar los spines de nuestros electrón y positrón de la misma forma que en el ejemplo anterior de una bola negra y una bola blanca sacadas de una caja? Hagámoslo de una forma totalmente general. En lugar de tener una bola negra y una bola blanca podríamos tener dos elementos mecánicos E y P que estuvieran inicialmente unidos y que luego separamos en direcciones opuestas. Supongamos que cada uno de estos E y P puede dar lugar a una respuesta **SÍ** o **NO** a una medida del spin en cualquier dirección dada. Esta respuesta podría estar completamente determinada, para cada elección de la dirección, por la constitución mecánica de E y P —o quizá el mecanismo produjera sólo respuestas probabilistas, con las probabilidades determinadas por la constitución mecánica— pero donde suponemos que, tras la separación, *cada uno de los E y P se comporta de forma totalmente independiente del otro.*

Disponemos de medidores de spin en cada lado, uno que mide el spin de E y otro que mide el spin de P . Supongamos que hay tres

posiciones para la dirección del spin en cada medidor, digamos A, B y C para el medidor-E, y A', B' y C' para el medidor-P. Las direcciones A', B', y C' serán paralelas respectivamente a las A, B y C, y consideraremos que A, B y C están en el mismo plano y formando ángulos iguales, es decir a 120° una de otra. (Ver Fig. 6.31). Imaginemos ahora que el experimento se repite muchas veces con varios valores diferentes de estas posiciones en cada lado. A veces el medidor-E registrará **SÍ** (es decir, el spin *está* en la dirección medida: A o B o C) y a veces registrará **NO** (spin en la dirección opuesta). Análogamente, el medidor-P registrará a veces **SI** y a veces **NO**. Tomemos nota de dos propiedades que deben tener las verdaderas probabilidades *cuánticas*:

(1) Si las posiciones en los dos lados son iguales (esto es, A y A', etc.) entonces los resultados de las dos medidas siempre *discrepan* (es decir, el medidor-E registra **SÍ** siempre que el medidor-P registre **NO**, y **NO** siempre que P de **SÍ**).

(2) Si los diales de las posiciones se giran y fijan *aleatoriamente*, completamente independientes uno de otro, entonces los dos medidores *tienen la misma probabilidad de coincidir que de discrepar*.

Podemos ver fácilmente que las propiedades (1) y (2) se siguen directamente de las reglas de probabilidad cuántica que hemos dado antes. Supongamos que el medidor-E actúa primero. El medidor-P encuentra entonces una partícula cuyo estado de spin es opuesto al medido por el medidor-E, de modo que la propiedad (1) se sigue inmediatamente. Para obtener la propiedad (2) notemos que, para medidas en direcciones que forman un ángulo de 120° , si el medidor-E da **SÍ**, entonces la dirección-P está a 60° respecto al estado de spin sobre el que actúa, y si da **NO**, entonces está a 120° respecto a este estado de spin. Por lo tanto existe una probabilidad $\frac{3}{4} = \frac{1}{2} (1 + \cos 60^\circ)$ de que las medidas coincidan y una probabilidad $\frac{1}{4} = \frac{1}{2} (1 + \cos 120^\circ)$

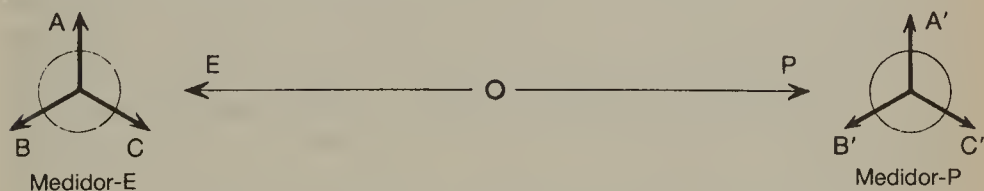


Fig. 6.31. Versión sencilla de David Mermin de la paradoja EPR y el teorema de Bell que demuestra que existe una contradicción entre una visión realista local de la Naturaleza y los resultados de la teoría cuántica. El medidor-E y el medidor-P, independientes uno de otro, tienen tres posiciones para la dirección en que pueden medir los spines de sus respectivas partículas.

de que discrepen. Por lo tanto, la probabilidad promediada para las tres posiciones de P, si E da **SÍ**, es $\frac{1}{3}(0 + \frac{3}{4} + \frac{3}{4}) = \frac{1}{2}$ para P dando **SÍ**, y $\frac{1}{3}(1 + \frac{1}{4} + \frac{1}{4}) = \frac{1}{2}$ para P dando **NO** —es decir, igualmente probable el acuerdo que la discrepancia—; y análogamente si da **NO**. Ésta es de hecho la propiedad (2). (Ver p. 337.)

¡Es un hecho notable que (1) y (2) sean *inconsistentes* con cualquier modelo realista local (esto es, con cualquier tipo de mecanismo del tipo propuesto)! Supongamos que tuvieramos tal modelo. La máquina E debe estar preparada para cada una de las medidas posibles A, B o C. Nótese que si estuviera preparada sólo para dar una respuesta *probabilista*, entonces no podría *asegurarse* que la máquina P estuviera en desacuerdo con ella, para A', B' y C' respectivamente, según afirma (1). En realidad, *ambas* máquinas deben tener decididamente preparadas por adelantado sus respuestas a cada una de las tres posibles medidas. Supongamos, por ejemplo, que estas respuestas van a ser **SÍ**, **SÍ**, **SÍ** respectivamente, para A, B, C; la partícula del lado derecho debe entonces estar preparada para dar **NO**, **NO**, **NO** para las tres posiciones correspondientes en el lado derecho. Si en lugar de esto, las respuestas preparadas en el lado izquierdo van a ser **SÍ**, **SÍ**, **NO**, entonces las respuestas del lado derecho van a ser **NO**, **NO**, **SÍ**. Todos los demás casos serían esencialmente similares a estos. Veamos ahora si esto puede ser compatible con (2). Las asignaciones **SÍ**, **SÍ**, **SÍ** / **NO**, **NO**, **NO** no son muy prometedoras ya que esto da 9 casos de discrepancia y 0 casos de acuerdo en todos los emparejamientos posibles A/A', A/B', A/C', B/A', etc. ¿Qué sucede con el caso **SÍ**, **SÍ**, **NO** / **NO**, **NO**, **SÍ** y similares? Estos dan 5 casos de discrepancia y 4 de acuerdo. (Para comprobarlo no hay más que contarlos: S/N, S/N, S/S, S/N, S/N, S/S, N/N, N/N, N/S, 5 de los cuales discrepan y 4 coinciden.) Esto está mucho más cerca de lo que se necesita para (2), pero no es *suficiente*, ¡ya que necesitamos tantas coincidencias como discrepancias! Cualquier otro par de asignaciones consistentes con (1) darían de nuevo 5 frente a 4 (excepto para **NO**, **NO**, **NO** / **SÍ**, **SÍ**, **SÍ** que es peor, pues da otra vez 9 a 0). ¡No existe conjunto de respuestas preparadas que pueda dar lugar a las probabilidades mecánico-cuánticas! ¡Los modelos realistas locales quedan descartados! ¹⁴.

Experimentos con fotones: ¿un problema para la relatividad?

Debemos preguntarnos si los experimentos reales han corroborado estas sorprendentes expectativas cuánticas. El experimento preciso re-

ción descrito es un experimento hipotético que no ha sido realizado realmente, pero se *han realizado* experimentos similares utilizando las polarizaciones de *pares* de fotones en lugar del spin de partículas con masa y de spin $\frac{1}{2}$. Aparte de esta diferencia, estos experimentos son, en sus aspectos esenciales, iguales al descrito antes —excepto que los ángulos que nos interesan (puesto que los fotones tienen spin uno en lugar de $\frac{1}{2}$) serán precisamente los ángulos mitad de los del experimento con partículas de spin $\frac{1}{2}$. La polarización de los pares de fotones se ha medido en varias combinaciones diferentes de direcciones y ¡los resultados están en completo acuerdo con las predicciones de la teoría cuántica, e inconsistentes con cualquier modelo realista local!

Los resultados más precisos y convincentes de los obtenidos hasta la fecha son los de Alain Aspect (1986) y sus colegas en París ¹⁵. Los experimentos de Aspect tienen otra característica interesante. Las «decisiones» sobre la forma de medida de las polarizaciones de los fotones se tomaban solamente después de que los fotones estuvieran en vuelo. Por lo tanto, si pensamos que alguna «influencia» no local está viajando desde un detector de fotones al fotón del lado opuesto, indicando la dirección en la que intenta medir la dirección de polarización del fotón que se aproxima, entonces vemos que esta «influencia» debe viajar ¡más rápida que la luz! Cualquier tipo de descripción realista del mundo cuántico que sea consistente con los hechos debe ser aparentemente *no causal*, en el sentido de que los efectos deben poder viajar más rápidos que la luz.

Pero vimos en el último capítulo que, mientras sea válida la relatividad general, el envío de señales más rápidas que la luz conduce a absurdos (y entra en conflicto con nuestras nociones de «libre albedrío», etc. cf. p. 271). Esto es ciertamente verdadero, pero las «influencias» no-locales que aparecen en experimentos tipo-EPR no son de las que se pueden utilizar para enviar mensajes —como se puede ver por la misma razón de que, si lo fueran, conducirían a absurdos semejantes. (Una demostración detallada de que estas «influencias» no pueden utilizarse para enviar mensajes ha sido llevada a cabo por Ghirardi, Rimini y Weber, 1980.) No sirve de nada que se nos diga que un fotón está polarizado «o vertical u horizontalmente», (como opuesto, digamos, a «o a 60° o 150°») hasta que se nos informe de *cuál* de las dos alternativas es la verdadera. Es el primer elemento de «información» (es decir, las *direcciones* de polarización alternativas) el que llega más rápido que la luz («instantáneamente»), mientras que el conocimiento sobre *en cuál* de estas direcciones deberá estar polarizado efectivamente llega más lentamente, por vía de una señal or-

dinaria que comunica el *resultado* de la primera medida de polarización.

Aunque los experimentos de tipo EPR no entran en conflicto, en el sentido ordinario del envío de señales, con la *causalidad* de la relatividad, existe un conflicto indudable con el *espíritu* de la relatividad en nuestra imagen de la «relatividad física». Veamos como la visión *realista* del vector de estado se aplica al experimento de tipo EPR anterior (con fotones). A medida que se separan los dos fotones, el vector de estado describe la situación como un *par* de fotones que actúan como una sola unidad. Ninguno de los fotones por separado tiene un estado objetivo: el estado cuántico se aplica sólo a los dos en conjunto. Cuando se mide la polarización de uno de estos dos fotones, el vector de estado *salta* de modo que el fotón no medido *tiene* ahora una polarización definida. Cuando *dicha* polarización del fotón es medida posteriormente, los valores de la probabilidad se obtienen correctamente aplicando las reglas cuánticas usuales a su estado de polarización. Esta manera de considerar la situación proporciona las respuestas correctas; es, en efecto, el modo en que aplicamos ordinariamente la mecánica cuántica. Pero es una visión esencialmente no-relativista; las dos medidas de polarización están separadas por lo que se conoce como un *intervalo de tipo-espacio*, lo que significa que cada una está fuera del cono de luz de la otra, como los puntos *R* y *Q* en la Figura 5.21. La pregunta de cuál de estas dos medidas ocurrió *primero* no es físicamente significativa sino que depende del estado de movimiento del «observador» (ver Fig. 6.32). Si el «observador» se mueve suficientemente rápido hacia la derecha, entonces él considera que la medida en el lado derecho ha ocurrido primero; y si se mueve hacia la izquierda, ¡entonces es la medida en el lado izquierdo! ¡Pero si consideramos que el fotón del lado derecho ha sido medido primero obtenemos una imagen de la realidad física completamente diferente de la obtenida si consideramos que el fotón del lado izquierdo se ha medido primero! (Es una medida diferente la que causa el «salto» no-local.) ¡Hay un conflicto esencial entre nuestra imagen espacio-temporal de la realidad física —incluso la mecánico-cuántica correctamente no-local— y la relatividad especial! Éste es un serio enigma, que los «realistas cuánticos» no han podido resolver adecuadamente (cf. Aharonov y Albert, 1981). Tendré necesidad de volver al tema más adelante.

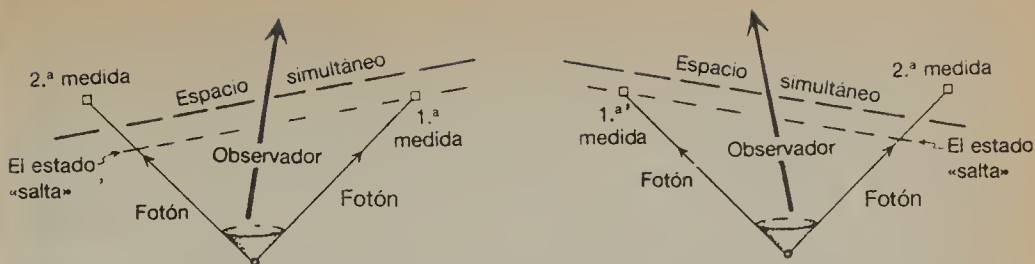


Fig. 6.32. Dos observadores diferentes forman imágenes de la «realidad» mutuamente inconsistentes en un experimento EPR en el que dos fotones son emitidos en direcciones opuestas a partir de un estado de spin-0. El observador que se mueve hacia la derecha juzga que la parte izquierda del estado salta *antes* de que sea medido, estando causado el salto por la medida en la derecha. ¡El observador que se mueve hacia la izquierda tiene la idea contraria!

La ecuación de Schrödinger y la ecuación de Dirac

Antes, en este mismo capítulo, me he referido a la ecuación de Schrödinger que es una ecuación determinista perfectamente definida, similar en muchos aspectos a las ecuaciones de la física clásica. Las reglas dicen que mientras no se hagan «medidas» (u «observaciones») sobre un sistema cuántico, la ecuación de Schrödinger será válida. El lector puede ser testigo de su forma real:

$$i\hbar \frac{\partial}{\partial t} |\psi\rangle = H|\psi\rangle.$$

Recordemos que $\hbar$ es una versión de Dirac de la constante de Planck ($h/2\pi$) (e $i = \sqrt{-1}$) y que el operador $\partial/\partial t$ (derivada parcial respecto al tiempo) actuando sobre $|\psi\rangle$ significa simplemente el *ritmo de cambio* de $|\psi\rangle$ con respecto al tiempo. La ecuación de Schrödinger establece que « $H|\psi\rangle$ » describe cómo evoluciona $|\psi\rangle$.

Pero ¿qué es « H »? Es la *función hamiltoniana* que consideramos en el capítulo anterior ¡pero con una diferencia fundamental! Recuerdese que el hamiltoniano clásico es la expresión para la *energía total* en términos de las diversas coordenadas de posición q_i y coordenadas de momento p_i , para todos los objetos físicos en el sistema. Para obtener el hamiltoniano *cuántico* tomamos la misma expresión, pero cada vez que aparezca el momento p_i , lo sustituimos por un múltiplo del *operador diferencial* «derivada parcial respecto a q_i ». Concretamente reemplazamos p_i por $-i\hbar\delta/\delta q_i$. Nuestro hamiltoniano cuántico H se convierte entonces en una (frecuentemente complicada) *ope-*

ración matemática que incluye derivadas y multiplicaciones, etc. —¡y ya no un simple número! ¡Esto parece una especie de abracadabra! Sin embargo, no es sólo un conjuro matemático: ¡es auténtica *magia* en acción! (Hay un poco de «arte» en la aplicación de este proceso de generar un hamiltoniano cuántico a partir de uno clásico, pero es curioso, en vista de su extravagante naturaleza, lo poco que parecen importar las ambigüedades inherentes al procedimiento.)

Algo importante que señalar acerca de la ecuación de Schrödinger (cualquiera que sea H) es que es *lineal*, es decir, si $|\psi\rangle$ y $|\varphi\rangle$ satisfacen ambos la ecuación, entonces también lo hace $|\psi\rangle + |\varphi\rangle$ —o, de hecho, cualquier superposición lineal compleja $w|\psi\rangle + z|\varphi\rangle$, donde w y z son números complejos fijos. Por lo tanto, una superposición lineal compleja se sigue manteniendo indefinidamente mediante la ecuación de Schrödinger. ¡Una superposición lineal (compleja) de dos estados alternativos posibles no puede des-superponerse simplemente por la acción de U ! Por esta razón es necesaria la acción de R , como un procedimiento *independiente* de U , para que finalmente sobreviva sólo *una* de las alternativas.

Al igual que el formalismo hamiltoniano para la física clásica, la ecuación de Schrödinger no es tanto una ecuación concreta como un marco para las ecuaciones mecano-cuánticas en general. Una vez que se ha obtenido el hamiltoniano cuántico apropiado, la evolución temporal del estado según la ecuación de Schrödinger tiene lugar como si $|\psi\rangle$ fuera un campo clásico sujeto a alguna ecuación clásica de campos como las de Maxwell. De hecho, si $|\psi\rangle$ describe el estado de un simple *fotón*, entonces resulta que la ecuación de Schrödinger *se convierte* realmente en las ecuaciones de Maxwell! La ecuación para un simple fotón es exactamente la misma que la ecuación * para un campo electromagnético completo. Este hecho es responsable de que el comportamiento de los *fonones individuales* que atisbamos antes sea parecido a las ondas del campo de Maxwell y la polarización. Como un ejemplo distinto, si $|\psi\rangle$ describe el estado de un simple *electrón*, entonces la ecuación de Schrödinger se transforma en la famosa ecuación de ondas de Dirac para el electrón —descubierta en 1928 después de que Dirac hubiera puesto mucha originalidad e intuición adicional.

De hecho, la ecuación de Dirac para el electrón debe colocarse,

* Sin embargo, hay una diferencia importante en el tipo de *solución* a la ecuación que es admisible. Los campos de Maxwell clásicos son necesariamente *reales* mientras que los estados del fotón son *complejos*. Existe también una condición llamada de «frecuencia positiva» que debe satisfacer el estado del fotón.

al lado de las ecuaciones de Maxwell y de Einstein, como una de las Grandes Ecuaciones de Campos de la física. Dar aquí una idea correcta de ella me obligaría a introducir ideas matemáticas que nos distraerían demasiado. Baste decir que en la ecuación de Dirac $|\psi\rangle$ tiene la curiosa propiedad «fermiónica» $|\psi\rangle \rightarrow -|\psi\rangle$ bajo la rotación de 360° que consideramos antes (p. 334). Las ecuaciones de Dirac y Maxwell juntas constituyen los ingredientes básicos de la electrodinámica cuántica, la más acertada de las teorías cuánticas de campos. Consideraremos esto en breve.

La teoría cuántica de campos

La materia conocida como «teoría cuántica de campos» ha aparecido como una unión de las ideas procedentes de la relatividad especial y la mecánica cuántica. Difiere de la mecánica cuántica estándar (es decir, no relativista) en que el número de partículas, de cualquier tipo, no necesita ser constante. Cada tipo de partícula tiene su *anti-partícula* (a veces, como en el caso de los fotones, la misma que la partícula original). Una partícula con masa y su antipartícula pueden aniquilarse para dar energía, y un par partícula-antipartícula puede crearse a partir de la energía. En realidad, el número de partículas no necesita estar definido; se permiten superposiciones lineales de estados con diferentes números de partículas. La teoría cuántica de campos suprema es la «electrodinámica cuántica» —básicamente la teoría de electrones y fotones. Esta teoría es notable por la precisión de sus predicciones (v.g. el valor preciso del momento magnético del electrón, citado en el último capítulo, p. 201). Sin embargo, es una teoría más bien desordenada —y no consistente globalmente— debido a que inicialmente da respuestas «infinitas» sin sentido. Éstas deben ser eliminadas mediante un proceso conocido como «renormalización». No todas las teorías cuánticas de campos son susceptibles de renormalización, y es difícil calcular con ellas incluso cuándo lo son.

Una aproximación popular a la teoría cuántica de campos es vía «integrales de camino», que supone la formación de superposiciones lineales cuánticas no sólo de diferentes estados de partículas (como con las funciones de onda ordinarias) sino de historias espacio-temporales completas de comportamiento físico (ver Feynman 1985, para una presentación popular). Sin embargo, esta aproximación tiene infinitos adicionales propios de ella, y sólo se puede entender vía la introducción de diversos «trucos matemáticos». A pesar de la potencia

indudable y la impresionante precisión de la teoría cuántica de campos (en aquellos pocos casos en los que la teoría se puede llevar hasta el final), nos quedamos con una sensación de que se necesita una comprensión más profunda antes de poder estar seguros de cualquier «imagen de la realidad física» a la que parezca conducir.

Debería dejar claro que la compatibilidad entre la teoría cuántica y la relatividad especial que proporciona la teoría cuántica de campos es sólo *parcial* —sólo afecta a U — y es sobre todo de naturaleza matemáticamente formal. La dificultad de una interpretación relativísticamente consistente de los «saltos cuánticos» que ocurren con R , la que nos dejaron los experimentos de tipo EPR, no es ni siquiera esbozada por la teoría cuántica de campos. Tampoco hay todavía ninguna teoría cuántica de campo gravitatorio consistente o creíble. Sugiriré, en el Capítulo 8, que estos temas pueden no estar completamente disociados.

El gato de Schrödinger

Volvamos finalmente a un tema que nos ha perseguido desde los inicios de nuestras descripciones. ¿Por qué no vemos superposiciones lineales de objetos a escala clásica, como bolas de cricket en dos lugares a la vez? ¿Qué es lo que hace que ciertas disposiciones de átomos constituyan «dispositivos de medida», de modo que el procedimiento R parezca reemplazar a U ? Ciertamente, cualquier pieza de un aparato de medida es ella misma parte del mundo físico, construida a partir de aquellos constituyentes mecano-cuánticos para cuyo examen del comportamiento ha sido diseñada. ¿Por qué no tratar el aparato de medida *junto* con el sistema físico examinado como un *sistema cuántico combinado*. Ninguna misteriosa medida «externa» está ahora involucrada. El sistema combinado debería evolucionar simplemente según U . ¿Pero lo hace? La acción de U sobre el sistema combinado es completamente determinista, ¡sin lugar para las incertidumbres probabilistas de tipo R implicadas en la «medida» u «observación» que el sistema combinado está realizando sobre sí mismo! Hay aquí una aparente contradicción, hecha especialmente gráfica en un famoso experimento mental introducido por Erwin Schrödinger (1935): *la paradoja del gato de Schrödinger*.

Imaginemos una habitación cerrada, construida de forma tan perfecta que ninguna influencia física puede atravesar sus paredes ni hacia dentro ni hacia afuera. Imaginemos que, dentro de la habitación,

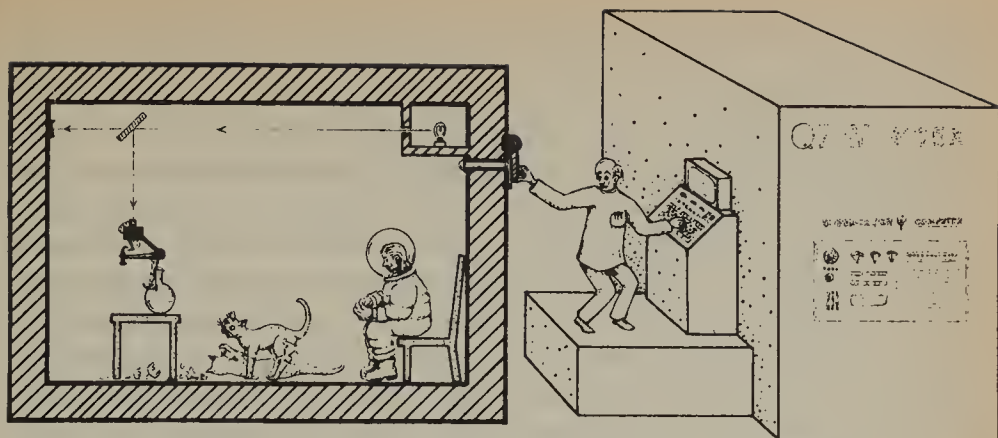


Fig. 6.33. El gato de Schrödinger —con aditamentos.

hay un gato y también un dispositivo que puede ser disparado mediante algún suceso cuántico. Si tiene lugar dicho suceso, entonces el dispositivo rompe una ampolla que contiene cianuro y el gato muere. Si el suceso no tiene lugar, el gato continúa vivo. En la versión original de Schrödinger el suceso cuántico era la desintegración de un átomo radiactivo. Permítaseme modificarla ligeramente y supongamos que el suceso cuántico es el disparo de una fotocélula por un fotón, fotón que ha sido emitido por alguna fuente luminosa en cierto estado predeterminado, y luego reflejado en un espejo semi-reflectante (ver Fig. 6.33). La reflexión en el espejo desdobra la función de onda del fotón en dos partes separadas, una de las cuales se refleja y la otra se transmite a través del espejo. La parte reflejada de la función de onda del fotón se focaliza sobre la fotocélula, de modo que si el fotón *es* registrado por la fotocélula ello significa que ha sido *reflejado*. En tal caso se libera el cianuro y el gato muere. Si la fotocélula *no* registra nada, el fotón fue transmitido a través del espejo semireflectante hasta la pared que hay detrás, y el gato se salva.

Desde el punto de vista (algo peligroso) de un observador en el *interior* de la habitación, ésta sería la descripción de lo que allí ocurriría. (¡Hubiéramos hecho mejor en proporcionar a este observador un traje protector adecuado!) *O bien* se considera que el fotón ha sido reflejado, porque se «observa» que la fotocélula ha registrado y el gato ha muerto, *o bien* se considera que el fotón ha sido transmitido, porque se «observa» que la fotocélula *no* ha registrado nada y el gato está vivo. O lo uno o lo otro tiene lugar *realmente*: **R** ha actuado y la probabilidad de cada alternativa es de un 50 por ciento (puesto que es un espejo *semi*-reflectante). Ahora adoptemos el punto

de vista de un físico en el *exterior* de la habitación. Podemos suponer que el vector de estado *inicial* de su contenido le es «conocido» antes de que la habitación sea cerrada. (No quiero decir que pudiera saberlo en la práctica, sino que no hay nada en la mecánica cuántica que diga que no pudiera saberlo *en principio*.) Según el observador exterior ninguna «medida» ha tenido lugar realmente, de modo que la evolución completa del vector de estado debería haber continuado según U. El fotón es emitido por su fuente en su estado predeterminado —ambos observadores coincidirán en esto— y su función de onda se desdobra en dos haces, con una amplitud de, pongamos por ejemplo, $1/\sqrt{2}$ de que el fotón esté en cada uno de ellos (de modo que el cuadrado del módulo diera realmente una probabilidad de $1/2$). Puesto que todo el contenido está siendo tratado como un solo sistema cuántico por el observador exterior, la superposición lineal de alternativas debe mantenerse hasta la escala del gato. Hay una amplitud $1/\sqrt{2}$ de que la fotocélula registre y una amplitud $1/\sqrt{2}$ de que no lo haga. *Ambas* alternativas deben estar presentes en el estado, con el mismo peso como parte de una superposición lineal cuántica. Según el observador exterior ¡el gato está en una superposición lineal de estar muerto y estar vivo!

¿Creemos realmente que sería así? El propio Schrödinger dejó claro que él no lo creía. Argumentaba, de hecho, que la regla U de la mecánica cuántica no sería aplicable a algo tan grande como un gato. Algo debe haber ido mal con la ecuación de Schrödinger en el camino. Por supuesto, Schrödinger tenía derecho a razonar de esta forma sobre su propia ecuación, ¡pero no es una prerrogativa que se nos conceda a los demás! Muchos físicos (probablemente la mayoría) mantendrían que, por el contrario, hay ahora tanta evidencia experimental a favor de U —y absolutamente ninguna en contra— que no tenemos ningún derecho a abandonar ese tipo de evolución, incluso en la escala de un gato. Si se acepta esto, entonces parece que estamos llevados a una visión muy *subjetiva* de la realidad física. Para el observador exterior el gato está realmente en una combinación lineal de estar vivo y muerto, y sólo cuando finalmente se abre la habitación colapsará el vector de estado del gato en uno u otro. Por el contrario, para un observador (adecuadamente protegido) dentro de la habitación, el vector de estado del gato habría colapsado mucho antes, y la combinación lineal del observador exterior

$$|\psi\rangle = \frac{1}{\sqrt{2}} \{ |\text{muerto}\rangle + |\text{vivo}\rangle \}$$

no tiene importancia. Parece que después de todo el vector de estado está ¡«todo en la mente»!

Pero ¿podemos adoptar realmente esta visión subjetiva del vector de estado? Supongamos que el observador exterior hiciera algo mucho más sofisticado que simplemente «mirar» dentro del contenedor. Supongamos que, a partir de su conocimiento del estado inicial del interior de la habitación, utiliza primero algún medio de cálculo importante disponible para él para *calcular*, mediante la ecuación de Schrödinger, cuál debe ser el estado en el interior del contenedor, obteniendo la respuesta (¡«correcta»!) $|\psi\rangle$ (donde $|\psi\rangle$ incluye la anterior superposición lineal de gato muerto y un gato vivo). Supongamos que entonces él realiza ese experimento *concreto* sobre los contenidos que distingue el propio estado $|\psi\rangle$ de cualquier cosa ortogonal a $|\psi\rangle$. (Como se ha descrito antes, según las reglas de la mecánica cuántica él puede *en principio* realizar dicho experimento, aunque fuera terriblemente difícil en la práctica.) Las dos resultados «sí, está en el estado $|\psi\rangle$ » y «no, es ortogonal a $|\psi\rangle$ » tendrían probabilidades respectivas del 100 por ciento y el 0 por ciento. En particular existe una probabilidad nula para el estado $|\chi\rangle = |\text{muerto}\rangle - |\text{vivo}\rangle$ que *es* ortogonal a $|\psi\rangle$. La imposibilidad de $|\chi\rangle$ como un resultado del experimento puede surgir sólo a causa de que *ambas* alternativas $|\text{muerto}\rangle$ y $|\text{vivo}\rangle$ *coexisten* e interfieren entre sí.

Lo mismo sería cierto si ajustásemos ligeramente las longitudes de los caminos del fotón (o la proporción reflectante del espejo) de modo que, en lugar del estado $|\text{muerto}\rangle + |\text{vivo}\rangle$, tuviéramos alguna otra combinación, pongamos por caso $|\text{muerto}\rangle - i|\text{vivo}\rangle$, etc. Todas estas diferentes combinaciones tienen consecuencias experimentales distintas —¡en principio! De modo que ya no se trata «simplemente» de cierto tipo de coexistencia entre muerte y vida que pudiera estar afectando a nuestro pobre gato. ¡Todas las diferentes combinaciones *complejas* están permitidas, y todas ellas son, en principio, distinguibles una de otra! Sin embargo, para el observador en el interior de la habitación todas estas combinaciones parecen irrelevantes. O el gato *está* vivo, o *está* muerto. ¿Cómo se puede entender este tipo de discrepancia? Señalaré en seguida algunos puntos de vista diferentes que se han manifestado sobre estas (y otras relacionadas) cuestiones —¡aunque sin duda no haré justicia a todas!

Diversas actitudes hacia la teoría cuántica existente

En primer lugar, existen dificultades obvias para realizar un experimento como el que distingue el estado $|\psi\rangle$ de cualquier otro ortogonal a $|\psi\rangle$. No hay duda de que un experimento semejante resulta *en la práctica* imposible para el observador externo. En particular, éste necesitaría conocer el vector de estado exacto de todo el contenido (incluyendo el observador interno) ¡antes de poder siquiera empezar a computar cuál sería realmente $|\psi\rangle$ en un tiempo posterior! Sin embargo, lo que nosotros exigimos es que este experimento sea imposible *en principio* —no simplemente en la práctica— puesto que de otro modo no tendríamos derecho a eliminar uno de los estados « $|\text{vivo}\rangle$ » o « $|\text{muerto}\rangle$ » de la realidad física. El problema es que la teoría cuántica, tal como está, no dice cómo trazar una línea clara entre medidas que son «posibles» y las que son «imposibles». Tal vez *debería* existir esta distinción tajante. Pero la teoría no lo permite. Para introducir una distinción semejante habría que *cambiar* la teoría cuántica.

En segundo lugar, existe el punto de vista no poco frecuente de que las dificultades desaparecerían si pudiéramos tener en cuenta de una manera adecuada el *entorno*. En realidad sería una imposibilidad práctica aislar *efectivamente* por completo el contenido del mundo externo. En cuanto el medio externo se mezcla con el estado en el interior de la habitación, el observador externo no puede considerar los contenidos como dados simplemente por un solo vector de estado. Incluso su *propio* estado se correlaciona con el entorno de una forma complicada. Además, habrá un enorme número de partículas inextricablemente entremezcladas y los efectos de las diferentes combinaciones lineales posibles se extenderán más y más por el universo sobre grandes números de grados de libertad. No hay modo *práctico* (por ejemplo, mediante la observación de efectos de interferencia adecuados) de distinguir estas superposiciones lineales complejas de las simples alternativas con pesos probabilistas. Esto ni siquiera tiene que ver con el aislamiento de los contenidos respecto al exterior. El propio gato incluye un gran número de partículas. En consecuencia, la combinación lineal compleja de un gato muerto y uno vivo debe ser tratada *como si* fuera simplemente una mezcla de probabilidades. Sin embargo, yo no encuentro esto nada satisfactorio. Como sucede con la concepción anterior podemos preguntar en qué etapa se juzga oficialmente que es «imposible» obtener efectos de interferencia —de modo que ahora pueda considerarse que los cuadrados de los módulos

de las amplitudes en la superposición compleja proporcionan una probabilidad ponderada de «muerto» y «vivo». Incluso si la «realidad» del mundo se transforma, en cierto sentido, *realmente* en un peso probabilístico en forma de número real, ¿cómo se resuelve esto en una sola alternativa o la otra? Yo no veo siquiera que la *realidad* pueda transformarse, sobre la sola base de la evolución U, de una *superposición* lineal compleja (o real) de dos alternativas en *una o la otra* de estas alternativas. ¡Parecemos llevados de nuevo a una visión subjetiva del mundo!

Hay quienes adoptan la postura de que los sistemas complicados no deberían describirse en realidad mediante «estados» sino mediante una generalización conocida como *matrices densidad* (von Neumann 1955). Éstas incluyen tanto probabilidades clásicas como amplitudes cuánticas. En efecto, muchos estados cuánticos diferentes se consideran en conjunto para representar la realidad. Las matrices densidad son útiles, pero no resuelven por ellas mismas los profundos problemas de la medida cuántica.

Podríamos tratar de adoptar la postura de que la evolución real es la U determinista, pero que las probabilidades surgen de las incertidumbres envueltas en saber cuál *es* realmente el estado cuántico del sistema combinado. Esto sería adoptar una visión muy «clásica» sobre el origen de las probabilidades —la de que todas surgen de las incertidumbres en el estado inicial. Podríamos imaginar que diferencias pequeñísimas en el estado inicial podrían dar lugar a diferencias enormes en la evolución, como el «caos» que puede ocurrir en los sistemas clásicos (v.g. la predicción del tiempo meteorológico; cf. Capítulo 5, p. 173). Sin embargo, tales efectos de «caos» sencillamente no pueden ocurrir con U por sí solo; puesto que es *lineal*: ¡las superposiciones lineales no deseadas persisten para siempre con U! Para resolver una superposición tal en una alternativa o la otra, se necesita algo *no-lineal*, de modo que U sólo no basta.

Para otro punto de vista, debemos tomar nota del hecho de que la única discrepancia completamente clara con la observación, en el experimento del gato de Schrödinger, parece surgir debido a que hay *observadores conscientes*, uno (¡o dos!) en el interior y otro en el exterior de la habitación. ¡Quizá las leyes de la superposición lineal cuántica compleja *no* se aplican a la consciencia! Un modelo matemático aproximado para este punto de vista fue propuesto por Eugene P. Wigner (1961). Él sugirió que la linealidad de la ecuación de Schrödinger podría fallar para entes conscientes (o simplemente «vivientes»), y debería ser reemplazada por algún procedimiento no-lineal de

acuerdo con el cual se resolvería una u otra alternativa. Pudiera parecer al lector que, puesto que estoy buscando algún tipo de papel para los fenómenos cuánticos en nuestro pensamiento consciente —como de verdad lo estoy haciendo— debería considerar este punto de vista como una atractiva posibilidad. Sin embargo, no me satisface en absoluto. Parece que lleva a una visión muy sesgada y perturbadora de la *realidad* del mundo. Los rincones del universo en donde reside la consciencia podrían ser más bien pocos y muy apartados. Desde esta perspectiva, *sólo* en aquellos rincones podrían resolverse las superposiciones lineales cuánticas complejas en alternativas reales. Puede ser que semejantes rincones tuvieran, para nosotros, la misma apariencia que el resto del universo, puesto que dondequiera que nosotros mismos *miráramos* (u observáramos de algún modo) haríamos, por el mismo acto de observación consciente, que se «resolviese en alternativas» *ya lo hubiese hecho antes o no*. Sea como fuere, este fuerte sesgo proporcionaría una imagen muy perturbadora de la *realidad* del mundo, y yo, por lo menos, lo aceptaría sólo si me viera obligado a ello.

Hay un punto de vista, relacionado en parte con el anterior, llamado el universo participatorio (sugerido por John A. Wheeler 1983), que lleva el papel de la consciencia a un (diferente) extremo. Notamos, por ejemplo, que la evolución de la vida consciente en nuestro planeta se debe a mutaciones apropiadas que han tenido lugar en distintos momentos. Éstas, presumiblemente, son sucesos cuánticos, de modo que sólo existirán en forma linealmente superpuesta hasta que finalmente conduzcan a la evolución de un ser consciente —¡cuya misma existencia depende de todas las mutaciones correctas que han tenido lugar «realmente»! Es nuestra propia presencia la que, en esta concepción, conjura a nuestro pasado a la existencia. La circularidad y paradoja que implica esta idea tiene atractivo para algunos, pero, por mi parte, la encuentro bastante preocupante— y, en realidad, escasamente creíble.

Otro punto de vista, también lógico a su manera pero que proporciona una imagen no menos extraña, es el de los *muchos universos*, publicado por primera vez por Hugh Everet III (1957). Según la interpretación de los muchos universos, **R** no tiene lugar nunca en absoluto. La evolución completa del vector de estado —que se considera de modo realista— está gobernada siempre por el procedimiento determinista **U**. Esto implica que el pobre gato de Schrödinger, junto con el observador protegido en el interior del contenedor, debe existir en alguna combinación lineal compleja con el gato en alguna

superposición de vida y muerte. Sin embargo, el estado de muerte está correlacionado con un estado de la consciencia del observador interno, y el estado vivo, con otro (y presumiblemente, en parte, con la consciencia del gato— y, en última instancia, también con el observador externo cuando el contenido le sea revelado). La consciencia de cada observador se «desdobla», de modo que ahora existe por duplicado, y cada uno de sus ejemplares tiene una experiencia diferente (es decir, uno que ve un gato vivo y otro que ve un gato muerto). De hecho, no sólo un observador sino todo el universo en el que habita se desdobla en dos (o más) en cada «medida» que hace del mundo. Este desdoblamiento ocurre una y otra vez —no simplemente a causa de «medidas» hechas por observadores, sino a causa de la amplificación macroscópica de estados cuánticos en general— de modo que estas «ramas» de universo proliferan incontroladamente. En realidad, todas las posibilidades alternativas coexistirán en una vasta superposición. Éste no es precisamente el más económico de todos los puntos de vista, pero mis propias objeciones a él no derivan de su falta de economía. En particular, no veo por qué un ser consciente necesita ser consciente de sólo «una» de las alternativas en una superposición lineal. ¿Qué consciencia es esa que exige que no podamos ser «conscientes» de esa seductora combinación lineal de un gato muerto y un gato vivo? Creo que sería necesaria una teoría de la consciencia antes de que la idea de los muchos universos pueda ser confrontada con lo que realmente observamos. No veo que relación existe entre el vector de estado «verdadero» (objetivo) del universo y el que se supone que observamos «realmente». Se ha alegado que la «ilusión» de R puede, en cierto sentido, deducirse efectivamente en esta imagen pero no creo que estas afirmaciones se sostengan. Cuando menos se necesitan más ingredientes para que el esquema funcione. Creo que la idea de los muchos universos introduce una multitud de problemas propios sin resolver realmente los *auténticos* enigmas de la medida cuántica. (Compárese con De Witt y Graham, 1973.)

¿Dónde nos deja todo esto?

Estos enigmas persisten, de una u otra forma, en *cualquier* interpretación de la mecánica cuántica tal como hoy existe. Repasemos brevemente lo que la teoría cuántica estándar nos ha dicho realmente acerca de cómo debemos describir el mundo, especialmente en rela-

ción con estos intrigantes temas; y luego preguntemos: ¿a dónde vamos desde aquí?

Recordemos, antes de nada, que las descripciones de la teoría cuántica parecen aplicarse acertadamente (¿útilmente?) sólo al llamado *nivel cuántico* —de moléculas, átomos o partículas subatómicas, pero también a dimensiones mayores siempre que las diferencias de energía entre posibilidades alternativas permanezcan muy pequeñas. En el nivel cuántico debemos tratar tales «alternativas» como cosas que pueden *coexistir* en alguna especie de superposición con pesos estadísticos complejos. Los números complejos que se utilizan como pesos se llaman *amplitudes de probabilidad*. Cada diferente totalidad de alternativas con pesos complejos define un diferente *estado cuántico*, y cualquier sistema cuántico debe describirse mediante uno de estos estados cuánticos. Con frecuencia, como sucedía muy claramente con el ejemplo del *spin*, no hay nada que decir sobre cuáles son las alternativas «reales» que componen un estado cuántico y cuáles son sólo «combinaciones» de alternativas. En cualquier caso, mientras el sistema *permanezca* en el nivel cuántico el estado cuántico evoluciona de forma completamente *determinista*. Esta evolución determinista es el proceso *U*, gobernado por la importante *ecuación de Schrödinger*.

Cuando los efectos de diferentes alternativas cuánticas se amplifican hasta el nivel clásico, de modo que las diferencias entre las alternativas son bastante grandes para que podamos percibir las directamente, entonces esas superposiciones con pesos complejos ya no parecen persistir más. En su lugar, deben formarse los cuadrados de los módulos de las amplitudes complejas (es decir, tomar los cuadrados de sus distancias al origen en el plano complejo), y estos números *reales* juegan ahora un nuevo papel como *probabilidades* reales para las alternativas en cuestión. Sólo *una* de las alternativas sobrevive en la realidad de la experiencia física siguiendo el proceso *R* (llamado reducción del vector de estado o colapso de la función de onda y que es completamente diferente de *U*). Es aquí, y sólo aquí, donde hace su entrada el no-determinismo de la teoría cuántica.

Puede defenderse con fuerza que el estado cuántico proporciona una imagen *objetiva*. Pero puede ser una imagen complicada y algo paradójica. Cuando varias partículas están involucradas, los estados cuánticos pueden (y normalmente lo «hacen») hacerse muy complicados. En tal caso, las partículas individuales no tienen «estados» por sí mismas sino que existen solamente en complicados «entramados» con otras partículas conocidos como *correlaciones*. Cuando una par-

tícula en una región es «observada», en el sentido de que desencadena algún efecto que se amplifica hasta el nivel clásico, entonces debe acudirse a **R** —pero en apariencia esto afecta *simultáneamente* a todas las demás partículas con la que esta partícula concreta está relacionada. Los experimentos del tipo Einstein-Podolsky-Rosen (EPR) (como el de Aspect, en el que una fuente cuántica emite un par de fotones en direcciones opuestas y luego se miden independientemente sus polarizaciones cuando están a muchos metros de distancia) dan una sustancia observacional evidente a este enigmático aunque esencial hecho de la física cuántica: ¡ésta es *no-local* (de modo que los fotones en el experimento de Aspect no pueden tratarse como entidades independientes)! Si se considera que **R** actúa de una manera objetiva (y eso parecería estar implicado en la objetividad del estado cuántico) entonces el espíritu de la mecánica cuántica es violado en consecuencia. ¡No parece existir *ninguna descripción espacio-temporal objetivamente real* del vector de estado (que se reduce) que sea consistente con los requisitos de la relatividad! Sin embargo, los efectos *observacionales* de la teoría cuántica no violan la relatividad.

La teoría cuántica guarda silencio sobre *cuándo y por qué* debería tener lugar (¿o aparentarlo?) **R**. Además, no explica adecuadamente, por sí sola, por qué el mundo del nivel clásico «parece» clásico. «La mayoría» de los estados cuánticos no se parecen en absoluto a los clásicos.

¿Dónde nos deja todo esto? Creo que debemos considerar seriamente la posibilidad de que la mecánica cuántica sea sencillamente *errónea* cuando se aplica a cuerpos macroscópicos —o, más bien que las leyes **U** y **R** sólo suministran excelentes aproximaciones a alguna teoría más completa aunque todavía desconocida. Es la *combinación* de estas dos leyes juntas, y no **U** por sí sola, la que ha proporcionado todo el maravilloso acuerdo con la observación de que goza la teoría actual. Si se extendiera la linealidad de **U** al mundo macroscópico tendríamos que aceptar la realidad física de combinaciones lineales complejas de diferentes posiciones (o diferentes spines, etc.) de bolas de cricket y similares. ¡El simple sentido común nos dice que el mundo no se comporta realmente de este modo! Las bolas de cricket se aproximan muy bien mediante las descripciones de la física *clásica*. Tienen posiciones razonablemente bien definidas, y no se las ve en dos lugares a la vez, como les permitirían estar las leyes de la mecánica cuántica. Si hay que reemplazar los procedimientos **U** y **R** por una ley más amplia, entonces, a diferencia de la ecuación de Schrödinger, esta nueva ley tendría que ser de carácter *no-lineal* (ya que

el propio **R** actúa de forma no lineal). Algunos presentan objeciones a esto, apuntando muy correctamente que buena parte de la elegancia matemática de la teoría cuántica estándar es resultado de su linealidad. Sin embargo, me sentiría sorprendido si la teoría cuántica no fuera a sufrir algún cambio fundamental en el futuro —hacia algo para lo que esta linealidad fuera sólo una aproximación. Ciertamente hay precedentes para este tipo de cambio. La poderosa y elegante teoría de Newton de la gravitación universal debía mucho al hecho de que las fuerzas de la teoría se suman de una manera *lineal*. Pero con la teoría de la relatividad general de Einstein se vio que esta linealidad era sólo una aproximación (si bien excelente) —¡y la elegancia de la teoría de Einstein supera incluso a la de Newton!

No me he andado con rodeos sobre el hecho de que creo que la resolución de los enigmas de la mecánica cuántica debe estar en el descubrimiento de una teoría mejorada. Aunque quizá no sea esta la opinión convencional tampoco es completamente no convencional. (Muchos de quienes dieron origen a la teoría cuántica eran también de este parecer. He citado las opiniones de Einstein, Schrödinger (1935), de Broglie (1956) y Dirac (1939) también consideraban provisional la teoría.) Pero incluso si se piensa que la teoría debe ser modificada de alguna manera, las restricciones sobre *cómo* hacerlo son enormes. Quizá algún tipo de punto de vista de «variables ocultas» resultaría aceptable finalmente. Pero la no-localidad que muestran los experimentos de tipo EPR desafían seriamente cualquier descripción «realista» del mundo que pueda ocurrir cómodamente en un espacio-tiempo ordinario —un espacio-tiempo del tipo particular que nos ha sido dado para coincidir con los principios de la relatividad— así que creo que se necesita un cambio mucho más radical. Además, nunca se ha encontrado ninguna discrepancia de ningún tipo entre la teoría cuántica y los experimentos— a menos, por supuesto, que consideremos como evidencia en contra la evidente ausencia de bolas de cricket superpuestas linealmente. ¡En mi propia opinión, la no existencia de bolas de cricket superpuestas linealmente *es* ciertamente evidencia en contra! Pero esto, en sí mismo, no es de gran ayuda. Sabemos que en el nivel sub-microscópico de las cosas las leyes cuánticas son válidas; pero en el nivel de las bolas de cricket es la física clásica la que vale. Sostendré que en algún lugar intermedio necesitamos comprender la nueva ley para ver cómo el mundo cuántico enlaza con el clásico. ¡Creo también que necesitamos esta nueva ley si queremos conocer alguna vez las mentes! Por todo esto pienso que debemos buscar nuevas claves.

En mis descripciones de la teoría cuántica en este capítulo he sido totalmente convencional, aunque con un énfasis quizá más geométrico y «realista» de lo que es usual. En el próximo capítulo trataremos de buscar algunas claves necesarias —claves que creo que nos pueden dar algunas ideas sobre una mecánica cuántica mejorada. Nuestro viaje se iniciará cerca de casa, pero nos veremos obligados a viajar muy lejos. ¡Resulta que necesitaremos explorar dominios del espacio muy diferentes, y retroceder incluso al propio comienzo del tiempo!

7. La cosmología y la flecha del tiempo

El flujo del tiempo

La sensación del paso del tiempo es central para nuestros sentimientos de consciencia. *Parece* que nos estemos moviendo siempre hacia adelante, desde un pasado definido hacia un futuro incierto. Sentimos que el pasado se ha ido y ya no hay nada que hacer con él. No se puede cambiar y, en cierto sentido, todavía está «ahí fuera». Nuestro conocimiento presente de él puede proceder de nuestros registros, de nuestra memoria y de nuestras deducciones de éstos, pero no tenemos tendencia a dudar de la *realidad* del pasado. El pasado era una cosa y sólo puede *ser* (ahora) una cosa. ¡Lo que sucedió, sucedió y ni nosotros, ni nadie más, podemos hacer nada por cambiarlo! Por el contrario, el futuro parece aún indeterminado. Podría resultar ser una cosa o podría ser otra. Quizá esta «elección» está completamente determinada por las leyes físicas, o quizá en parte por nuestras propias decisiones (o por Dios); pero *parece* que esta «elección» esté aún por hacerse. Da la impresión de que sólo hay simples *potencialidades* para cualquier cosa que la «realidad» del futuro pueda decidirse a ser. A medida que percibimos conscientemente que el tiempo pasa, la parte más inmediata de este vasto y aparentemente indeterminado futuro se realiza en acto y, de este modo, hace su entrada en el pasado fijo. A veces podemos tener la sensación de que *nosotros* hemos sido incluso personalmente «responsables» al influir de algún modo en esta elección del futuro potencial concreto que se realiza efectivamente y se hace permanente en la realidad del pasado. Más frecuentemente nos sentimos como espectadores inútiles —quizá agradecidos por este alivio de responsabilidad— de cómo, inexorable-

mente, la frontera del pasado determinado se mueve hacia el futuro incierto.

Pese a ello, la física, tal como la conocemos, nos cuenta una historia diferente. Todas las ecuaciones fructíferas de la física son simétricas respecto al tiempo. Se pueden utilizar tanto en una dirección del tiempo como en la otra. El futuro y el pasado parecen estar físicamente en pie de igualdad. Las leyes de Newton, las ecuaciones de Hamilton, las ecuaciones de Maxwell, la relatividad general de Einstein, la ecuación de Dirac, la ecuación de Schrödinger,... todas permanecen inalteradas si invertimos la dirección del tiempo (lo que equivale a reemplazar la coordenada t , que representa el tiempo, por $-t$.) Toda la mecánica clásica, junto con la parte «U» de la mecánica cuántica, es completamente reversible en el tiempo. Hay una cuestión abierta sobre si la parte «R» de la mecánica cuántica es realmente reversible o no. Esta cuestión será capital para los argumentos que presentaré en el próximo capítulo. Por el momento, dejemos de lado este tema y refirámonos a lo que puede considerarse el «saber convencional» sobre la materia, según el cual, a pesar de las apariencias iniciales, la operación de **R** debe ser aceptada también como tiempo-simétrica (cf. Aharonov, Bergmann y Lebowitz, 1964). Si aceptamos esto, parece que tendremos que buscar en otra parte si queremos encontrar el lugar donde, según nuestras leyes físicas, debe residir la diferencia entre pasado y futuro.

Antes de abordar este punto, consideremos otra discrepancia intrigante entre nuestra percepción del tiempo y lo que la teoría física moderna nos dice que debemos creer. Según la relatividad, no existe en absoluto algo como el «ahora». Lo más cercano que tenemos de tal concepto es el «espacio simultáneo» de un observador en el espacio-tiempo, como se representaba en la Figura 5.21, p. 259, ¡pero éste depende del *movimiento* del observador! El «ahora» de un observador no coincidiría con el de otro ¹. Con respecto a dos sucesos espaciotemporales A y B , un observador U podría considerar que B pertenece al pasado fijo y A al futuro incierto, mientras que para un segundo observador V , ¡sería A el que perteneciera al pasado fijo y B al futuro incierto! (Ver Fig. 7.1.) No tiene sentido afirmar que uno cualquiera de los sucesos A y B permanece incierto en tanto que el otro está definido.

Recordemos la discusión de la p. 260 y Figura 5.22. Dos personas se cruzan en la calle; según una de ellas, una flota espacial de Andrómeda ha iniciado ya su viaje, mientras que según la otra, todavía no se ha tomado la decisión de realizar o no dicho viaje. ¿Cómo

puede haber todavía alguna incertidumbre sobre el resultado de dicha decisión? Si para *una cualquiera* de las dos personas la decisión ya ha sido tomada, entonces es seguro que *no puede haber* ninguna incertidumbre. El lanzamiento de la flota espacial es algo inevitable. En realidad, ninguna de las dos personas puede *saber* todavía del lanzamiento de la flota espacial. Ellas sólo lo pueden saber más tarde, cuando las observaciones telescópicas desde la Tierra revelen que la flota está realmente en camino. Entonces ellas podrían recordar su encuentro y llegar a la conclusión de que en *ese* momento, según una de ellas, la decisión estaba en el futuro incierto, mientras que según la otra, estaba en el pasado cierto. ¿Había *entonces* alguna incertidumbre sobre dicho futuro? ¿O estaba «fijo» ya el futuro de *ambas* personas?

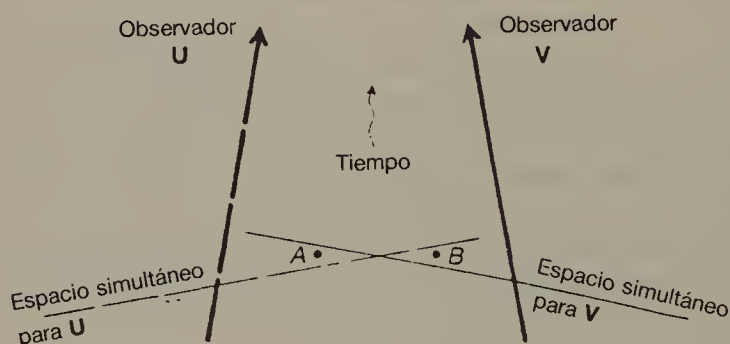


Fig. 7.1. ¿Puede el tiempo «fluir» realmente? Para un observador U, B puede estar en el pasado «fijo» mientras que A está todavía en el futuro «incierto». ¡El observador V mantendrá la opinión contraria!

¡Empieza a parecer que basta con que algo esté definido para que todo el espacio-tiempo deba estar definido! No puede haber futuro «incierto». La *totalidad* del espacio-tiempo debe estar fijada, sin ninguna perspectiva para la incertidumbre. De hecho, ésta parece haber sido la propia conclusión de Einstein (cf. Pais, 1982, p. 444). Además, no hay flujo del tiempo en absoluto. Sólo tenemos «espacio-tiempo» ¡y ninguna perspectiva para un futuro cuyo dominio está siendo invadido inexorablemente por un pasado determinado! (El lector puede estar preguntándose cuál es el papel de las «incertidumbres» de la mecánica cuántica en todo esto. Volveré más adelante, en este mismo capítulo, a las cuestiones que plantea la mecánica cuántica. Por el momento será mejor que pensemos en términos de una imagen puramente clásica.)

Tengo la impresión de que hay serias discrepancias entre lo que sentimos conscientemente, con relación al flujo del tiempo, y lo que nuestras teorías (maravillosamente precisas) afirman sobre la realidad del mundo físico. Seguramente estas discrepancias nos están diciendo algo profundo acerca de la física que presumiblemente debe subyacer a nuestras percepciones conscientes —suponiendo (como creo) que lo que subyace a estas percepciones sea inteligible mediante algún tipo apropiado de física. Al menos parece evidente que, cualquiera que sea la física que esté actuando, ella debe tener un ingrediente esencialmente tiempo-asimétrico, esto es, debe hacer una distinción entre el pasado y el futuro.

Si las ecuaciones de la física no parecen hacer distinción entre futuro y pasado —y si incluso la misma idea del «presente» no encaja perfectamente en la relatividad— entonces ¿dónde diablos tenemos que buscar para encontrar leyes físicas más en consonancia con lo que parecemos percibir del mundo? En realidad, las cosas no son tan discrepantes como parece que he estado dando a entender. Nuestro conocimiento de la física contiene realmente ingredientes importantes *además de* las solas ecuaciones de evolución temporal —y algunos de estos implican asimetrías temporales. El más importante de éstos es lo que se conoce como la *segunda ley de la termodinámica*. Intentemos sacar alguna idea de lo que esta ley significa.

El incremento inexorable de la entropía

Imaginemos un vaso de agua en equilibrio en el borde de una mesa. Si se le empuja ligeramente es probable que el vaso caiga al suelo —sin duda para hacerse añicos, con el agua salpicada por un área considerable, quizá absorbida en una alfombra o en las ranuras entre las baldosas del suelo. En esto, nuestro vaso de agua no ha hecho más que seguir fielmente las ecuaciones de la física. Las descripciones de Newton bastarán. Cada uno de los átomos del vaso y del agua está obedeciendo individualmente las leyes de Newton (Fig. 7.2). Ahora hagamos pasar estas imágenes al revés en la dirección inversa del tiempo. Por la reversibilidad temporal de estas leyes, el agua podría brotar de la alfombra y de las ranuras entre las baldosas, meterse en el vaso que está ocupado en construirse a sí mismo a partir de los numerosos trozos dispersos, para saltar luego todo el conjunto desde el suelo hasta la altura exacta de la mesa y colocarse en reposo equilibrado en el borde de la mesa. ¡Todo esto está de

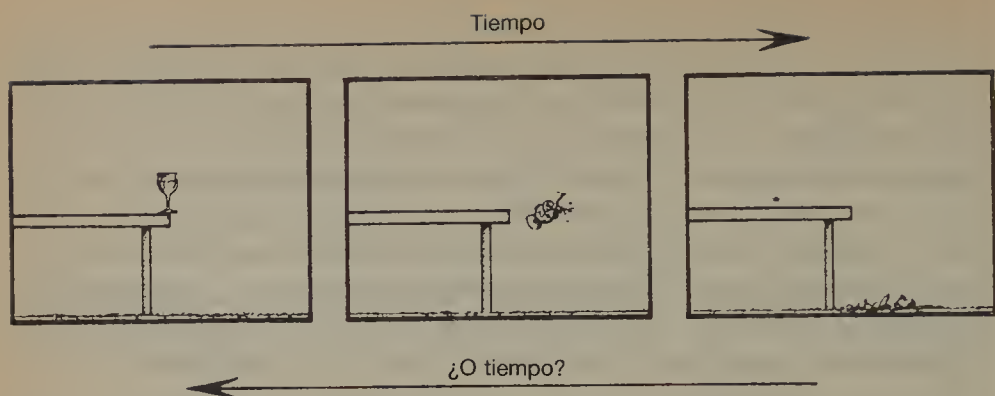


Fig. 7.2. Las leyes de la mecánica son reversibles respecto al tiempo; pese a ello nunca se tiene experiencia de una escena con un orden temporal que vaya desde la imagen derecha hacia la izquierda, mientras que la escena con orden temporal de izquierda a derecha sería algo muy común.

acuerdo con las leyes de Newton, lo mismo que la caída y rotura del vaso!

El lector puede estar preguntándose quizá de dónde sale la energía que eleva el vaso desde el suelo a la mesa. *Esto* no es problema. No puede haber problema con la energía puesto que en la situación en la que el vaso *cae* desde la mesa, la energía que adquiere de la caída debe *ir* a otra parte. De hecho, la energía del vaso que cae se transforma en *calor*. Los átomos de los fragmento del vaso, del agua, de la alfombra y de las baldosas se están moviendo de manera aleatoria justo un poco más rápido de lo que lo hacían en el momento en que el vaso chocó contra el suelo; es decir, los fragmentos del vaso, agua, alfombra y baldosas estarán apenas un poco *más calientes* de lo que estaban antes (ignorando posibles pérdidas de calor por evaporación —*aunque* también eso es reversible en principio). Por la *conservación de la energía*, esta energía calorífica es exactamente igual a la energía perdida por el vaso de agua al caer desde la mesa. Por lo tanto, esta pequeña cantidad de energía calorífica ¡será justo suficiente para elevar el vaso de nuevo hasta la mesa! Es importante darse cuenta de que debe ser incluida la energía calorífica cuando consideramos la conservación de la energía. La ley de la conservación de la energía, cuando se tiene en cuenta la energía calorífica, se llama *primera ley de la termodinámica*. Al ser una deducción de la mecánica newtoniana, la primera ley de la termodinámica es tiempo-simétrica. La primera ley *no* impone ninguna limitación al vaso y al agua que excluya

su recomposición, el llenado con agua y el salto milagroso a la mesa.

La razón de que no veamos que suceden estas cosas es que el movimiento «térmico» de los átomos en los fragmentos del vaso, agua, baldosas y alfombra será completamente desordenado, de modo que la mayoría de los átomos se estarán moviendo en las direcciones inadecuadas. Sería necesaria una coordinación absurdamente precisa de sus movimientos para recomponer el vaso, con todas las salpicaduras de agua recuperadas dentro, y alzarlo delicadamente hasta la mesa. ¡Es una certeza efectiva que semejante movimiento coordinado no estará presente! Semejante coordinación sólo podría ocurrir mediante el más inverosímil golpe de azar —¡de un tipo que llamaríamos «mágico» si se diera alguna vez!

Sin embargo, en la otra dirección del tiempo semejante movimiento coordinado es un lugar común. De ningún modo consideramos un golpe de suerte el que las partículas se muevan de forma coordinada, siempre que lo hagan *después* de que haya tenido lugar algún cambio a gran escala en el sistema físico (en este caso la ruptura y derramamiento del vaso de agua), y no *antes* de tal cambio. Los movimientos de las partículas deben efectivamente estar fuertemente coordinados después de tal suceso, pues estos movimientos son de tal naturaleza que si tuviéramos que invertir de una manera exacta el movimiento de cada átomo individual, el comportamiento resultante sería exactamente el necesario para recomponer, llenar y levantar el vaso hasta su exacta configuración de partida.

El movimiento fuertemente coordinado es aceptable si se considera como un *efecto* de un cambio a gran escala, y no como la *causa* de éste. Sin embargo, las palabras «causa» y «efecto» encierran de algún modo una petición de principio sobre el problema de la asimetría temporal. En nuestra forma de hablar normal acostumbramos a aplicar estos términos en el sentido de que la causa debe preceder al efecto. Pero, si estamos tratando de comprender la diferencia física entre pasado y futuro, tenemos que tener mucho cuidado de no introducir inconscientemente en la discusión nuestras sensaciones cotidianas sobre pasado y futuro. Debo advertir al lector que es extremadamente difícil evitarlo, pero es imperativo que tratemos de hacerlo. Debemos tratar de utilizar las palabras de tal modo que no prejuzguen el resultado de la diferencia física entre pasado y futuro. En consecuencia, si las circunstancias lo juzgan apropiado ¡tendríamos que permitirnos tomar las causas de las cosas como estando en el futuro y los efectos como estando en el pasado! Las ecuaciones deterministas de la física clásica (o, en su caso, la operación de U en la

física cuántica) no tienen preferencia por evolucionar en la dirección del futuro. Pueden utilizarse igualmente para evolucionar hacia el pasado. El futuro determina el pasado exactamente de la misma forma que el pasado determina el futuro. Podemos especificar un estado de un sistema de algún modo arbitrario en el futuro y luego utilizar este estado para calcular cómo hubiera debido ser en el pasado. Si se nos permite ver el pasado como «causa» y el futuro como «efecto», cuando seguimos las ecuaciones del sistema en la dirección normal del futuro, entonces cuando aplicamos el procedimiento igualmente válido de seguir la ecuaciones en la dirección del pasado debemos considerar aparentemente el futuro como «causa» y el pasado como «efecto».

Sin embargo, hay algo más implícito en nuestro uso de los términos «causa» y «efecto» que no es cuestión realmente de cuál de los sucesos referidos esté en el pasado y cuál en el futuro. Imaginemos un universo hipotético en el que se aplican las mismas ecuaciones clásicas con simetría temporal que en nuestro propio universo, pero en el que el comportamiento de tipo familiar (v.g. la rotura y derramamiento de los vasos de agua) coexiste con acontecimientos como los inversos de éstos en el tiempo. Supongamos que, a la par con nuestras experiencias más familiares, los vasos de agua a veces se *recomponen* a partir de los pedazos rotos, se llenan misteriosamente a partir de salpicaduras de agua y luego trepan a las mesas; supongamos también que, en ocasiones, los huevos revueltos se separan y des-fríen mágicamente, para volver finalmente a sus cáscaras rotas que se recomponen perfectamente y se cierran; que los terrones de azúcar pueden formarse por sí solos a partir del azúcar disuelto en un café azucarado y luego saltan espontáneamente desde la taza hasta la mano de alguien. Si viviéramos en un mundo en el que tales acontecimientos fueran un lugar común, seguramente atribuiríamos las «causas» de tales sucesos no a coincidencias azarosas fantásticamente improbables respecto al comportamiento correlacionado de los átomos individuales, sino a algún «efecto teleológico» por el que los objetos auto-formantes luchan a veces por conseguir alguna configuración macroscópica deseada. «¡Mira!», diríamos, «está ocurriendo otra vez. ¡Este revoltijo va a recomponerse en otro vaso de agua!» Sin duda aceptaríamos la idea de que los átomos se dirigían a sí mismos de forma tan precisa *debido a que* ésta era la forma de producir el vaso de agua sobre la mesa. El vaso sobre la mesa sería la «causa», y la aparentemente aleatoria colección de átomos en el suelo sería el «efecto» —pese al hecho de que el «efecto» ocurre ahora en un tiempo anterior a la

«causa». Análogamente, el movimiento minuciosamente organizado de los átomos en el huevo revuelto no es la «causa» del levantamiento hasta la cáscara recompuesta, sino el «efecto» de este acontecer futuro; y el terrón de azúcar no se reúne por sí mismo y se sale de la taza «a causa de que» los átomos se muevan con tan extraordinaria precisión, sino debido al hecho de que ¡alguien —aunque en el futuro— sostendrá más tarde ese terrón de azúcar en su mano!

Por supuesto no vemos que tales cosas sucedan en nuestro mundo —o, mejor dicho, lo que no vemos es la *coexistencia* de tales cosas con las de nuestro tipo normal. Si *todo* lo que viésemos fueran acontecimientos del tipo anómalo recién descrito, entonces no tendríamos problema. Podríamos intercambiar simplemente los términos «pasado» y «futuro», «antes» y «después», etc., en todas nuestras descripciones. El tiempo podría ser considerado avanzando en la dirección inversa de la especificada originalmente, y dicho mundo podría describirse como si fuera exactamente igual que el nuestro. Sin embargo, aquí estoy considerando una posibilidad diferente —pero igual de consistente con las ecuaciones tiempo-simétricas de la física— en la que la ruptura y la auto-recomposición de los vasos de agua pueden *coexistir*. En un mundo semejante no podemos recuperar nuestras descripciones familiares simplemente mediante una inversión de nuestras convenciones sobre la dirección de avance del tiempo. Por supuesto, nuestro mundo no parece ser así, pero ¿por qué no lo es? Para empezar a comprender este hecho he estado pidiéndoles que traten de imaginarse un mundo semejante y preguntarse cómo describiríamos los acontecimientos que tienen lugar en él. Les estoy pidiendo que acepten que, en semejante mundo, ciertamente describiríamos las grandes configuraciones macroscópicas —tales como vasos de agua enteros, huevos intactos, o un terrón de azúcar sostenido en una mano— como si fueran las «causas», y los detallados, y quizá estrechamente correlacionados, movimientos de los átomos individuales como «efectos», estén o no las «causas» en el futuro o en el pasado de los «efectos».

¿Por qué, en el mundo que nos ha tocado vivir, son las causas las que *preceden* a los efectos?; o, por poner las cosas de otra forma, ¿por qué los movimientos de partículas exactamente coordinados ocurren sólo *después* de algún cambio a gran escala en el sistema físico y no *antes* de él? Para dar una mejor descripción física de tales cosas necesitaré introducir el concepto de *entropía*. En términos generales, la entropía de un sistema es una medida de su *desorden* manifiesto. (Más adelante seré un poco más preciso.) Así, el cristal roto y el agua desparramada por el suelo están en un estado de mayor entropía que

el del vaso entero y lleno de agua en la mesa; el huevo revuelto tiene una entropía más alta que el huevo fresco intacto; el café azucarado tiene una entropía más alta que el terrón de azúcar sin disolver en un café amargo. El estado de baja entropía parece «particularmente ordenado», de algún modo manifiesto, y el estado de alta entropía, menos «particularmente ordenado».

Es importante darse cuenta de que cuando nos referimos a la «particularidad» de un estado de baja entropía nos estamos refiriendo en realidad a una particularidad *manifiesta*. De hecho, en un sentido más sutil, el estado de mayor entropía, en estas situaciones, *está* tan «particularmente ordenado» como el estado de menor entropía, debido a la muy precisa coordinación de movimientos de las partículas individuales. Por ejemplo, los movimientos aparentemente aleatorios de las moléculas de agua que se han escurrido entre las baldosas después de que el vaso se ha roto son realmente muy especiales: los movimientos son tan precisos que si todos ellos fueran *invertidos* exactamente se recobraría el estado original de baja entropía en el que el vaso está sobre la mesa entero y lleno de agua. (Esto debe ser así puesto que la inversión de todos estos movimientos correspondería simplemente a invertir la dirección del tiempo —según lo cual el vaso se recompondría por sí mismo y saltaría a la mesa.) Pero semejante movimiento coordinado de todas las moléculas de agua no es el tipo de «particularidad» que llamamos baja entropía. La entropía se refiere al desorden *manifiesto*. El orden que está presente en la coordinación exacta de movimientos de partículas no es orden manifiesto, así que no cuenta para disminuir la entropía de un sistema. Por lo tanto, el orden en las moléculas del agua desparramada no cuenta en este sentido y la entropía es alta. Sin embargo, el orden *manifiesto* en el vaso de agua *compuesto* da un bajo valor de la entropía. Este se refiere al hecho de que relativamente pocas configuraciones posibles diferentes de movimientos de partículas son compatibles con la configuración manifiesta de un vaso de agua entero y lleno; mientras que existen muchos más movimientos que son compatibles con la configuración manifiesta del agua ligeramente calentada que fluye entre las ranuras de las baldosas.

La *segunda ley de la termodinámica* afirma que *la entropía de un sistema aislado aumenta con el tiempo (o, en el caso de un sistema reversible, permanece constante)*. Está bien que no contemos los movimientos coordinados de partículas como baja entropía pues, si lo hiciéramos, la «entropía» de un sistema, según esa definición, siempre permanecería constante. El concepto de entropía debe referirse sólo

al desorden que es realmente manifiesto. Para un sistema aislado del resto del Universo, su entropía total crece, de modo que si se parte de algún estado con algún tipo de organización manifiesta esta organización será erosionada en el curso del tiempo, y estas especiales características manifiestas se convertirán en «inútiles» movimientos coordinados de partículas. Podría parecer, tal vez, que la segunda ley es una especie de último recurso, que nos dice que la organización se rompe continuamente por necesidad. ¡Veremos más adelante que esta conclusión pesimista no es completamente apropiada!

¿Qué es la entropía?

Pero ¿qué es exactamente la entropía de un sistema físico? Hemos visto que es algún tipo de medida del desorden manifiesto, pero podría parecer, dada mi utilización de términos tan imprecisos como «desorden» y «manifiesto», que el concepto de entropía no fuera realmente una cantidad científicamente muy clara. Existe también otro aspecto de la segunda ley que parece señalar un elemento de imprecisión en el concepto de entropía: es sólo en los llamados sistemas *irreversibles* en donde la entropía crece en lugar de permanecer constante. ¿Qué significa «irreversible»? Si tenemos en cuenta los movimientos detallados de todas las partículas, entonces ¡*todos* los sistemas son reversibles! *En la práctica* diríamos que el vaso que cae de la mesa y se rompe, o el revolver del huevo, o la disolución del azúcar en el café son todos irreversibles; mientras que el rebotar de un pequeño número de partículas unas en otras sería considerado reversible, como lo serían diversas situaciones cuidadosamente controladas en las que la energía no se transforma en calor. Básicamente, el término «irreversible» se refiere simplemente al hecho de que no ha sido posible seguir, ni controlar, todos los detalles relevantes de los movimientos de las partículas individuales del sistema. Estos movimientos incontrolados son designados como «calor». Así, la irreversibilidad parece ser simplemente una cuestión «práctica». No podemos *en la práctica* des-revolver un huevo, aunque es un procedimiento perfectamente admitido según las leyes de la mecánica. ¿Depende nuestro concepto de entropía de lo que es prácticamente factible y lo que no lo es?

Recordemos, del Capítulo 5, que se *puede* dar una definición matemática precisa del concepto físico de *energía*, tanto como de los de momento o momento angular, en términos de posiciones, veloci-

des, masas y fuerzas sobre las partículas. Pero ¿cómo podemos esperar que también se pueda hacer esto para el concepto de «desorden manifiesto» que es necesario para hacer matemáticamente preciso el concepto de entropía? Ciertamente, lo que es manifiesto para un observador puede no serlo para otro. ¿Dependería de la precisión con que cada observador sea capaz de medir el sistema bajo examen? Un observador con mejores instrumentos de medida podría obtener una información sobre los constituyentes microscópicos de un sistema mucho más detallada que la que pudiera obtener otro observador. Para un observador podría ser manifiesta una mayor parte del «orden oculto» que para otro —y, consecuentemente, concebiría una entropía más baja que el otro. Parece también que los juicios estéticos de los diversos observadores podrían estar involucrados en lo que ellos consideran que es «orden» más que «desorden». Podríamos pensar en algún artista que opina que la colección de fragmentos del vaso roto está ordenada de forma mucho más bella de lo que estaba el vaso horriblemente feo que estuvo una vez en el borde de la mesa. ¿Se habría reducido realmente la entropía a juicio de semejante observador con sensibilidad artística?

A la vista de estos problemas de subjetividad resulta notable que el concepto de entropía tenga siquiera alguna utilidad en descripciones científicamente precisas —¡como ciertamente la tiene! La razón de esta utilidad es que los cambios de orden a desorden en un sistema, en términos de posiciones y velocidades detalladas de las partículas, son absolutamente enormes y (en casi todas las circunstancias) desbordan completamente cualesquiera diferencias razonables de punto de vista sobre lo que es o no es «orden manifiesto» a escala macroscópica. En particular, el juicio del artista o del científico acerca de si es el vaso entero o el fragmentado el que tiene una configuración más ordenada no tiene apenas consecuencias respecto a su medida de la entropía. La mayor contribución a la entropía procede, con mucho, del movimiento aleatorio de las partículas que suponen el pequeñísimo incremento de temperatura y la dispersión del agua cuando vaso y agua golpean el suelo.

Para ser más precisos sobre el concepto de entropía volvamos a la idea de *espacio de fases* que fue introducida en el Capítulo 5. Recordemos que el espacio de fases de un sistema es un espacio, normalmente de un enorme número de dimensiones, cada uno de cuyos puntos representa un estado físico completo en sus más mínimos detalles. Un *simple* punto en el espacio de fases proporciona todas las coordenadas de posición y momento de todas las partículas individua-

les que constituyen el sistema físico en cuestión. Lo que necesitamos para el concepto de entropía es una manera de agrupar todos los estados que parezcan idénticos desde el punto de vista de sus propiedades *manifiestas* (esto es, macroscópicas). Por lo tanto, necesitamos dividir nuestro espacio de fases en un número de compartimentos (cf. Fig. 7.3), de manera que los diferentes puntos pertenecientes a un compartimento concreto representan sistemas físicos que, aunque diferentes en los detalles menudos de las configuraciones y movimientos de sus partículas, se consideran idénticos con respecto a sus características macroscópicamente observables. Desde el punto de vista de lo que es manifiesto, todos los puntos de un simple compartimento deben considerarse representando el *mismo* sistema físico. Esta división del espacio de fases en compartimentos se conoce como división de *grano-grueso* del espacio de fases.

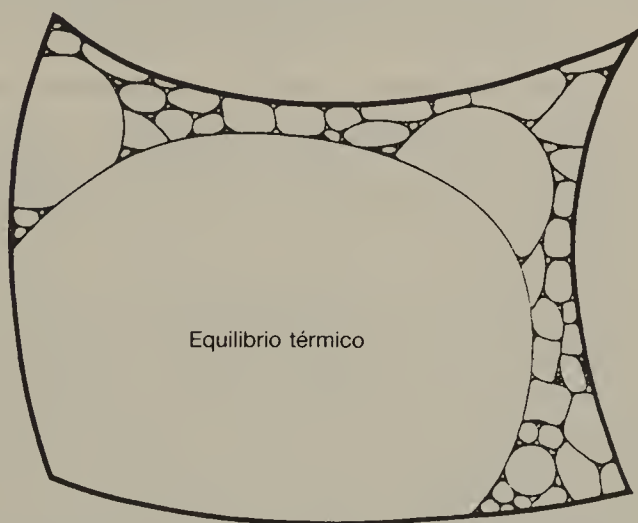


Fig. 7.3. Una división de grano-grueso del espacio de fases en regiones correspondientes a estados que son macroscópicamente indistinguibles uno de otro. La *entropía* es proporcional al logaritmo del volumen en el espacio de fases.

Ahora resultará que algunos de estos compartimentos serán inmensamente mayores que otros. Por ejemplo, consideremos el espacio de fases de un gas en una caja. La mayor parte del espacio de fases corresponderá a estados en los que el gas está muy uniformemente distribuido por toda la caja, con las partículas moviéndose de una forma característica que proporciona una temperatura y presión uniformes. Este tipo característico de movimiento es, en cierto sentido, el más «aleatorio» posible, y se conoce como *distribución maxwe-*

lliana —por el mismo James Clerk Maxwell que ya hemos encontrado antes. Cuando el gas está en tal estado aleatorio se dice que está en *equilibrio térmico*. Existe un volumen enormemente vasto de puntos del espacio de fases que corresponden al equilibrio térmico; los puntos de este volumen describen todas las diferentes configuraciones detalladas de posiciones y velocidades de las partículas individuales que son consistentes con el equilibrio térmico. Este vasto volumen es uno de nuestros compartimentos del espacio de fases —¡el más grande de todos con mucho, que ocupa la totalidad del espacio de fases! Consideremos otro posible estado del gas, por ejemplo aquel en el que todo el gas está concentrado en una esquina de la caja. De nuevo habrá muchos estados detallados individuales distintos, todos los cuales describen el gas concentrado de la misma forma en la esquina de la caja. Todos estos estados son macroscópicamente indistinguibles uno de otro, y los puntos del espacio de fases que los representan constituyen otro compartimento del espacio de fases. Sin embargo, el volumen de este compartimento resulta ser muchísimo más pequeño que el de los estados que representan el equilibrio térmico —¡en un factor de alrededor de $10^{10^{25}}$ si consideramos una caja de un metro cúbico que contiene aire a temperatura ambiente y presión atmosférica en equilibrio y tomamos la región en la esquina de un centímetro cúbico!

Para empezar a apreciar este tipo de discrepancia entre volúmenes del espacio de fases, imaginemos una situación simplificada en la que cierto número de bolas van a distribuirse entre varias celdas. Supongamos que cada celda está vacía o contiene una sola bola. Las bolas representarán las moléculas del gas, y las celdas representan las diferentes posiciones que podrían ocupar las moléculas en la caja. Seleccionemos un pequeño subconjunto de celdas como *especial*; éstas representan las posiciones de las moléculas del gas que corresponden a la región en la esquina de la caja. Supongamos, para ser más precisos, que exactamente una décima parte de las celdas son especiales —digamos que hay n celdas especiales y $9n$ no especiales. (Ver Fig. 7.4.) Queremos distribuir m bolas entre las celdas de una manera aleatoria y encontrar la probabilidad de que todas ellas estén en celdas especiales. Si hay sólo una bola y diez celdas (de modo que tenemos una celda especial) esta probabilidad es evidentemente de un décimo. Lo mismo es válido si hay una bola y cualquier número $10n$ de celdas (con n celdas especiales por lo tanto). Así, para un «gas» con sólo *un* átomo, el compartimento especial que corresponde al gas «concentrado en la esquina» tendrá un volumen de sólo *un décimo* del volumen

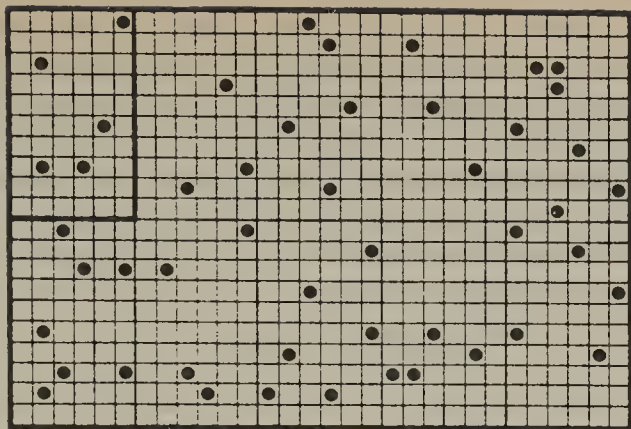


Fig. 7.4. Un modelo para un gas en una caja: cierto número de minúsculas bolas se distribuyen entre un número mucho mayor de celdas. Una décima parte de las celdas están etiquetadas como *especiales*. Éstas son las recuadradas en la esquina superior izquierda.

total del «espacio de fases». Pero si aumentamos el número de bolas, la probabilidad de encontrarlas *todas* en celdas especiales disminuye muy considerablemente. Para *dos* bolas, y digamos veinte celdas * (de las cuales dos son especiales) ($m=2$, $n=2$), la probabilidad es $1/190$, o con cien celdas (con diez especiales) ($m=2$, $n=10$) es de $1/110$; con un número muy grande de celdas la probabilidad se hace $1/100$. Así, el volumen del compartimento especial para un «gas» de-*dos*-átomos es sólo *una centésima* parte del volumen total del espacio de fases. Para tres bolas y treinta celdas ($m=3$, $n=3$), es $1/4060$; y con un número muy grande de celdas se hace $1/1000$ —de modo que para un «gas» de-*tres*-átomos, el volumen del compartimento especial es ahora de *una milésima* parte del volumen del «espacio de fases». Para cuatro bolas y un número muy grande de celdas, la probabilidad se hace $1/10000$. Para cinco bolas y un número muy grande de celdas, la probabilidad se hace $1/100000$, y así sucesivamente. Para m bolas y un número muy grande de celdas, la probabilidad se hace $1/10^m$, de modo que para un «gas» de- m -átomos, el volumen de la región especial es $1/10^m$ del volumen del «espacio de fases». (Esto sigue siendo cierto si se incluye el «momento».)

Podemos aplicar esto a la situación considerada arriba de un gas real en una caja, pero ahora, en lugar de ser sólo una décima parte

* Para m , n cualesquiera la probabilidad es $10^m C_m \div {}^n C_m = (10n)!(n-m)!/n!(10n-m)!$

del total, la región especial ocupa sólo una millonésima (esto es, $1/1000000$) de este total (es decir, un centímetro cúbico en un metro cúbico). Esto significa que en lugar de ser la probabilidad $1/10^m$, ahora es $1/(1000000)^m$, esto es, $1/10^{6m}$. Para el aire ordinario habrá unas 10^{25} moléculas en total en nuestra caja, así que tomamos $m=10^{25}$. Por consiguiente, el compartimento especial del espacio de fases que representa la situación en la que todo el gas está concentrado en la esquina tiene un volumen de ¡sólo

$$1/10^{60\,000\,000\,000\,000\,000\,000\,000\,000\,000}$$

de todo el espacio de fases!

La *entropía* de un estado es una medida del *volumen* V del compartimento que contiene los puntos del espacio de fases que representan a dicho estado. En vista de las enormes diferencias entre estos volúmenes, como antes señalamos, es bueno quizá que no tomemos la entropía como proporcional a dicho volumen sino al *logaritmo* del volumen:

$$\text{entropía} = k \log V.$$

El tomar un logaritmo ayuda a hacer estos números más razonables. El logaritmo * de $10\,000\,000$, por ejemplo, es sólo alrededor de 16. La cantidad k es una constante, llamada *constante de Boltzmann*. Su valor es de alrededor de 10^{-23} julios por grado Kelvin. La razón esencial aquí para tomar un logaritmo es hacer de la entropía una cantidad *aditiva* para sistemas independientes. Así, para dos sistemas físicos completamente independientes, la entropía total de los dos sistemas combinados será la *suma* de la entropías de cada sistema por separado. (Esto es una consecuencia de la propiedad algebraica elemental de la función logarítmica: $\log AB = \log A + \log B$. Si los dos sistemas pertenecen a compartimentos de volumen A y B , en sus respectivos espacios de fases, entonces el volumen del espacio de fases para los dos juntos será su producto AB , puesto que cada posibilidad para un sistema debe contarse una vez por cada posibilidad del otro; por lo tanto la entropía del sistema combinado será realmente la suma de la dos entropías individuales.)

Las enormes discrepancias entre los tamaños de los compartimen-

* El logaritmo utilizado aquí es un logaritmo *natural*, es decir, tomado con base $e=2.7182818285\dots$ en lugar de 10, pero esta diferencia apenas es importante. El logaritmo natural, $x = \log n$, de un número n es la potencia a la que debemos elevar e para obtener n , es decir, la solución de $e^x = n$ (véase la nota a pie de página en p. 124).

tos del espacio de fases se ven más razonables en términos de entropía. La entropía de nuestra caja de gas de un metro cúbico de tamaño, como se describió antes, resultará ser de sólo unos 1400 JK^{-1} ($= 14k \cdot 10^{25}$) veces mayor que la entropía del gas concentrado en la región «especial» de un centímetro cúbico de tamaño (puesto que $\log_e(10^{6 \times 10^{25}})$ es alrededor de 14×10^{25}).

Para dar los valores *reales* de la entropía para estos compartimentos tendríamos que ocuparnos un poco de la cuestión de las unidades a escoger (metros, julios, kilogramos, grados Kelvin, etc.). Esto estaría fuera de lugar aquí y, de hecho, para los tremendamente enormes valores de la entropía que acabo de dar no supone ninguna diferencia esencial la elección particular de unidades. Sin embargo, para ser algo más preciso (para los expertos), advierto que tomaré unidades *naturales*, tal como las proporcionan las reglas de la mecánica cuántica, para las que la constante de Boltzmann resulta ser la *unidad*:

$$k = 1.$$

La segunda ley en acción

Supongamos, ahora, que partimos de un sistema en alguna situación muy especial, como sucedía cuando todo el gas estaba en una esquina de la caja. En los instantes siguientes el gas se expandirá y ocupará rápidamente volúmenes cada vez mayores. Tras un cierto intervalo de tiempo se establecerá el equilibrio térmico. ¿Cuál es nuestra imagen de este proceso en términos del espacio de fases? En cada paso, el estado detallado completo de posiciones y movimientos de todas las partículas del gas vendrá descrito por un simple punto en el espacio de fases. A medida que el gas evoluciona, este punto se mueve por el espacio de fases y sus recorridos precisos describen la historia entera de todas las partículas del gas. El punto parte de una región pequeñísima —la región que representa la colección de posibles estados iniciales para los que todo el gas está en una esquina particular de la caja. A medida que el gas comienza a expandirse, nuestro punto en movimiento entrará en un volumen considerablemente mayor del espacio de fases, que corresponde a los estados en los que el gas se ha extendido algo por toda la caja. El punto en el espacio de fases sigue invadiendo volúmenes cada vez mayores a medida que el gas se extiende más y más, de manera que cada nuevo volumen empequeñece totalmente a todos aquellos en los que el pun-

to había estado antes —¡por factores absolutamente extraordinarios! (ver Fig. 7.5). En cada caso, una vez que el punto ha entrado en el volumen mayor ya no existe la posibilidad (efectiva) de que pueda encontrar cualquiera de los volúmenes anteriores más pequeños. Finalmente se pierde en el volumen más enorme de todos en el espacio de fases: el correspondiente al equilibrio térmico. Este volumen ocupa prácticamente la totalidad del espacio de fases. Podemos asegurar virtualmente que, en su deambular prácticamente aleatorio, nuestro punto en el espacio de fases no encontrará ninguno de los volúmenes más pequeños en un intervalo de tiempo plausible. Una vez que se haya alcanzado el estado de equilibrio térmico, el estado permanecerá allí efectivamente para siempre. Vemos así que la entropía del sistema, que proporciona simplemente una medida logarítmica del volumen del compartimento apropiado del espacio de fases, tendrá esta tendencia inexorable a aumentar * a medida que avanza el tiempo.



Fig. 7.5. La segunda ley de la termodinámica en acción: a medida que transcurre el tiempo, el punto en el espacio de las fases entra en compartimentos de volúmenes cada vez mayores. Por consiguiente, la entropía aumenta continuamente.

* Por supuesto no es cierto que nuestro punto en el espacio de fases no encontrará nunca de nuevo uno de los compartimentos más pequeños. Si esperamos el tiempo suficiente llegará a re-entrar en estos volúmenes relativamente minúsculos. (Esto se conoce como *recurrencia de Poincaré*.) Sin embargo, las escalas de tiempo serían ridículamente largas en la mayoría de las circunstancias, v.g. alrededor de $10^{10^{26}}$ años para el caso del gas que se concentra en un centímetro cúbico en una esquina de la caja. ¡Esto es muchísimo mayor que el tiempo de existencia del Universo! Descartaré esta posibilidad en la discusión que sigue ya que no es realmente relevante para el problema que consideramos.

¡Parece ahora que tuviéramos una *explicación* para la segunda ley! En efecto, podemos suponer que nuestro punto en el espacio de fases no se mueve de ninguna manera especialmente concebida, y si parte de un volumen pequeñísimo en el espacio de fases, correspondiente a una entropía *pequeña*, entonces a medida que el tiempo avanza será extraordinariamente probable que entre en volúmenes del espacio de fases sucesivamente mayores que corresponden a valores gradualmente crecientes de la entropía.

Sin embargo, hay algo un poco extraño en lo que parecemos haber deducido mediante este argumento. Parece que hemos deducido una conclusión con *asimetría temporal*. La entropía *crece* en la dirección *positiva* del tiempo, y por consiguiente debe *disminuir* en la dirección *inversa* del tiempo. ¿De dónde procede esta asimetría temporal? Es seguro que no hemos introducido ninguna ley física con asimetría temporal. La asimetría temporal procede simplemente del hecho de que el sistema *ha partido* de un estado muy especial (esto es, de baja entropía); y al haber empezado así, lo hemos visto evolucionar en la dirección del *futuro* y hemos encontrado que la entropía crece. Este incremento de la entropía está realmente de acuerdo con el comportamiento de los sistemas en nuestro universo real. Pero podríamos perfectamente haber aplicado este mismo argumento en la dirección inversa del tiempo. Especificaremos de nuevo que el sistema está en un estado de baja entropía en algún instante dado, pero ahora preguntamos cuál es la secuencia más probable de estados que *preceden* a éste.

Ensayemos el argumento en forma inversa. Consideremos, como antes, que el estado de baja entropía es el de todo el gas en una esquina de la caja. Nuestro punto en el espacio de fases está ahora en la misma región minúscula de la que partimos antes, pero ahora trataremos de seguir su historia *hacia atrás*. Si imaginamos el punto en el espacio de fases deambulando de una forma completamente aleatoria como antes, entonces esperamos que, a medida que seguimos este movimiento hacia atrás en el tiempo, alcanzará pronto el mismo volumen considerablemente mayor del espacio de fases que alcanzó antes, correspondiente al gas extendido un poco por la caja pero no en equilibrio térmico, y seguirá alcanzando volúmenes cada vez mayores tales que cada nuevo volumen empequeñece totalmente a los previos; y retrocediendo más en el tiempo lo encontraremos en el volumen más grande de todos, que representa el equilibrio térmico. *Ahora* parece que hemos deducido que, dado que todo el gas estaba en cierto instante concentrado en la esquina de la caja, entonces el

modo más probable de que haya llegado ahí es que partió del equilibrio térmico, luego empezó a concentrarse por sí mismo en un extremo de la caja, más y más, hasta que finalmente se agrupó en el pequeño volumen especificado de la esquina. La entropía tendría que haber estado *decreciendo* todo el tiempo: ¡partió del valor alto de equilibrio y luego decreció gradualmente hasta que alcanzó el valor muy bajo correspondiente al gas concentrado en la pequeña esquina de la caja!

¡Por supuesto, esto no se parece nada a lo que realmente sucede en nuestro universo! La entropía no disminuye de esta forma sino que *aumenta*. Si supiéramos que todo el gas estaba concentrado en una esquina de la caja en un instante particular, entonces una situación mucho más probable para *preceder* a aquella hubiera sido la del gas mantenido firmemente en la esquina mediante un tabique que fue rápidamente eliminado. O quizá el gas estuviera en un estado congelado o líquido y fue calentado rápidamente para volverse gaseoso. Para cualquiera de estas posibles alternativas la entropía era aún *menor* en los estados previos. La segunda ley realmente se satisfacía y la entropía estaba creciendo todo el tiempo —dicho de otra forma, en la dirección *inversa* del tiempo estaba *disminuyendo* realmente. ¡Ahora vemos que nuestro argumento nos ha dado una respuesta totalmente equivocada! Nos ha dicho que la forma más probable de tener el gas en la esquina de la caja sería partir del equilibrio térmico y entonces, con la entropía reduciéndose continuamente, el gas se agruparía en la esquina; mientras que, de hecho, es extraordinariamente *improbable* que suceda de esta forma en nuestro mundo real. En nuestro mundo, el gas partiría de un estado aún *menos* probable (esto es, de menor entropía), y la entropía *aumentaría* continuamente hasta el valor que tiene posteriormente cuando el gas está concentrado en la esquina.

Nuestro argumento parece ser bueno cuando se aplica en la dirección del futuro, pero no en la dirección del pasado. Para la dirección del *futuro* prevemos correctamente que, dondequiera que parta el gas en la esquina, lo más probable que suceda en el futuro es que se *alcance* el equilibrio térmico, y *no* que aparezca un tabique repentinamente, o que el gas repentinamente se congele o se haga líquido. Semejantes alternativas extrañas representarían precisamente el tipo de comportamiento con disminución de entropía en la dirección del futuro que nuestro argumento del espacio de fases parece correctamente descartar. Pero en la dirección del *pasado*, semejantes alternativas «anómalas» son las que sería probable que sucedieran —y no

nos parecen ser en absoluto anómalas. ¡Nuestro argumento del espacio de fases nos dio la respuesta totalmente equivocada cuando tratamos de aplicarlo en la dirección inversa en el tiempo!

Evidentemente esto arroja algunas dudas sobre el argumento original. *No* hemos deducido la segunda ley. Lo que realmente mostraba el argumento era que para un estado de baja entropía dado (digamos para un gas concentrado en una esquina de una caja), entonces, *en ausencia* de otros factores que impongan restricciones al sistema, se espera que la entropía aumente en *ambas* direcciones en el tiempo a partir del estado dado (ver Fig. 7.6). El argumento no ha funcionado en la dirección del pasado precisamente porque *existían* tales factores. Había algo realmente que en el pasado limitaba al sistema. Algo *obligaba* a la entropía a ser baja en el pasado. La tendencia hacia una entropía alta en el futuro no es una sorpresa. Los estados de alta entropía son, en cierto sentido, los estados «naturales» que no necesitan más explicación. Pero los estados de baja entropía en el pasado son un enigma. ¿Qué obligaba a la entropía de nuestro mundo a ser tan baja en el pasado? La presencia común de estados para los que la entropía es absurdamente baja es un hecho sorprendente del universo real en el que habitamos —aunque tales estados son tan comunes y familiares para nosotros que no tendemos normalmente a considerarlos como sorprendentes. ¡Nosotros mismos somos configuraciones de entropía ridículamente ínfima! El argumento anterior muestra que no deberíamos sorprendernos si, *dado* un estado de baja entropía, la entropía resulta ser mayor en un instante posterior. ¡Lo que

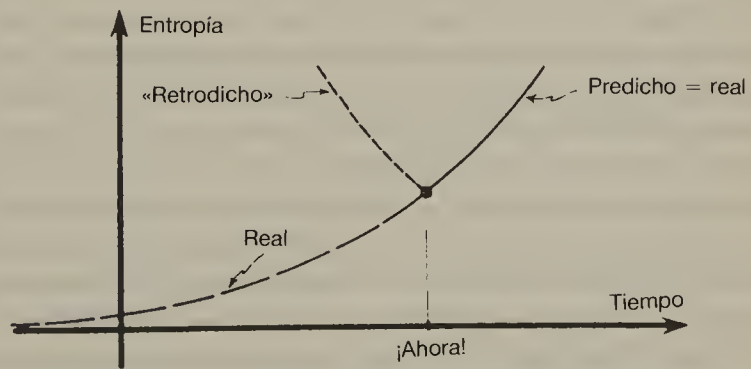


Fig. 7.6. Si utilizamos el argumento representado en la Fig. 7.5 en la dirección inversa en el tiempo «retrodecimos» que la entropía aumentará también hacia el *pasado* desde su valor presente. Esto está en abierta contradicción con lo observado.

debería sorprendernos es que la entropía se haga cada vez más ridículamente pequeña a medida que nos adentramos cada vez más en el pasado!

El origen de la baja entropía en el Universo

Trataremos de entender de dónde procede esta «sorprendentemente» baja entropía en el mundo real que habitamos. Empecemos por nosotros mismos. Si podemos entender de dónde procede nuestra propia baja entropía, entonces seríamos capaces de ver de dónde procede la baja entropía del gas mantenido por el tabique —o del vaso de agua en la mesa, o del huevo mantenido sobre la sartén, o del terrón de azúcar mantenido sobre la taza de café. En cada caso una persona o colección de personas (¡o quizá una gallina!) era responsable directa o indirectamente. Era, en gran medida, una pequeña parte de nuestra propia baja entropía la que realmente se utilizaba para establecer estos otros estados de baja entropía. Podrían haber estado involucrados otros factores. Quizá se utilizó una bomba de vacío para aspirar el gas hasta la esquina de la caja tras el tabique. Si la bomba no se operó manualmente pudo haberse quemado algún «combustible fósil» (v.g. petróleo) para proporcionar la energía de baja entropía necesaria para su operación. Tal vez la bomba fue operada eléctricamente y dependía, en alguna medida, de la energía de baja entropía almacenada en el uranio de alguna central nuclear. Más tarde volveré a estas fuentes de baja entropía, pero consideremos primero simplemente la baja entropía en nosotros mismos.

¿De dónde *procede* realmente nuestra propia baja entropía? La organización de nuestros cuerpos procede del alimento que comemos y del oxígeno que respiramos. Con frecuencia se oye decir que obtenemos *energía* de nuestra ingestión de alimentos y oxígeno, pero hay un sentido evidente en el que esto no es correcto. Es cierto que el alimento que consumimos se combina con el oxígeno que introducimos en nuestros cuerpos, y que esto nos proporciona energía. Pero esta energía, en su mayor parte, escapa de nuevo de nuestros cuerpos, principalmente en forma de calor. Puesto que la energía se conserva, y puesto que el contenido real de energía de nuestros cuerpos permanece más o menos constante a lo largo de nuestra vida adulta, no hay necesidad de *añadir* nada al contenido de energía de nuestros cuerpos. No *necesitamos* más energía dentro de nosotros de la que ya tenemos. De hecho añadimos algo a nuestro contenido energético cuando au-

mentamos de peso, ¡pero normalmente esto no se considera deseable! Cuando crecemos en la infancia también incrementamos considerablemente nuestro contenido energético a medida que formamos nuestros cuerpos; no es esto en lo que estoy interesado aquí. La cuestión es cómo nos mantenemos *vivos* a lo largo de nuestra vida normal (principalmente adulta). Para esto no *necesitamos* añadir nada a nuestro contenido energético.

Sin embargo, necesitamos reemplazar la energía que perdemos continuamente en forma de calor. En realidad, cuanto más «energéticos» somos, más energía perdemos de esta forma. Toda esta energía debe ser reemplazada. El calor es la forma más *desordenada* de energía que existe, es decir, es la forma de energía con mayor entropía. Tomamos energía en forma de *baja-entropía* (alimento y oxígeno) y la desechamos en una forma de *alta-entropía* (calor, dióxido de carbono, excrementos). No necesitamos obtener energía de nuestro medio ambiente puesto que la energía *se conserva*; pero estamos luchando continuamente contra la segunda ley de la termodinámica. La entropía *no* se conserva; está *aumentando* todo el tiempo. Para mantenernos vivos necesitamos reducir la entropía que hay en nosotros. Lo hacemos alimentándonos con una combinación de comida y oxígeno atmosférico, combinándolos dentro de nuestro cuerpo, y desechando la energía, que de no hacerlo hubiéramos ganado, en una forma de alta-entropía. De este modo, podemos impedir que aumente la entropía en nuestros cuerpos, y podemos mantener (e incluso incrementar) nuestra organización interna. (Ver Schrödinger, 1967.)

¿De dónde procede este suministro de baja entropía? Si el alimento que estamos comiendo es carne (¡o setas!), entonces *este* alimento, al igual que nosotros, tendría que depender de una fuente externa adicional de baja-entropía que proporcione y mantenga su estructura de baja-entropía. Esto simplemente desplaza a alguna otra parte el problema del origen de la baja entropía externa. Supongamos que nosotros (o el animal, o la seta) estamos consumiendo una *planta*. Debemos estar todos profundamente agradecidos a las plantas verdes —ya sea directa o indirectamente— por su inteligencia: tomar dióxido de carbono atmosférico, separar el oxígeno del carbono, y utilizar el carbono para formar su propia substancia. Este procedimiento, la *fotosíntesis*, produce una gran reducción de la entropía. Nosotros mismos hacemos uso efectivo de esta separación de baja-entropía mediante la recombinación simple del oxígeno y el carbono dentro de nuestros cuerpos. ¿Cómo es que las plantas verdes son capaces de conseguir esta mágica reducción de entropía? Lo hacen utilizando la

luz del Sol. La luz que procede del Sol trae energía a la Tierra en una forma de relativamente baja-entropía: en los fotones de la luz visible. La Tierra, incluyendo sus habitantes, no *retiene* esta energía sino que la re-irradia (después de algún tiempo) toda hacia el espacio. Sin embargo, la energía re-irradiada está en una forma de *alta*-entropía, que se llama «calor radiante» —lo que significa fotones infrarrojos. ¡Contrariamente a la impresión común, la Tierra (junto con sus habitantes) *no* gana energía del Sol! Lo que hace la Tierra es tomar energía en una forma de *baja*-entropía y luego arrojarla *toda* al espacio, pero en una forma de *alta*-entropía (Fig. 7.7). Lo que el Sol ha hecho por nosotros es suministrarnos una fuente enorme de baja entropía. Nosotros (por vía de la «inteligencia» de las plantas) hacemos uso de ésta, extrayendo finalmente una mínima parte de su baja entropía y transformándola en las intrincadas estructuras organizadas que somos nosotros mismos.



Fig. 7.7. Cómo hacemos uso del hecho de que el Sol es un punto caliente en la oscuridad del espacio.

Veamos, desde el punto de vista global con relación al Sol y la Tierra, qué ha sucedido con la energía y la entropía. El Sol emite energía en forma de fotones de luz visible. Algunos de estos fotones son absorbidos por la Tierra y su energía es re-irradiada en forma de fotones infrarrojos. Ahora bien, la diferencia crucial entre los fotones de luz visible y los infrarrojos es que los primeros tienen una frecuencia mayor y, por lo tanto, tienen más energía individualmente que los

últimos. (Recordemos la fórmula de Planck $E=h\nu$, dada en p. 292. Ésta nos dice que cuanto mayor sea la frecuencia de un fotón, mayor será su energía.) Puesto que cada uno de los fotones de luz visible tiene mayor energía que cada uno de los infrarrojos, los fotones de luz visible que llegan a la Tierra deben hacerlo en número *menor* que los fotones infrarrojos que dejan la Tierra, de modo que la *energía* que entra en la Tierra compense la que la abandona. La energía que la Tierra arroja al espacio se distribuye sobre muchos más grados de libertad que la energía que se recibe del Sol. Puesto que hay muchos más grados de libertad involucrados cuando la energía es devuelta, el volumen del espacio de fases es mucho mayor, y la *entropía* ha aumentado enormemente. Las plantas verdes, al tomar energía en forma de baja-entropía (relativamente *pocos* fotones de luz visible) han sido capaces de alimentarse de esta baja entropía y proporcionarnos esta separación oxígeno-carbono que necesitamos.

¡Todo esto se hace posible por el hecho de que el Sol es un *punto-caliente* en el cielo! El cielo está en un estado de temperatura desigual: una pequeña región, la ocupada por el Sol, está a una temperatura mucho más alta que el resto. Este hecho nos proporciona la poderosa fuente de baja entropía deseada. La Tierra obtiene energía de este punto-caliente en una forma de baja-entropía (pocos fotones), y la re-irradia a las regiones frías en una forma de alta-entropía (muchos fotones).

¿Por qué es el Sol ese punto-caliente? ¿Cómo ha podido conseguir esa desigualdad de temperatura y proporcionarnos así un estado de baja entropía? La respuesta es que se ha formado por contracción gravitatoria a partir de una previa distribución uniforme de gas (principalmente hidrógeno). A medida que se contrae, en los primeros pasos de su formación, el Sol se calienta. Seguiría contrayéndose y calentándose aún más si no fuera porque, cuando su temperatura y su presión alcanzan un cierto punto, encuentra otra fuente de energía, además de la contracción gravitatoria, en las *reacciones termonucleares*: la fusión de núcleos de hidrógeno en núcleos de helio para dar energía. Sin las reacciones termonucleares el Sol se hubiera hecho mucho *más caliente* y más pequeño de lo que es ahora, hasta que finalmente hubiera muerto. Las reacciones nucleares han impedido que el Sol se vuelva demasiado caliente, deteniendo su contracción, y han estabilizado el Sol a una temperatura que es apropiada para nosotros, haciendo posible que siga brillando durante un tiempo mucho mayor de lo que lo hubiera hecho en otro caso.

Es importante notar que, aunque las reacciones termonucleares

son indudablemente muy importantes al determinar la naturaleza y la cantidad de la energía radiada por el Sol, es la *gravitación* lo que es crucial. (En realidad la potencialidad para las reacciones termonucleares *da* una contribución altamente significativa al bajo valor de la entropía solar, pero las cuestiones que plantea la entropía de la fusión son delicadas y una completa discusión de ellas sólo serviría para complicar el argumento sin afectar a la conclusión final.) ² ¡Sin gravedad ni siquiera existiría el Sol! El Sol brillaría aun sin reacciones termonucleares —aunque no de una forma apropiada para nosotros— pero *no* estaría brillando en absoluto sin la gravedad que se necesita para mantenerlo unido y proporcionar las temperaturas y presiones necesarias. Sin gravedad, todo lo que tendríamos sería un gas frío y difuso en el lugar del Sol y *¡no* habría punto-caliente en el cielo!

No he discutido todavía la fuente de la baja entropía en los «combustibles fósiles» de la Tierra, pero las consideraciones son básicamente las mismas. Según la teoría convencional, todo el petróleo (y gas natural) en la Tierra procede de la vida vegetal prehistórica. De nuevo son las plantas las que resultan haber sido responsables de esta fuente de baja entropía. Las plantas prehistóricas obtuvieron su energía del Sol —de modo que debemos volver otra vez a la acción gravitatoria que formó el Sol a partir de un gas difuso. Hay una interesante teoría «inconformista» alternativa sobre el origen del petróleo en la Tierra, debida a Thomas Gold, que cuestiona esta opinión convencional sugiriendo que hay mucho más petróleo en la Tierra del que podría haber surgido a partir de plantas prehistóricas. Gold cree que el petróleo estaba atrapado en el interior de la Tierra cuando ésta se formó, y desde entonces ha estado rezumando continuamente hacia fuera acumulándose en bolsas subterráneas ³. No obstante, según la teoría de Gold el petróleo habría sido en cualquier caso sintetizado por la luz del Sol, aunque esta vez en el espacio, antes de que la Tierra se formase. De nuevo sería el Sol formado gravitatoriamente el responsable.

¿Qué hay de la energía nuclear de baja entropía en el isótopo de uranio-235 que se utiliza en las centrales nucleares? Ésta no procedía originalmente del Sol (aunque podría haber pasado perfectamente por el Sol en alguna etapa) sino de alguna otra estrella que ¡explotó hace muchos miles de millones de años en una explosión de supernova! En realidad, el material fue recogido de *muchas* de estas estrellas explosivas. El material de estas estrellas fue arrojado al espacio por la explosión y parte de él se juntó finalmente (mediante la actuación del Sol) para proporcionar los elementos pesados en la Tierra, incluyendo

todo este uranio-235. Cada núcleo, con su reserva de energía de baja-entropía, procede de los violentos procesos nucleares que tuvieron lugar durante la explosión de la supernova. La explosión ocurrió a consecuencia del colapso gravitatorio ⁴ de una estrella cuya masa era demasiado grande para poder ser mantenida por las fuerzas debidas a la presión térmica. Como resultado de este colapso, y subsiguiente explosión, quedó un pequeño núcleo —probablemente en la forma de lo que se conoce como una *estrella de neutrones* (volveré a esto más adelante). La estrella se habría contraído gravitatoriamente en origen a partir de una difusa nube de gas y mucho de este material original, incluyendo nuestro uranio-235, habría sido arrojado al espacio. Sin embargo, hubo una enorme ganancia en entropía debida a la contracción gravitatoria, a causa de la estrella de neutrones que quedó. De nuevo fue la *gravedad* la responsable última —esta vez al causar la condensación (finalmente violenta) de gas difuso en una estrella de neutrones.

Parece que hemos llegado a la conclusión de que toda esta notable pequeñez de la entropía que encontramos —y que proporciona el aspecto más enigmático de la segunda ley— debe atribuirse al hecho de que se pueden ganar grandes cantidades de entropía mediante la contracción gravitatoria de gas difuso en estrellas. ¿De dónde procede todo este gas difuso? Es precisamente el hecho de que este gas empezó siendo *difuso* el que nos proporciona una enorme reserva de baja entropía. Aún estamos viviendo de esta reserva de baja entropía, y continuaremos haciéndolo durante mucho tiempo. Es la potencialidad del agrupamiento gravitatorio de este gas la que nos ha dado la segunda ley. Además, no es solamente la segunda ley lo que este agrupamiento gravitatorio ha producido, sino algo mucho más preciso y detallado que el simple enunciado: «la entropía del mundo empezó siendo muy baja». La entropía podía habernos sido dada «baja» de muchas *otras* formas diferentes, es decir, podría haber habido mucho «orden manifiesto» en el Universo primitivo, pero muy diferente de ese «orden» que se nos presenta en el mundo real. (Imaginemos que el universo hubiera sido un dodecaedro regular —como hubiera gustado a Platón— o alguna otra forma geométrica improbable. Esto sería realmente «orden manifiesto», ¡pero no del tipo que esperamos encontrar en el universo primitivo *real*!) Debemos comprender de dónde procede todo este gas difuso —y para ello tendremos que volver a nuestras teorías cosmológicas.

La cosmología y el big bang

Hasta donde podemos asegurar mediante la utilización de nuestros telescopios más potentes —tanto ópticos como radiotelescopios— el Universo aparece bastante uniforme a gran escala; pero, lo que es más notable, está en *expansión*. Cuanto más lejos miramos, más rápido parece que las galaxias distantes (y los cuásares aún más distantes) se alejan de nosotros. Es como si el Universo mismo se hubiera creado en una gigantesca explosión, un suceso conocido como el *big bang*, que ocurrió hace unos diez mil millones de años *. Un impresionante apoyo adicional para esta uniformidad, y para la existencia real del big bang, viene de lo que se conoce como *radiación de fondo de cuerpo negro*. Ésta es radiación térmica —fotones que se mueven aleatoriamente, sin ninguna fuente discernible— correspondiente a una temperatura de aproximadamente $2,7^\circ$ absolutos ($2,7\text{ K}$), esto es $-270,3^\circ$ Celsius, o $454,5^\circ$ bajo cero Fahrenheit. Esto puede parecer una temperatura *muy* fría —¡como de hecho lo es!— pero parece ser el residuo del flash del propio big bang. Puesto que el Universo se ha expandido en un factor tan enorme desde el momento del big bang, esta bola de fuego inicial se ha extendido multiplicándose por un factor absolutamente enorme. Las temperaturas durante el big bang excedían cualquier temperatura que pueda darse en el tiempo presente pero, debido a la expansión, esta temperatura se ha enfriado hasta el pequeñísimo valor que tiene ahora la radiación negra de fondo. La presencia de esta radiación de fondo fue *predicha* por el físico ruso-americano George Gamow en 1948 sobre la base de la imagen del big bang ahora estándar. Fue observada (accidentalmente) por Penzias y Wilson en 1965.

Abordaré una cuestión que a menudo desconcierta a la gente. Si las galaxias distantes en el Universo están todas alejándose de nosotros, ¿no significa esto que nosotros mismos estamos ocupando algún lugar central muy especial? ¡Nada de eso! La misma recesión de galaxias distantes se vería *dondequiera* que pudiéramos estar situados en el Universo. La expansión es uniforme a gran escala, y ninguna localización particular es preferida a otra. Esto se representa a me-

* Actualmente hay una enconada discusión a propósito del valor de esta cifra, que está entre 6×10^9 y 1.5×10^{10} años. Estos valores son considerablemente mayores que los 10^9 años que originalmente parecieron adecuados después de que las observaciones iniciales de Edwin Hubble, alrededor de 1930, demostraron que el Universo está expandiéndose.

nudo en términos de un globo que se infla (Fig. 7.8). Supongamos que hay manchas en el globo que representan las diferentes galaxias, y tomemos la propia superficie bi-dimensional del globo para representar el universo espacial tri-dimensional completo. Es evidente que con referencia a *cada* punto del globo *todos* los demás puntos se alejan de él. No hay ningún punto del globo preferido, a este respecto, a cualquier otro punto. Análogamente, como se ve a partir desde la posición de cada galaxia en el Universo, todas las demás galaxias parecen alejarse de ella igualmente en todas direcciones.



Fig. 7.8. Se puede hacer una analogía entre la expansión del Universo y la superficie de un globo que se hincha. Todas las galaxias se alejan una de otra.

Este globo en expansión proporciona una imagen bastante buena de uno de los tres modelos de universo estándar del tipo llamado de *Friedman-Robertson-Walker* (FRW): el modelo FRW con *curvatura positiva* espacialmente cerrado. En los otros dos modelos-FRW (con curvatura nula o negativa) el universo se expande de un modo semejante, pero en lugar de tener un universo espacialmente finito, como indica la superficie del globo, tenemos un universo *infinito* con un número infinito de galaxias.

En el más fácil de comprender de estos dos modelos infinitos, la geometría es *euclídea*, esto es, tiene curvatura *nula*. Pensemos en un plano ordinario, que representa el universo espacial completo, en donde marcamos puntos que representan galaxias. A medida que el universo evoluciona con el tiempo, estas galaxias se alejan unas de otras de manera uniforme. Pensemos en esto en términos de *espacio-tiempo*. Según esto, tenemos un plano euclídeo diferente para cada «instante de tiempo», y todos estos planos se conciben como apilados uno sobre otro, de modo que tenemos una imagen de todo el espacio-tiempo a la vez (Fig. 7.9). Las galaxias se representan ahora como

curvas —las *líneas de universo* de las historias de las galaxias— y estas curvas divergen en la dirección del futuro. De nuevo no existe ninguna línea-de-universo de una galaxia preferida.

Para el modelo FRW restante, el modelo con curvatura *negativa*, la geometría espacial es la geometría *lobachevskiana* no-euclídea descrita en el Capítulo 5 e ilustrada con el grabado de Escher mostrado en la Figura 5.2 (p. 205). Para la descripción del espacio-tiempo necesitamos uno de estos espacios de Lobachevsky para cada «instante de tiempo», y apilamos todos ellos uno encima de otro para dar una imagen del espacio-tiempo completo (Fig. 7.10) ⁵. Una vez más, las líneas-de-universo de las galaxias son curvas que se separan unas de otras en la dirección del futuro, y no hay ninguna galaxia preferida.

Por supuesto, en todas estas descripciones hemos suprimido una de las tres dimensiones espaciales (como hicimos en el Capítulo 5, cf. p. 251) para dar un espacio-tiempo tri-dimensional más cercano a lo que sería necesario para la imagen completa del espacio-tiempo tetra-dimensional. ¡Aún así es difícil visualizar el espacio-tiempo con curvatura positiva sin descartar una dimensión espacial más! Hagámoslo y representemos el universo espacialmente cerrado con curvatura positiva mediante un *círculo* (uni-dimensional), en lugar de la superficie esférica (bi-dimensional) que había sido la superficie del globo. A medi-

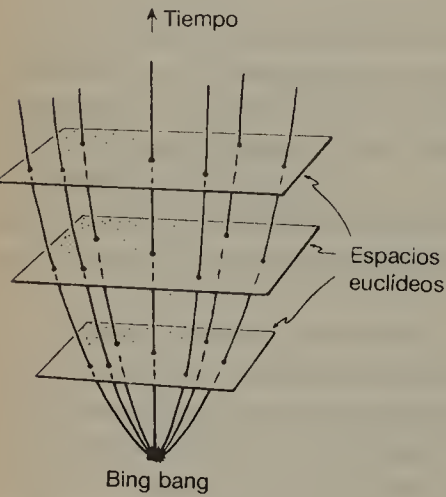


Fig. 7.9. Representación espacio-temporal de un universo en expansión con secciones espaciales euclídeas (se muestran dos dimensiones espaciales).

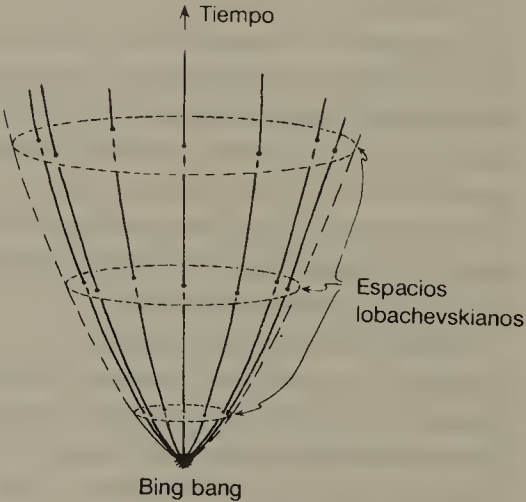


Fig. 7.10. Representación espacio-temporal de un universo en expansión con secciones espaciales lobachevskianas (se muestran dos dimensiones espaciales).

da que el universo se expande, este círculo aumenta de tamaño y podemos representar el espacio-tiempo apilando estos círculos (un círculo para cada «instante de tiempo») uno encima de otro para obtener una especie de cono curvo (Fig. 7.11(a)). Ahora bien, de la ecuaciones de la relatividad general de Einstein se deduce que este universo cerrado positivamente no puede seguir expandiéndose para siempre. Después de alcanzar una etapa de máxima expansión colapsa sobre sí mismo para llegar finalmente otra vez al tamaño nulo en una especie de big bang al revés (Fig. 7.11(b)). Este big bang invertido en el tiempo se denomina a veces el *big crunch* *. Los modelos FRW de universo (infinito) con curvatura negativa y nula no colapsan de esta forma. En lugar de llegar a un big crunch, continúan expandiéndose para siempre.

Esto es cierto al menos para la relatividad general *estándar* en la que la llamada *constante cosmológica* se considera nula. Con valores

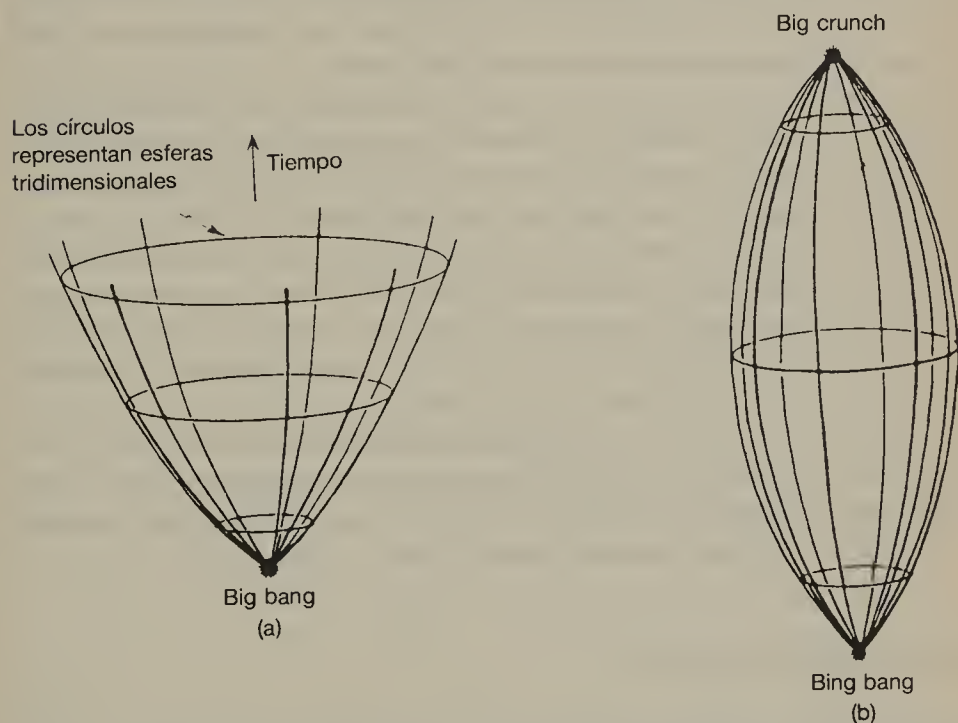


Fig. 7.11. Representación espacio-temporal de un universo en expansión con secciones espaciales esféricas (sólo se muestra una dimensión espacial). (b) Este universo puede llegar a recolapsar en un big crunch final.

* Literalmente «gran crujido». [Nota del traductor.]

no nulos apropiados para la constante cosmológica es posible tener modelos de universo espacialmente infinitos que colapsan a un big crunch, o modelos finitos con curvatura positiva que se expanden indefinidamente. La presencia de una constante cosmológica no nula complicaría ligeramente la discusión, aunque no de manera significativa para nuestros propósitos. Por simplicidad, consideraré nula la constante cosmológica *. En el momento de escribir este libro se sabe, a partir de las observaciones, que la constante cosmológica debe ser muy pequeña, y los datos son consistentes con que sea nula. (Para más información sobre los modelos cosmológicos, ver Rindler, 1977.)

Desgraciadamente los datos no son todavía suficientemente buenos para apuntar claramente a uno u otro de los modelos cosmológicos propuestos (ni para determinar si la presencia o no de una pequeñísima constante cosmológica pudiera tener algún efecto global significativo). Por otra parte, los datos parecerían indicar que el universo tiene curvatura espacial negativa (con geometría de Lobachevsky a gran escala) y que continuará expandiéndose indefinidamente. Esto se basa principalmente en observaciones sobre la cantidad real de materia que parece estar presente en forma visible. Sin embargo, puede haber enormes cantidades de materia invisible distribuida por todo el espacio, en cuyo caso el universo podría estar curvado positivamente y podría colapsar finalmente en un big crunch —aunque sólo en una escala de tiempo mucho mayor que la de los 10^{10} años, aproximadamente, que lleva existiendo el universo. Para que este colapso fuera posible tendría que haber esparcida por el espacio, en esta forma invisible (la postulada «materia oscura»), una cantidad de materia unas treinta veces mayor que la que puede ser discernida a través de los telescopios. Existe cierta buena evidencia indirecta de que realmente hay presente una cantidad substancial de material oscuro, pero es todavía una cuestión abierta el que sea o no *suficiente* para «cerrar el universo» (o hacerlo espacialmente plano) y colapsarlo.

La bola de fuego primordial

Volvamos a nuestra búsqueda del origen de la segunda ley de la termodinámica. Hemos rastreado este origen hasta la presencia del gas difuso a partir del cual se condensaron las estrellas. ¿Qué es este

* Einstein introdujo la constante cosmológica en 1917, pero se retractó hacia 1931, considerando su introducción anterior como ¡su «mayor error»!

gas? ¿De dónde procede? En su mayor parte es hidrógeno, pero también hay alrededor de un 23 % (en masa) de helio y pequeñas cantidades de otros materiales. Según la teoría estándar este gas fue escudido como consecuencia de la explosión que creó el universo: el big bang. Sin embargo, es importante que no pensemos en esta explosión como en una explosión ordinaria del tipo conocido, en la que el material se expele a partir de algún punto central hacia el espacio pre-existente. ¡Aquí, el propio espacio es *creado* en la explosión y no existe, ni existió, punto central! Quizá sea más fácil visualizar la situación en el caso de curvatura positiva. Consideremos de nuevo la Figura 7.11, o el globo inflado de la Figura 7.8. No hay ningún «espacio vacío preexistente» en el que se disperse el material producido por la explosión. El espacio mismo, esto es, la «superficie del globo», nace de la explosión. Debemos darnos cuenta de que, si bien en nuestras representaciones para el caso de curvatura positiva hemos utilizado un «espacio circundante» —el espacio euclídeo en el que se encuentra el globo, o el espacio tri-dimensional en el que se muestra el espacio tiempo en la figura 7.11—, no debemos considerar que estos espacios circundantes tengan una realidad física. El espacio en el interior o el exterior del globo está allí sólo para ayudarnos a visualizar la superficie del globo. Es *únicamente* la superficie del globo la que representa el espacio físico del universo. Vemos ahora que no hay punto central del que emane la materia del big bang. El «punto» que parece estar en el centro del globo no es parte del universo, sino que es simplemente una ayuda para nuestra visualización del modelo. El material que explotó en el big bang sencillamente está ¡distribuido de modo uniforme por *todo* el espacio del universo!

La situación es la misma (aunque quizá un poco más difícil de visualizar) para los otros dos modelos estándar. El material no estuvo nunca concentrado en un punto del espacio. Llenaba uniformemente la *totalidad* del espacio ¡desde el mismo comienzo!

Esta imagen está subyacente en la teoría del *big bang caliente* conocida como el *modelo estándar*. En esta teoría, el universo, instantes después de su creación, estaba en un estado térmico extremadamente caliente: la *bola de fuego primordial*. Se han realizado cálculos detallados acerca de la naturaleza y proporciones de los constituyentes iniciales de esta bola de fuego, y de cómo cambiaron estos constituyentes a medida que la bola de fuego (que era el universo entero) se expandía y enfriaba. Puede parecer extraño que se puedan realizar cálculos fiables para describir un estado del universo que era tan diferente del de nuestra era presente. Sin embargo, la física sobre la

que se basan estos cálculos no se pone en duda, ¡siempre que no nos preguntemos qué sucedió *antes* de 10^{-4} segundos tras la creación! Desde ese instante, una diezmilésima de segundo después de la creación, hasta unos tres minutos más tarde se ha calculado el comportamiento con gran detalle (cf. Weinberg, 1977) —y, curiosamentè, nuestras teorías físicas bien establecidas, derivadas a partir de un conocimiento experimental de un universo ahora en un estado muy diferente, resultan muy adecuadas para esto ⁶. Estos cálculos implican finalmente que, distribuidos de manera uniforme por todo el universo, debe haber muchos fotones (es decir, luz), electrones y protones (los dos constituyentes del hidrógeno), algunas partículas α (los núcleos de helio), un número aún menor de deuterones (los núcleos del deuterio, un isótopo pesado del hidrógeno), y trazas de otros tipos de núcleos —con quizá también un gran número de partículas «invisibles», tales como los neutrinos, que apenas dejarían sentir su presencia. Los constituyentes *materiales* (principalmente protones y electrones) se combinarían para producir el gas (principalmente hidrógeno) a partir del cual se formaron las estrellas unos 10^8 años después del big bang.

Sin embargo, las estrellas no se formaron inmediatamente. Tras la posterior expansión y enfriamiento del gas fueron necesarias concentraciones de gas en ciertas regiones para que los efectos gravitatorios locales pudieran superar la expansión global. Aquí entramos en la controvertida y no resuelta cuestión de cómo se formaron realmente las galaxias y qué irregularidades debían estar presentes para que fuera posible su formación. No quiero entrar en esta discusión. Aceptemos que debía estar presente algún tipo de irregularidad en la distribución inicial del gas, y que el tipo correcto de agrupamiento gravitatorio se inició de tal modo que las galaxias pudieron formarse con sus cientos de miles de millones de estrellas constituyentes.

Hemos descubierto de dónde procede el gas difuso. Procedía de la misma bola de fuego que constituyó el propio big bang. El hecho de que este gas se haya distribuido de manera muy uniforme por todo el espacio es lo que nos ha dado la segunda ley —en la forma detallada en que esta ley ha llegado a nosotros— una vez que se haya hecho disponible el procedimiento de agrupamiento gravitatorio que incrementa la entropía. ¿Hasta qué punto está uniformemente distribuido el material que constituye el universo real? Hemos señalado que las estrellas están agrupadas en galaxias. Las galaxias, por su parte, también están agrupadas en cúmulos de galaxias; y los cúmulos lo están en los llamados supercúmulos. Hay incluso alguna evidencia de que estos supercúmulos se reúnen en unos agrupamientos aún ma-

yores que se conocen como complejos de supercúmulos. Es importante señalar, no obstante, que toda esta irregularidad y agrupamiento es una «minucia» en comparación con la impresionante uniformidad de la estructura del universo en general. Cuanto más atrás ha sido posible mirar en el universo, y mayor es la porción del universo que ha sido posible explorar, más uniforme parece. La radiación de fondo de cuerpo negro proporciona la evidencia más impresionante a este respecto. Nos dice, en concreto, que cuando el Universo tenía apenas un millón de años, y sobre un dominio que se extendía a 10^{23} kilómetros (a una distancia de nosotros que abarcaría unas 10^{10} galaxias), el universo y todo su contenido material era *uniforme* hasta una parte en cien mil (cf. Davies, 1987). El universo, pese a sus orígenes violentos, era realmente muy uniforme en sus etapas primitivas.

Por consiguiente, fue la bola de fuego inicial la que dispersó tan uniformemente este gas por todo el espacio. Aquí es donde nos ha llevado nuestra búsqueda.

¿Explica el big bang la segunda ley?

¿Ha concluido nuestra búsqueda? ¿Está «explicado» el hecho enigmático de que la entropía de nuestro universo empezara tan baja —el hecho que nos ha dado la segunda ley de la termodinámica— sólo por la circunstancia de que el universo comenzó en un big bang? Un poco de reflexión sugiere que hay algo paradójico implícito en esta idea. No puede ser la respuesta verdadera. Recordemos que la bola de fuego primordial era una estado *térmico* —un gas caliente en equilibrio térmico en expansión. Recordemos también que el término «equilibrio térmico» designa el estado de *máxima* entropía. (Así es como designábamos al estado de máxima entropía de un gas en una caja.) Sin embargo, la segunda ley exige que en su estado inicial la entropía de nuestro universo esté en alguna especie de *mínimo*, ¡no en un máximo!

¿Qué es lo que anda mal? Una respuesta estándar diría aproximadamente lo siguiente:

Cierto, la bola de fuego estaba efectivamente en equilibrio térmico en el inicio, pero el universo era muy pequeño en ese tiempo. La bola de fuego representaba el estado de entropía máxima permitido para un universo de *tal* minúsculo tamaño, pero la entropía entonces permitida habría sido minúscula en comparación con la que es posible para un universo del tamaño del que hoy encontramos. A medida que el universo se expandía la entropía máxima

permitida incrementó con el tamaño del universo, pero la entropía real del universo quedó muy por debajo de este máximo permitido. La segunda ley aparece debido a que la entropía real está siempre tratando de alcanzar este máximo permitido.

Sin embargo, un pequeño examen nos dice que ésta no puede ser la explicación correcta. Si lo fuera, entonces, en el caso de un modelo de universo (cerrado espacialmente) que colapse finalmente en un big crunch, el argumento se aplicaría otra vez en la dirección *inversa* del tiempo. Cuando el universo alcanzara finalmente un tamaño minúsculo comenzaría de nuevo a haber un techo para los posibles valores de la entropía. La misma limitación que servía para darnos una baja entropía en las etapas muy primitivas del universo en expansión se aplicaría de nuevo en las etapas finales del universo en contracción. Era un límite para la baja entropía «en el principio del tiempo» el que nos daba la segunda ley, según la cual la entropía está creciendo con el tiempo. Si este mismo límite para la baja entropía tuviera que aplicarse al «final del tiempo» entonces ¡entraríamos en un gran conflicto con la segunda ley de la termodinámica!

Por supuesto, también pudiera suceder que nuestro universo *real* no vaya a colapsar nunca más. Quizá estamos viviendo en un universo con una curvatura espacial globalmente nula (caso euclídeo) o con curvatura negativa (caso de Lobachevsky). O quizá *estamos* viviendo en un universo (con curvatura positiva) que colapsa, pero este colapso ocurrirá en un tiempo tan remoto que ninguna violación de la segunda ley sería discernible en nuestra época presente —pese al hecho de que, desde esta perspectiva, llegaría un momento en que el incremento de la entropía *global* del universo se detendría y finalmente empezaría a disminuir hasta un valor pequeñísimo— y la segunda ley, como hoy la entendemos, sería violada.

De hecho, existen muy buenas razones para dudar de que pudiera haber tal inversión de la entropía en un universo colapsante. Algunas de las más poderosas de estas razones tienen que ver con esos misteriosos objetos conocidos como *agujeros negros*. En un agujero negro tenemos un microcosmos de un universo colapsante, de modo que si la entropía se invirtiera en un universo colapsante entonces deberían darse también groseras violaciones de la segunda ley en las proximidades de un agujero negro. Sin embargo, existen todas las razones para creer que la segunda ley sigue firmemente en vigor en los agujeros negros. La teoría de los agujeros negros proporcionará un dato de partida vital para nuestra discusión de la entropía, así que será necesario que consideremos estos extraños objetos con cierto detalle.

Agujeros negros

Consideremos primero lo que nos dice la teoría sobre el destino final de nuestro Sol. El Sol lleva existiendo desde hace unos cinco mil millones de años. Dentro de otros 5 ó 6 mil millones de años empezará a expandirse en tamaño, hinchándose inexorablemente hasta que su superficie alcance la órbita de la Tierra. Entonces se habrá convertido en un tipo de estrella conocido como una *gigante roja*. Muchas gigantes rojas se pueden observar en otros lugares del cielo, siendo dos de las más conocidas Aldebarán en Tauro y Betelgeuse en Orión. Mientras su superficie esté expandiéndose, en su mismo núcleo habrá una pequeña concentración de materia excepcionalmente densa que crece continuamente. Este núcleo denso tendrá la naturaleza de una estrella *enana blanca* (Fig. 7.12).

Las estrellas enanas blancas, propiamente dichas, son auténticas estrellas cuyo material está concentrado a una densidad tan extremadamente alta que una bola de ping pong llena de este material pesaría ¡varios cientos de toneladas! Estas estrellas se observan en el cielo en número bastante considerable: quizá un diez por ciento de las estrellas de nuestra Vía Láctea sean enanas blancas. La enana blanca más famosa es la compañera de Sirio, cuya alarmantemente alta densidad supuso un gran enigma observacional para los astrónomos de principios de este siglo. Con el paso del tiempo, sin embargo, esta misma estrella proporcionó una maravillosa confirmación de una teoría física

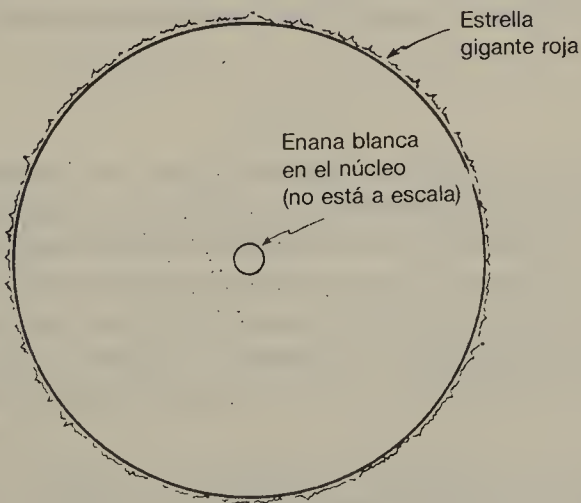


Fig. 7.12. Una estrella gigante roja con un núcleo de enana blanca.

(debida originalmente a R. H. Fowler, alrededor de 1926) según la cual algunas estrellas podrían tener tal alta densidad y se mantendrían por «presión de degeneración electrónica», lo que quiere decir que es el principio mecano-cuántico de exclusión de Pauli (p. 351), aplicado a los electrones, el que impediría que la estrella colapsara gravitatoriamente hacia dentro.

Toda gigante roja tendrá una enana blanca en su núcleo central, y este núcleo estará absorbiendo continuamente material del cuerpo principal de la estrella. Finalmente, la gigante roja será consumida completamente por este núcleo parásito y todo lo que queda es una auténtica enana blanca —de un tamaño similar al de la Tierra. Se espera que nuestro Sol existirá como gigante roja durante «sólo» unos pocos miles de millones de años. Después de eso, en su última encarnación «visible» —como un rescoldo de enana blanca que se enfría y agoniza lentamente *— el Sol persistirá durante unos pocos miles de millones de años más hasta llegar finalmente a una oscuridad total como una invisible *enana negra*.

No todas las estrellas compartirán el destino del Sol. Para algunas su sino es mucho más violento, y su destino está decidido por lo que se conoce como el *límite de Chandrasekhar*: el máximo valor posible para la masa de una estrella enana blanca. Según un cálculo realizado en 1929 por Subrahmanyan Chandrasekhar, las enanas blancas no pueden existir si sus masas son de más de aproximadamente una vez y media la masa del Sol. (Cuando hizo este cálculo, viajando en barco desde la India a Inglaterra, él era un joven indio que se preparaba para ser investigador.) El cálculo fue repetido independientemente por el ruso Lev Landau hacia 1930. El valor moderno algo refinado para el límite de Chandrasekhar es de aproximadamente

$$1,4 M_{\odot}$$

donde $M_{\odot}$ es la masa del Sol, es decir $M_{\odot} =$ una *masa solar*.

Nótese que el límite de Chandrasekhar no es mucho mayor que la masa del Sol, mientras que se conocen muchas estrellas ordinarias cuya masa es considerablemente mayor que este valor. ¿Cuál sería el destino final de una estrella de masa $2M_{\odot}$, por ejemplo? De nuevo, según la teoría establecida, la estrella se hincharía hasta hacerse una gigante roja y su núcleo de tipo enana blanca adquiriría masa lenta-

* De hecho, en sus etapas finales la enana brillará tenuemente como una estrella roja —¡aunque lo que se conoce como «enana roja» es una estrella de un carácter muy diferente!

mente igual que antes. Sin embargo, en algún momento crítico el núcleo alcanzará el límite de Chandrasekhar y el principio de exclusión de Pauli será insuficiente para mantenerla contra las enormes presiones inducidas por la gravitación ⁷. En este punto, o próximo a él, el núcleo colapsa catastróficamente hacia dentro y las temperaturas y presiones se incrementan enormemente. Se producen violentas reacciones nucleares y una enorme cantidad de energía se libera del núcleo en forma de neutrinos. Éstos calientan las regiones más externas de la estrella, que habían estado colapsando hacia dentro, y se produce una extraordinaria explosión. ¡La estrella se ha transformado en una supernova!

¿Qué sucede entonces con el núcleo que aún sigue colapsando? La teoría nos dice que alcanza densidades enormemente más grandes incluso que las densidades alarmantemente altas que se daban en el interior de una enana blanca. El núcleo puede estabilizarse como una *estrella de neutrones* (p. 402), en donde ahora es la *presión de degeneración neutrónica* —esto es, el principio de Pauli aplicado a los neutrones— la que mantiene a la estrella. La densidad será tal que la misma bola de ping pong conteniendo material de estrellas de neutrones pesaría tanto como el asteroide Hermes (o quizá como Deimos, uno de los satélites de Marte). ¡Éste es el tipo de densidad que se encuentra en el interior de los propios núcleos atómicos! (Una estrella de neutrones es como un núcleo atómico gigantesco, quizá de unos diez kilómetros de radio, que, sin embargo, ¡es extremadamente pequeño para los niveles estelares!) Pero ahora existe un *nuevo* límite, análogo al de Chandrasekhar (conocido como el límite de Landau-Oppenheimer-Volkov), cuyo valor moderno (revisado) es muy aproximadamente

$$2,5 M_{\odot}$$

por encima del cual no puede mantenerse una estrella de neutrones.

¿Qué le sucede a este núcleo colapsante si la masa de la estrella original es tan grande que incluso excede *este* límite? Por ejemplo, se conocen muchas estrellas de masas comprendidas entre $10M_{\odot}$ y $100M_{\odot}$. Parecería muy poco probable que arrojaran invariablemente tanta masa que el núcleo resultante quedara por debajo del límite para una estrella de neutrones. En lugar de ello, lo que se espera es que resulte un *agujero negro*.

¿Qué es un agujero negro? Es una región del espacio —o del espacio-tiempo— en la que el campo gravitatorio es tan intenso que ni siquiera la luz puede escapar de ella. Recordemos que una conse-

cuencia de los principios de la relatividad es que la velocidad de la luz es una velocidad límite: ningún objeto material o señal puede superar la velocidad local de la luz (pp. 251, 272). En consecuencia, si la luz no puede escapar de un agujero negro, *nada* puede escapar.

Quizá el lector esté familiarizado con el concepto de *velocidad de escape*. Ésta es la velocidad que debe alcanzar un objeto para escapar de algún cuerpo con masa. Supongamos que este cuerpo fuera la Tierra; entonces la velocidad de escape sería de aproximadamente 40.000 kilómetros por hora, algo más de 11 kilómetros por segundo. Una piedra que se lanzara desde la superficie de la Tierra (alejándose del suelo en cualquier dirección) con una velocidad que superara este valor, escaparía por completo de la Tierra (suponiendo que podamos despreciar los efectos de la resistencia del aire). Si la arrojamamos con una velocidad menor que ésta, entonces caerá de nuevo a la Tierra. (No es cierto, por lo tanto, que «todo lo que sube debe bajar»; un objeto retorna ¡sólo si es arrojado con una velocidad *menor* que la de escape!) Para Júpiter la velocidad de escape es de 220.000 kilómetros por hora; y para el Sol es de 2.200.000 kilómetros por hora. Imaginemos ahora que la masa del Sol estuviera concentrada en una esfera de sólo *un cuarto* de su radio actual, lo que supondría una velocidad de escape que es el *doble* de la actual; si el Sol estuviera más concentrado, digamos en una esfera de *una centésima parte* de su radio actual, entonces la velocidad sería *diez veces* mayor. Podemos imaginar que para un cuerpo concentrado y de suficiente masa, ¡la velocidad de escape superaría incluso a la velocidad de la luz! Cuando esto sucede, tenemos un agujero negro ⁸.

En la Figura 7.13 he dibujado un diagrama espacio-temporal que muestra el colapso de un cuerpo para formar un agujero negro (donde estoy suponiendo que el colapso tiene lugar de un modo que mantiene aproximadamente la simetría esférica, y donde he suprimido una de las dimensiones espaciales). Se han dibujado los conos de luz; como recordamos de la discusión de la relatividad general en el Capítulo 5 (cfr. p. 269), éstos indican los límites absolutos para el movimiento de un objeto material o señal. Nótese que los conos empiezan a inclinarse hacia dentro y la inclinación se hace mayor cuanto más próximos están al centro.

Hay una distancia crítica a partir del centro, llamada *radio de Schwarzschild*, en la que los límites exteriores de los conos se hacen *verticales* en el diagrama. A esta distancia, lo más que puede hacer la luz (que debe seguir a lo largo de los conos de luz) es mantenerse sobre el cuerpo colapsado: toda la velocidad hacia afuera que la luz

pueda reunir es apenas suficiente para contrarrestar la enorme atracción gravitatoria. La (3-)superficie en el espacio-tiempo que describe, en el radio de Schwarzschild, esta luz que se mantiene sobre el cuerpo (es decir, la historia entera de la luz) se conoce como el *horizonte de sucesos (absoluto)* del agujero negro. Cualquier cosa que se encuentre dentro del horizonte de sucesos no puede escapar, ni siquiera comunicar con el mundo exterior. Esto puede verse a partir de la inclinación de los conos y del hecho fundamental de que todos los movimientos y señales están obligados a propagarse dentro (o a lo largo)

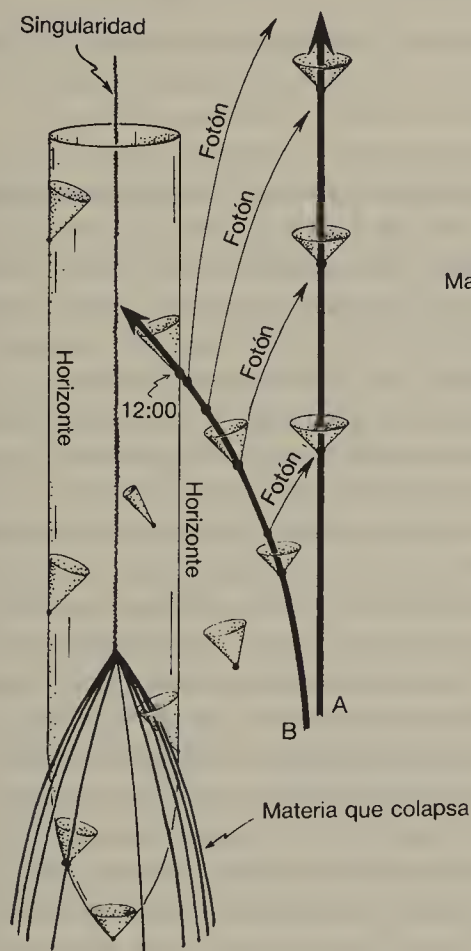


Fig. 7.13. Diagrama espacio-temporal que representa el colapso de un agujero negro. El radio de Schwarzschild figura marcado como «Horizonte».

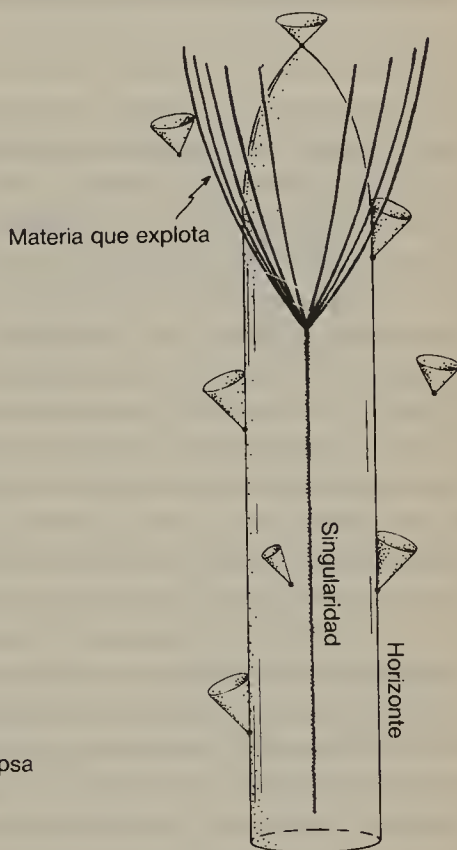


Fig. 7.4. Una hipotética configuración espacio-temporal: un agujero blanco que finalmente explota en materia (el inverso en el tiempo del diagrama de la Fig. 7.13).

de estos conos. Para un agujero negro formado por el colapso de una estrella de unas pocas masas solares el radio del horizonte será de unos pocos kilómetros. Se cree que puede haber agujeros negros mucho mayores en los centros de las galaxias. Nuestra propia Vía Láctea podría contener un agujero negro de aproximadamente un millón de masas solares, y el radio del agujero sería entonces de unos pocos millones de kilómetros.

El cuerpo material real que colapsa para formar el agujero negro acabará en su totalidad dentro del horizonte, y por lo tanto no puede comunicarse con el exterior. Consideraremos dentro de poco el destino probable de este cuerpo. Por el momento, es solamente la geometría espacio-temporal creada por su colapso lo que nos interesa: una geometría espacio-temporal con curiosas y profundas implicaciones.

Imaginemos un valiente (¿o temerario?) astronauta **B** que decide viajar al interior del agujero negro, mientras su compañero **A**, más tímido (¿o prudente?), permanece a salvo fuera del horizonte de sucesos. Supongamos que **A** se esfuerza por mantenerse a la vista de **B** el mayor tiempo posible. ¿Qué ve **A**? Puede pensarse a partir de la Figura 7.13 que la porción de la historia de **B** (esto es, la línea-de-universo de **B**) que está *en el interior* del horizonte nunca será vista por **A**, mientras que la porción *fuera* del horizonte se hará finalmente visible a **A** —aunque los momentos inmediatamente anteriores a que **B** se adentre a través del horizonte serán vistos por **A** sólo tras períodos de espera cada vez mayores. Supongamos que **B** cruza el horizonte cuando su propio reloj marca las 12 en punto. El astronauta **A** nunca podrá ser testigo de este acontecer, pero las lecturas del reloj 11:30, 11:45, 11:52, 11:56, 11:58, 11:59, 11:59½, 11:59¾, 11:59⅞, etc. serán vistas sucesivamente por **A** (a intervalos aproximadamente iguales según el punto de vista de **A**). En principio, **B** permanecerá siempre visible para **A**, y parecerá estar inmóvil para siempre justo en el borde del horizonte, con su reloj acercándose cada vez más lentamente a la hora fatídica de las 12:00 aunque sin alcanzarla nunca. Pero, en realidad, la imagen de **B** que es percibida por **A** se haría rápidamente demasiado tenue para ser discernible. Esto se debe a que la luz de la pequeñísima porción de la línea-de-universo de **B** que queda fuera del horizonte es toda la que percibirá **A** durante todo el resto del tiempo. De hecho, **B** habrá desaparecido de la vista de **A**, y lo mismo sería cierto de todo el cuerpo colapsante original. ¡Todo lo que **A** pueda ver será precisamente un «agujero negro»!

¿Qué le sucede al pobre **B**? ¿Cuál será *su* experiencia? Debe se-

ñalarse en primer lugar que no ocurrirá nada especial para **B** en el momento de su paso del horizonte. El mira su reloj cuando va a marcar las 12 en punto y ve que los minutos pasan regularmente: 11:57, 11:58, 11:59, 12:00, 12:01, 12:02, 12:03,... Nada resulta especial en torno a las 12:00. Puede volverse a mirar a **A** y encontrará que **A** permanece continuamente a la vista durante todo el tiempo. Puede mirar el propio reloj de **A**, que a **B** le parecerá estar avanzando de una forma ordenada y uniforme. A menos que **B** haya *calculado* que debe haber cruzado el horizonte, él no tendrá medio de saberlo ⁹. El horizonte ha sido extremadamente perverso. Una vez cruzado ya no hay escape para **B**. ¡Su universo local colapsará finalmente sobre él, y él estará destinado a encontrar rápidamente su propio «big crunch» privado!

O quizá no sea tan privado. Toda la materia del cuerpo colapsado, que formó el agujero negro en primer lugar, estará compartiendo, en cierto sentido, el «mismo crujido» con él. De hecho, si el universo *fuera* del agujero está cerrado espacialmente, de modo que la materia exterior acabará también por ser engullida en un big-crunch que englobe todo, entonces también sería de esperar que dicho «crujido» fuera el «mismo» que el «crujido privado» de **B** ^{*}.

A pesar del destino poco agradable de **B**, no esperamos que la física local que él experimenta hasta ese punto pudiera estar reñida con la física que hemos llegado a conocer y comprender. En particular, no esperamos que él experimente violaciones locales de la segunda ley de la termodinámica, ni mucho menos una completa inversión del comportamiento creciente de la entropía. La segunda ley seguirá siendo tan válida en el interior del agujero negro como lo es en cualquier otra parte. La entropía en la vecindad de **B** sigue en continuo incremento hasta el mismo instante de su crujido final.

Para entender cómo la entropía en un «big crunch» (ya sea «privado» u «omni-comprensivo») puede ser enormemente alta, mientras que la entropía en el big bang tuvo que haber sido mucho más baja, necesitaremos ahondar todavía un poco más en la geometría espacio-temporal de un agujero negro. Pero antes de hacerlo, el lector debe echar una ojeada también a la Figura 7.14 que muestra el hipotético

^{*} Al hacer esta afirmación estoy aceptando dos hipótesis. La primera es que la posible desaparición final del agujero negro —debida a su (extraordinariamente lenta) «evaporación» por radiación de Hawking que consideramos más adelante (cf. p. 428)— quedaría impedida por el colapso del Universo; la segunda es una hipótesis (muy plausible) conocida como «censura cósmica» (p. 278).

inverso temporal de un agujero negro: un agujero *blanco*. Los agujeros blancos probablemente *no* existan en la Naturaleza, pero su posibilidad teórica tendrá considerable importancia para nosotros.

La estructura de las singularidades del espacio-tiempo

Recordemos del Capítulo 5 (p. 264) que la curvatura del espacio-tiempo se manifiesta como un *efecto de marea*. Una superficie esférica, constituida por partículas en caída libre en el campo gravitatorio de algún cuerpo grande, será estirada en una dirección (a lo largo de la línea dirigida hacia el cuerpo) y aplastada en direcciones perpendiculares a aquélla. Esta distorsión de marea aumenta a medida que se acerca al cuerpo gravitatorio (Fig. 7.15), variando de forma inversamente proporcional al cubo de la distancia. Un semejante efecto de marea en aumento será sentido por el astronauta **B** a medida que cae hacia el agujero negro. Para un agujero negro de unas pocas masas solares este efecto de marea sería enorme —demasiado enorme para que el astronauta pudiera sobrevivir a cualquier aproximación al agujero, y mucho menos al cruce del horizonte. Para agujeros más grandes el tamaño del efecto de marea en el horizonte sería en realidad más pequeño. Para el agujero negro de millones de masas solares que muchos astrónomos creen que puede haber en el centro de la Vía Láctea, el efecto de marea sería suficientemente pequeño para que el astronauta cruzara el horizonte, aunque lo bastante grande para que se sintiera algo incómodo. Si embargo, este efecto de marea no se mantiene pequeño durante toda la caída del astronauta sino que crece rápidamente a infinito ¡en cuestión de pocos segundos! No sólo el desgraciado astronauta sería despedazado por esta fuerza de marea rápidamente creciente sino que también lo serían, en rápida sucesión, las moléculas de las que estaba compuesto, sus átomos constituyentes,



Fig. 7.15. El efecto de marea debido a un cuerpo esférico gravitante aumenta, a medida que el cuerpo se acerca, como el inverso del cubo de la distancia al centro del cuerpo.

sus núcleos e incluso, finalmente, ¡todas las partículas subatómicas! Así es como el «crujido» lleva a cabo su último estrago.

No sólo toda la materia se destruye de este modo sino que ¡incluso el propio espacio-tiempo debe encontrar su final! Semejante catástrofe final se conoce como una *singularidad del espacio-tiempo*. El lector puede preguntarse cómo sabemos que deben ocurrir tales singularidades, y bajo qué circunstancias la materia y el espacio-tiempo están destinados a sufrir este destino. Éstas son consecuencias que se deducen de las ecuaciones clásicas de la relatividad general, en cualquier circunstancia en que se forme un agujero negro. El modelo original de agujero negro de Oppenheimer y Snyder (1939) mostraba un comportamiento de este tipo. Sin embargo, los astrofísicos han mantenido durante muchos años la esperanza de que este comportamiento singular fuera un artificio de las simetrías espaciales que tenían que suponerse en el modelo. Quizá en situaciones realistas (asimétricas) la materia colapsante podría enroscarse de alguna forma complicada y luego escapar de nuevo hacia afuera. Pero tales esperanzas se fueron por tierra cuando se pudo disponer de argumentos matemáticos de tipo más general que proporcionan lo que se conocen como *teoremas de singularidad* (cf. Penrose, 1965, Hawking y Penrose, 1970). Estos teoremas establecieron, dentro de la teoría clásica de la relatividad con fuentes materiales razonables, que las singularidades en el espacio-tiempo son *inevitables* en situaciones de colapso gravitatorio.

Análogamente, utilizando la dirección inversa del tiempo resulta de nuevo inevitable llegar a una correspondiente singularidad *inicial* en el espacio-tiempo, que representa ahora al big bang, en cualquier universo en (apropiada) expansión. Aquí, en lugar de representar la *destrucción* final de toda la materia del espacio-tiempo, la singularidad representa la *creación* de espacio-tiempo y materia. Podría parecer que hay una simetría temporal exacta entre estos dos tipos de singularidad: el tipo *inicial*, en el que se crean el espacio-tiempo y la materia; y el tipo *final*, en el que se destruyen el espacio-tiempo y la materia. Existe, realmente, una importante analogía entre estas dos situaciones, pero cuando las examinamos en detalle encontramos que *no* son exactamente inversas en el tiempo. ¡Es importante que comprendamos las diferencias geométricas puesto que ellas contienen la clave del origen de la segunda ley de la termodinámica!

Volvamos a las experiencias de nuestro auto-sacrificado astronauta. Él encuentra que las fuerzas de marea crecen rápidamente hacia infinito. Puesto que está viajando en un espacio vacío, él experimenta los efectos de conservación de volumen, aunque con *distorsión*, que

son debidos al tipo de tensor de curvatura espacio-temporal que he llamado **WEYL** (ver Capítulo 5, pp. 264, 271). La parte restante del tensor de curvatura espacio-temporal, la parte que representa una compresión global y a la que llamamos **RICCI**, es nula en el espacio vacío. Pudiera ser que, en efecto, **B** encontrara materia en alguna etapa, pero incluso si así fuera (y él mismo está constituido, después de todo, de materia), seguiríamos encontrando en general que la medida de **WEYL** es mucho *mayor* que la de **RICCI**. En realidad, esperamos encontrar que la curvatura cerca de una singularidad *final* esté dominada completamente por el tensor **WEYL**. Este tensor tiende a *infinito*, en general:

$$\text{WEYL} \rightarrow \infty$$

(aunque podría hacerlo perfectamente de manera oscilatoria). Ésta parece ser la situación *genérica* con una singularidad en el espacio-tiempo ¹⁰. Tal comportamiento está asociado con una singularidad de *alta entropía*.

Sin embargo la situación con el big bang parece ser bastante diferente. Los modelos estándar del big bang se derivan de los espacio-tiempos altamente simétricos de Friedmann-Robertson-Walker que consideramos anteriormente. Ahora el efecto de marea distorsionante que proporciona el tensor **WEYL** está totalmente *ausente*. En su lugar hay una aceleración simétrica hacia adentro actuando en cualquier superficie esférica de partículas de prueba (ver Fig. 5.26). Ésta es un efecto del tensor **RICCI** más que del **WEYL**. En cualquier modelo FRW siempre se satisface la ecuación tensorial

$$\text{WEYL}=0$$

A medida que nos aproximamos cada vez más a la singularidad inicial encontramos que es **RICCI** el que se hace infinito, en lugar de **WEYL**, de modo que es **RICCI** el que domina cerca de la singularidad inicial, por encima de **WEYL**. Esto nos proporciona una singularidad de *baja entropía*.

Si examinamos la singularidad del *big crunch* en los modelos FRW *exactamente* recolapsantes encontraremos ahora **WEYL** = 0 en el momento final del «crujido» mientras que **RICCI** tiende a infinito. Sin embargo, ésta es una situación muy especial y *no* es lo que esperamos para un modelo completamente realista en el que se tenga en cuenta el agrupamiento gravitatorio. A medida que avanza el tiempo, el material, originalmente en forma de gas difuso, se agrupará en galaxias de estrellas. A su debido tiempo un gran número de estas estrellas se

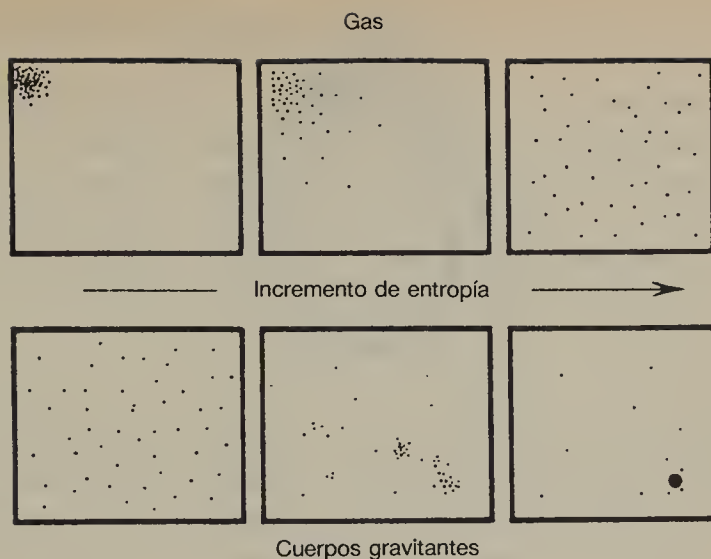


Fig. 7.16. Para un gas ordinario, el incremento de la entropía tiende a hacer más uniforme la distribución. Para un sistema de cuerpos gravitantes sucede lo contrario. La entropía alta se logra mediante agrupamiento gravitatorio —y la más alta de todas, mediante el colapso de un agujero negro.

contraerán gravitatoriamente para dar enanas blancas, estrellas de neutrones y agujeros negros, y podrían darse perfectamente enormes agujeros negros en los centros de las galaxias. El agrupamiento —particularmente en el caso de los agujeros negros— representa un enorme incremento de entropía (ver Fig. 7.16). Puede parecer enigmático, al principio, que los estados agrupados representen *alta* entropía y los más uniformes *baja* entropía, cuando recordamos que, en el caso del gas en una caja, los estados agrupados (como el del gas cuando estaba en una esquina de la caja) eran de *baja* entropía, mientras que en el estado *uniforme* de equilibrio térmico la entropía era *alta*. Cuando se tiene en cuenta la gravedad se produce una *inversión* de esto debido a la naturaleza universalmente atractiva del campo gravitatorio. El agrupamiento se hace cada vez más extremo conforme pasa el tiempo y, finalmente, se coagulan muchos agujeros negros y sus singularidades se unen en la muy compleja singularidad final del big crunch. La singularidad final no se parece en modo alguno al big crunch idealizado del modelo FRW recolapsante, con su ligadura $W_{EYL} = 0$. A medida que se produce un agrupamiento cada vez mayor, existe, durante todo el tiempo, una tendencia del tensor Weyl a crecer cada vez más ¹¹ y, en general, $W_{EYL} \rightarrow \infty$ en cualquier singularidad final.



Fig. 7.17. La historia completa de un universo cerrado que comienza en un big bang uniforme de baja entropía, con $\text{WEYL} = 0$, y termina en un big crunch —que representa la coagulación de muchos agujeros negros— con $\text{WEYL} \rightarrow \infty$.

Véase la figura 7.17 para una imagen espacio-temporal que representa toda la historia de un universo cerrado de acuerdo con esta descripción general.

Vemos ahora cómo es posible que un universo recolapsado *no* necesite tener una pequeña entropía. La «pequeñez» de la entropía en el big bang —que nos dio la segunda ley— *no* era, por lo tanto, una mera consecuencia de la «pequeñez» del Universo en el tiempo del big bang. Si invirtiéramos en el tiempo la imagen del big crunch que obtuvimos antes, obtendríamos un «big bang» con una entropía enormemente *alta*, ¡y no habría existido segunda ley! Por alguna razón, el universo fue creado en un estado muy especial (de baja entropía), en el que se impuso algo parecido a la ligadura $\text{WEYL} = 0$ de los modelos FRW. Si no fuera por una ligadura de esta naturaleza sería «mucho más probable» tener una situación en la que *ambas* singularidades inicial y final fueran del tipo $\text{WEYL} \rightarrow \infty$ de alta entropía (ver Fig. 7.18). ¡En un universo «probable» semejante *no* existiría, de hecho, segunda ley de la termodinámica!

¿Hasta qué punto fue especial el big bang?

Tratemos de comprender hasta qué punto era restrictiva para el big bang una condición como $WEIL = 0$. Por simplicidad (como en la discusión anterior) supongamos que el universo es cerrado. Para poder obtener algunas cifras redondas supondremos, además, que el número B de *bariones* —es decir el número de protones y neutrones juntos— en el universo viene dado aproximadamente por

$$B = 10^{80}$$

(No existe ninguna razón especial para esta cifra, aparte del hecho de que observacionalmente B debe ser de *al menos* ese orden; Eddington afirmó en cierta ocasión haber calculado B *exactamente*, ¡obteniendo una cifra que estaba próxima al valor anterior! Nadie más parece creer este cálculo particular, pero parece que se ha retenido el valor 10^{80} .) Si se tomara B *mayor* que este valor (y quizá en realidad tengamos $B = \infty$) ¡entonces las cifras que obtendríamos serían aún *más* sorprendentes que las extraordinarias cifras a que vamos a llegar en un minuto!



Fig. 7.18. Si se levanta la ligadura $WEYL = 0$ entonces tenemos también un big bang con alta entropía, en el que $WEYL \rightarrow \infty$. Un universo semejante estaría plagado de agujeros blancos y no habría segunda ley de la termodinámica, en abierta contradicción con la experiencia.

¡Tratemos de imaginar el espacio de fases (cf. p. 230) del universo *entero*! Cada punto en este espacio de fases representa una diferente forma posible en que pudiera haber comenzado el Universo. Vamos a representar al Creador provisto de una «aguja» que será colocada en algún punto del espacio de fases (Fig. 7.19). Cada colocación diferente de la aguja proporciona un universo diferente. Ahora bien, la precisión necesaria para la puntería del Creador depende de la entropía del universo que se vaya a crear. Sería relativamente «fácil» producir un universo de alta entropía, puesto que para ello habría un gran volumen del espacio de fases disponible para colocar la aguja. (Recordemos que la entropía es proporcional al logaritmo del volumen que interesa en el espacio de fases.) Pero para iniciar el Universo en un estado de baja entropía —de modo que realmente exista una segunda ley de la termodinámica— el Creador debe apuntar a un volumen muchísimo más pequeño del espacio de fases. ¿Cómo de pequeña debería ser esta región para que el resultado fuera un universo que se pareciera estrechamente al universo en que vivimos? Para responder a esta respuesta debemos ir antes a una fórmula muy importante, debida a Jacob Bekenstein (1972) y Stephen Hawking (1975), que nos dice cuál debe ser la entropía de un agujero negro.



Fig. 7.19. Para producir un universo parecido al que habitamos, el Creador tendría que haber apuntado a un volumen absurdamente minúsculo del espacio de fases de los universos posibles —aproximadamente $1/10^{10^{123}}$ del volumen total, para la situación considerada. (¡La aguja, y el punto apuntado, no están dibujados a escala!)

Consideremos un agujero negro, y supongamos que el área de su horizonte es A . La fórmula de Bekenstein-Hawking para la entropía del agujero negro es entonces:

$$S_{an} = \frac{A}{4} \times \left(\frac{kc^3}{G\hbar} \right),$$

donde k es la constante de Boltzmann, c es la velocidad de la luz, G es la constante gravitatoria de Newton, y $\hbar$ es la constante de Planck dividida por 2π . La parte esencial de esta fórmula es el factor $A/4$. La parte entre paréntesis consta simplemente de las constantes físicas apropiadas. En definitiva, la entropía del agujero negro es proporcional al área de su superficie. Para un agujero negro con simetría esférica, este área resulta ser proporcional al cuadrado de la masa del agujero.

$$A = m^2 \times 8\pi(G^2/c^4).$$

Introduciendo esto en la fórmula de Bekenstein-Hawking encontramos que la entropía de un agujero negro es proporcional al cuadrado de su masa:

$$S_{an} = m^2 \times 2\pi(kG/\hbar c).$$

Así, la *entropía por unidad de masa* (S_{an}/m) de un agujero negro es proporcional a su masa y, por lo tanto, es cada vez mayor para agujeros negros cada vez más grandes. En consecuencia, para una cantidad dada de masa (o equivalentemente, por la relación $E = mc^2$ de Einstein, para una cantidad dada de *energía*) se alcanza la mayor entropía ¡cuando todo el material ha colapsado completamente en un agujero negro! Más aún, ¡dos agujeros negros ganan (enormemente) en entropía cuando se engullen mutuamente para dar lugar a un único agujero negro unido! Los agujeros negros grandes, como los que probablemente se encuentren en los centros de las galaxias, proporcionarán cantidades de entropía absolutamente enormes, muchísimo mayores que los otros tipos de entropía que encontramos en los otros tipos de situación física.

En realidad es necesario hacer una ligera puntualización a la afirmación de que la mayor entropía se logra cuando toda la masa está concentrada en un agujero negro. El análisis de Hawking de la termodinámica de los agujeros negros muestra que existirá también una *temperatura* no nula asociada a un agujero negro. Una consecuencia de esto es que no toda la energía puede estar contenida dentro del

agujero negro en el estado de máxima entropía, pues la máxima entropía se alcanza para un agujero negro en equilibrio con un «baño térmico de radiación». La temperatura de esta radiación es realmente ínfima para un agujero negro de tamaño razonable. Por ejemplo, para un agujero negro de una masa solar esta temperatura sería de unos 10^{-7}K , que es algo más pequeña que la temperatura más baja que se ha medido en un laboratorio hasta la fecha, y muchísimo menor que la temperatura de 2,7 K del espacio intergaláctico. Para agujeros negros mayores la temperatura de Hawking es ¡aún más pequeña!

La temperatura de Hawking se haría importante para nuestra discusión sólo en uno de estos dos casos: (i) podrían existir en nuestro universo agujeros negros muchísimos más pequeños, conocidos como *mini-agujeros negros*; (ii) el universo no colapsa antes del *tiempo de evaporación de Hawking* —el tiempo que tardaría el agujero negro en evaporarse completamente. Con respecto a (i), los mini-agujeros negros podrían producirse únicamente en un big bang adecuadamente caótico. Tales mini-agujeros no pueden ser muy numerosos en nuestro universo real, o de otra forma sus efectos ya se hubieran observado; más aún, de acuerdo con el punto de vista que estoy exponiendo aquí estos mini-agujeros deberían estar totalmente ausentes. Con respecto a (ii), para un agujero negro de masa solar el tiempo de evaporación de Hawking sería de unas 10^{54} veces la edad actual del Universo, y para agujeros negros más grandes sería considerablemente más largo. No parece que estos efectos pudieran modificar substancialmente los argumentos anteriores.

Para hacernos alguna idea de la enormidad de la entropía de los agujeros negros consideremos lo que previamente pensábamos que suministraba la mayor contribución a la entropía del Universo: la radiación de fondo de cuerpo negro de 2,7 K. Los astrofísicos habían quedado sorprendidos por la enorme cantidad de entropía que contiene esta radiación, que está muy por encima de las cifras normales de entropía que encontramos en otros procesos (v.g. en el Sol). La entropía de la radiación de fondo es del orden de 10^8 por cada barión (en donde estoy escogiendo «unidades naturales» de modo que la constante de Boltzmann es la unidad). (Esto significa, de hecho, que hay 10^8 fotones en la radiación de fondo por cada barión.) Así, con 10^{80} bariones en total tendremos una entropía de

$$10^{88}$$

para la entropía de la radiación de fondo en el universo.

En realidad, si no fuera por los agujeros negros esta cifra repre-

sentaría la entropía *total* del Universo, puesto que la entropía de la radiación de fondo supera con mucho la de cualesquiera otros procesos ordinarios. La entropía por barión en el Sol, por ejemplo, es del orden de la unidad. Por el contrario, para los niveles de los *agujeros negros*, la entropía de la radiación de fondo es el «chocolate del loro». En efecto, la fórmula de Bekenstein-Hawking nos dice que la entropía por barión en un agujero negro de masa solar es del orden de 10^{20} , en unidades naturales, de modo que si el universo hubiera consistido solamente en agujeros negros de masa solar, la cifra global habría sido mucho mayor que la dada más arriba, a saber

$$10^{100}$$

Por supuesto, el universo no está constituido de esta forma pero esta cifra empieza a decirnos qué «pequeña» debe ser considerada la entropía de la radiación de fondo cuando los efectos implacables de la gravedad empiezan a ser tenidos en cuenta.

Tratemos de ser un poco más realistas. En lugar de poblar completamente las galaxias con agujeros negros, consideremos que constan principalmente de estrellas ordinarias —unas 10^{11} de ellas— y que cada galaxia tiene en su núcleo un agujero negro de un millón (esto es, 10^6) de masas solares (como podría ser razonable para nuestra propia Vía Láctea). Los cálculos muestran que la entropía por barión será entonces algo mayor incluso que la enorme cifra anterior, a saber 10^{21} , dando una entropía total, en unidades naturales, de

$$10^{101}$$

Podemos prever que, después de un tiempo largo, una fracción importante de las masas de las galaxias estarán incorporadas en los agujeros negros de sus centros. Cuando esto sucede la entropía por barión será 10^{31} , dando un total monstruoso de

$$10^{111}$$

Sin embargo, estamos considerando un universo cerrado que finalmente colapsa; y no es irrazonable estimar la entropía del «crujido» final utilizando la fórmula de Bekenstein-Hawking como si todo el universo hubiera formado un agujero negro. Esto da una entropía por barión de 10^{43} , y el total, absolutamente inimaginable, para todo el big crunch sería

$$10^{123}$$

Esta cifra nos dará una estimación del volumen total V del espacio de fases disponible al Creador, ya que esta entropía representaría el logaritmo del volumen del (con mucho) mayor compartimento. Puesto que 10^{123} es el *logaritmo* del volumen, este volumen debe ser la *exponencial* de 10^{123} , es decir

$$V = 10^{10^{123}}$$

¡en unidades naturales! (Algunos lectores perspicaces pueden pensar que debería haber utilizado la cifra $e^{10^{123}}$, pero para números de esta magnitud, el e y el 10 son prácticamente intercambiables) ¿Qué tamaño tenía el volumen W del espacio de fases original al que tuvo que apuntar el Creador para proporcionar un universo compatible con la segunda ley de la termodinámica y con el que ahora observamos? No importa mucho si tomamos el valor

$$W = 10^{10^{101}} \text{ o } W = 10^{10^{88}}$$

dado por los agujeros negros galácticos o por la radiación de fondo, respectivamente, o una cifra mucho más pequeña (y, de hecho, más adecuada) que hubiera sido la cifra *real* en el big bang. En cualquier caso, la razón de V a W será, aproximadamente

$$V/W = 10^{10^{123}}$$

(Inténtelo: $10^{10^{123}} \div 10^{10^{101}} = 10^{(10^{123}-10^{101})} = 10^{10^{123}}$ muy aproximadamente.)

Esto nos dice lo precisa que debía haber sido la puntería del Creador: una precisión de

$$\text{una parte en } 10^{10^{123}}$$

Ésta es una cifra extraordinaria. Ni siquiera podríamos *escribir el número* completo en la notación decimal ordinaria: sería un «1» ¡seguido de 10^{123} «0»s! Incluso si escribiéramos un «0» en cada protón y en cada neutrón del universo entero —y añadiéramos también todas las demás partículas— nos quedaríamos muy cortos para escribir la cifra requerida. Se puede ver que la precisión necesaria para poner al universo en su curso no es en modo alguno inferior a la extraordinaria precisión a la que ya nos habíamos acostumbrado en las ecuaciones dinámicas soberbias (las de Newton, las de Maxwell, las de Einstein) que gobiernan el comportamiento de las cosas en cada instante.

Pero, ¿por qué estaba el big bang tan exactamente organizado,

mientras que el big crunch (o las singularidades de los agujeros negros) sería totalmente caótico? Parecería que esta pregunta pueda ser enunciada en términos del comportamiento de la parte WEYL del tensor de curvatura espacio-temporal en las singularidades del espacio-tiempo. Lo que parece que encontramos es que existe una ligadura

$$\text{WEYL} = 0$$

(o algo muy parecido a esto) en las singularidades *iniciales* del espacio-tiempo —pero no en las singularidades finales— y esto parece ser lo que confina la elección del Creador a esta región minúscula del espacio de fases. Yo he llamado *hipótesis de curvatura de Weyl* a la hipótesis de que esta ligadura se aplica a cualquier singularidad inicial (pero no final) del espacio-tiempo. Parecería así que necesitamos entender por qué debería aplicarse esta hipótesis temporalmente asimétrica si tuviéramos que comprender de dónde procede la segunda ley ¹².

¿Cómo podemos obtener una comprensión mayor del origen de la segunda ley? Parece que hemos llegado a un punto muerto. Necesitamos comprender *por qué las singularidades del espacio-tiempo* tienen las estructuras que parecen tener; pero las singularidades del espacio-tiempo son regiones en las que nuestra comprensión de la física ha llegado al límite. El punto muerto que supone la existencia de singularidades en el espacio-tiempo se compara a veces con otro punto muerto: el que encontraron los físicos a principios de siglo a propósito de la estabilidad de los átomos (cf. p. 290). En cada caso, la teoría clásica bien establecida había llegado a la respuesta «infinito», y por consiguiente se había mostrado inadecuada para la tarea. El comportamiento singular del colapso electromagnético de los átomos era predicho por la teoría *cuántica*; y análogamente debería ser la teoría cuántica la que diera una teoría finita en lugar de las singularidades «infinitas» clásicas en el espacio-tiempo para el colapso gravitatorio de las estrellas. Pero no puede ser una teoría cuántica ordinaria. Debe ser una teoría cuántica de la propia estructura del espacio-tiempo. Esta teoría, si existiera, sería conocida como «*gravitación cuántica*». El hecho de que no exista una gravitación cuántica no se debe a falta de esfuerzo, capacidad o ingenio por parte de los físicos. Muchos cerebros científicos de primera fila se han dedicado a la construcción de una teoría semejante, aunque sin éxito. Éste es el punto muerto al que hemos llegado finalmente en nuestros intentos por comprender la direccionalidad y el flujo del tiempo.

El lector puede preguntarse qué es lo que hemos sacado en limpio

de nuestro viaje. En nuestra búsqueda de comprensión de por qué el tiempo parece fluir en sólo una dirección, y no en la otra, hemos tenido que viajar a los mismos confines del tiempo, donde las nociones mismas del espacio se disuelven. ¿Qué hemos aprendido de todo esto? Hemos aprendido que nuestras teorías no son todavía adecuadas para proporcionar respuestas, pero ¿de qué nos ha servido en nuestro intento de comprender la mente? Pese a la falta de una teoría adecuada, creo que hay importantes lecciones que aprender de nuestro viaje. Ahora debemos encaminarnos de vuelta a casa. Nuestro viaje de regreso será más especulativo de lo que fue el de la ida pero, en mi opinión, ¡no existe otro camino de vuelta posible!

8. En busca de la gravitación cuántica

¿Por qué la gravitación cuántica?

¿Qué cosas nuevas acerca de cerebros o mentes podemos aprender de lo que hemos visto en el último capítulo? Aunque podamos haber captado algunos de los principios físicos omnicomprensivos que subyacen a la direccionalidad de nuestro «flujo del tiempo» percibido, parece que hasta ahora no hemos sacado ninguna idea sobre la cuestión de por qué percibimos que el tiempo fluye o, en realidad, por qué percibimos siquiera cualquier cosa. En mi opinión se necesitan ideas mucho más radicales. Hasta ahora mi presentación no ha sido especialmente radical, aunque en ocasiones he acentuado aspectos diferentes de los usuales. Hemos tenido conocimiento de la segunda ley de la termodinámica y he intentado persuadir al lector de que el origen de esta ley —que nos presenta la Naturaleza en la forma particular en que ha sido escogida— puede rastrearse hasta una enorme ligadura geométrica en el big bang origen del universo: *la hipótesis de curvatura de Weyl*. Algunos cosmólogos preferirán caracterizar esta condición inicial de forma algo diferente, pero se necesita realmente una restricción semejante en la singularidad inicial. Las *deducciones* que voy a obtener de esta hipótesis serán mucho menos convencionales que la propia hipótesis. ¡Afirmando que necesitamos un cambio en el propio marco de la teoría cuántica!

Este cambio jugará su papel cuando la mecánica cuántica se unifique apropiadamente con la relatividad general, es decir, en la tan deseada teoría de la *gravitación cuántica*. La mayoría de los físicos no cree que la teoría cuántica tenga que cambiar cuando se unifique con la relatividad general. Además, estos mismos físicos mantendrán que,

en la escala relevante para nuestros cerebros, los efectos de *cualquier* gravitación cuántica ¡deben ser completamnete insignificantes! Dirán (muy razonablemente) que aunque tales efectos físicos podrían ser importantes en la absurdamente minúscula escala de la distancia conocida como *longitud de Planck* * —que es de 10^{-35} m, unas 100 000 000 000 000 000 000 veces más pequeña que el tamaño de la más pequeña partícula subatómica— estos efectos no tendrían ninguna relevancia directa para fenómenos en las escalas muchísimo más grandes de, pongamos por caso, hasta sólo 10^{-12} m, donde tienen lugar los procesos químicos o eléctricos que son importantes para la actividad cerebral. De hecho, ni siquiera la gravitación *clásica* (esto es, no-cuántica) tiene apenas importancia en estas actividades eléctricas y químicas. Si la gravitación clásica no tiene consecuencias, entonces ¿cómo demonios podría suponer la más mínima diferencia cualquier minúscula «corrección cuántica» a la teoría clásica? Además, puesto que *nunca* se han observado *desviaciones* de la teoría cuántica, ¡parecería todavía menos razonable el imaginar que cualquier supuesta minúscula desviación de la teoría cuántica estándar pudiera tener ningún papel concebible que jugar en los fenómenos mentales!

Yo voy a argumentar de forma muy diferente. De hecho, no estoy interesado tanto en los efectos que la mecánica cuántica pudiera tener en nuestra teoría (la relatividad general de Einstein) de la estructura del espacio-tiempo, como en el caso *inverso*, a saber: los efectos que la teoría del espacio-tiempo de Einstein pudiera tener sobre la propia estructura de la mecánica cuántica. Recalcaré que es un punto de vista *no convencional* el que voy a desarrollar. ¡No es convencional el que la relatividad general pudiera tener alguna influencia sobre la estructura de la mecánica cuántica! Los físicos convencionales han sido muy reacios a creer que la estructura estándar de la mecánica cuántica debiera ser alterada de alguna forma. Aunque es cierto que la aplicación directa de las reglas de la teoría cuántica a la teoría de Einstein ha encontrado dificultades aparentemente insuperables, la reacción de quienes trabajan en este campo ha tendido a utilizar esto como una razón para modificar la teoría *de Einstein*, no la teoría cuántica ¹. Mi punto de vista es prácticamente el opuesto. Creo que los problemas internos a la propia teoría cuántica son de un carácter fundamental.

* Esta es la distancia ($10^{-35}\text{m} = \sqrt{(\hbar G c^{-3})}$) para las que las llamadas «fluctuaciones cuánticas» de la propia métrica del espacio-tiempo serían tan grandes que la idea normal de un continuo espacio-temporal suave deja de aplicarse. (Las fluctuaciones cuánticas son una consecuencia del principio de incertidumbre de Heisenberg cf. pp. 314-315.)

Recordemos la incompatibilidad entre los dos procedimientos básicos **U** y **R** de la mecánica cuántica (**U** obedece la completamente determinista *ecuación de Schrödinger* —llamada evolución *unitaria*— y **R** era la probabilista *reducción del vector de estado* que debemos aplicar siempre que se estime que se ha hecho alguna «observación»). En mi opinión, esta incompatibilidad es algo que *no puede* resolverse adecuadamente mediante la simple adopción de una «interpretación» apropiada de la mecánica cuántica (aunque parece que es opinión común que algo debe poder hacer), sino sólo mediante alguna teoría radicalmente nueva, según la cual los dos procedimientos **U** y **R** se verán como dos aproximaciones diferentes (y excelentes) a un procedimiento *único* más general y exacto. Mi opinión, en consecuencia, es que incluso la maravillosamente exacta teoría de la mecánica cuántica tendrá que cambiarse, y que poderosos indicios sobre la naturaleza de este cambio tendrán que venir de la relatividad general de Einstein. Iré todavía más lejos y diré que es realmente la deseada teoría de la *gravitación cuántica* la que debe contener, como uno de sus ingredientes fundamentales, este supuesto proceso combinado **U/R**.

En el punto de vista *convencional*, por el contrario, cualquier implicación directa de la gravitación cuántica sería de una naturaleza más esotérica. He mencionado las expectativas de una alteración fundamental de la estructura del espacio-tiempo en la dimensión ridículamente pequeña de la longitud de Planck. Existe también la creencia (justificada, en mi opinión) de que la gravitación cuántica debería estar involucrada de modo fundamental en la determinación en última instancia de la naturaleza del zoo de las «partículas elementales» actualmente observadas. Por el momento no existe, por ejemplo, ninguna buena teoría que explique por qué las masas de las partículas debieran ser las que son —cuando la «masa» es un concepto íntimamente ligado con el concepto de gravitación. (De hecho, la masa actúa unívocamente como la «fuente» de la gravitación.) Existen también buenas expectativas de que (según una idea desarrollada alrededor de 1955 por el físico sueco Oskar Klein) la teoría correcta de la gravitación cuántica sirviera para eliminar los infinitos que plagan la teoría cuántica de campos convencional (cf. p. 365). La física es una unidad, y la *verdadera* teoría de la gravitación cuántica, cuando finalmente llegue, debe constituir ciertamente una parte profunda de nuestra comprensión detallada de las leyes universales de la Naturaleza.

Estamos, sin embargo, lejos de tal comprensión. Además, cualquier supuesta teoría de la gravitación cuántica debería ciertamente estar muy alejada de los fenómenos que gobiernan el comportamiento

de los cerebros. *Especialmente* alejado de la actividad cerebral parecería estar ese papel (generalmente aceptado) de la gravitación cuántica que se necesita para resolver el punto muerto al que llegamos en el último capítulo: el problema de las *singularidades del espacio-tiempo* —las singularidades de la teoría clásica de Einstein que aparecen en el big bang y en los *agujeros negros*— y también en el *big crunch*, si nuestro universo decide finalmente colapsar. Sí, este papel podría *parecer* muy lejano. Argumentaré, sin embargo, que existe un hilo de conexión lógica sutil pero importante. Tratemos de ver cuál es esta conexión.

¿Qué hay detrás de la hipótesis de curvatura de Weyl?

Como ya he señalado, incluso el punto de vista convencional nos dice que debería ser la gravitación cuántica la que viniera en ayuda de la teoría clásica de la relatividad general y resolver el misterio de las singularidades del espacio-tiempo. Así, la gravitación cuántica tiene que proporcionarnos alguna física coherente en lugar de la absurda respuesta «infinito» que nos da la teoría clásica. Ciertamente coincido en esta opinión: éste es realmente un lugar en donde la gravitación cuántica debe dejar clara su huella. Sin embargo, los teóricos no parecen haberse puesto de acuerdo en el hecho sorprendente de que la huella de la gravitación cuántica es ¡descaradamente tiempo-asimétrica! En el big bang —la *singularidad del pasado*— la gravitación cuántica debe decirnos que debe darse una condición similar a

$$\text{WEYL} = 0$$

en el momento que se haga significativo el hablar en términos de conceptos clásicos de geometría del espacio-tiempo. Por otro lado, en las singularidades en el interior de los agujeros negros y en el (posible) big crunch —*singularidades futuras*— no existe tal restricción, y esperamos que el tensor de Weyl se haga infinito:

$$\text{WEYL} \rightarrow \infty$$

a medida que nos acercamos a la singularidad. En mi opinión, ésta es una indicación clara de que la teoría real que buscamos debe ser asimétrica respecto al tiempo:

*nuestra deseada gravitación cuántica debe ser una teoría
tiempo-asimétrica*

Se advierte aquí al lector que esta conclusión, pese a su aparente necesidad obvia según el modo que estoy presentando las cosas, ¡no es un saber aceptado! La mayoría de los que trabajan en este campo parecen muy reacios a aceptar esta idea. La razón parece ser el hecho de que no hay una manera clara mediante la que los procedimientos convencionales y bien establecidos de cuantización (hasta donde puedan llegar) pudieran producir una teoría cuantizada con asimetría temporal, cuando la propia teoría clásica a la que se aplican estos procedimientos (la relatividad general estándar o una de sus modificaciones más conocidas) tiene simetría temporal. En consecuencia (cuando ellos consideran estos temas —¡lo que no es muy frecuente!), estos cuantizadores de la gravitación necesitarán tratar de buscar en otra parte la «explicación» del bajo valor de la entropía en el big bang.

Quizá muchos físicos sostendrían que una hipótesis tal como la de una curvatura de Weyl inicialmente nula, al ser una elección de «condición de contorno» y no una ley dinámica, no es algo cuya explicación esté dentro de los poderes de la física. En realidad, están alegando que estamos ante un «acto divino», y no somos nosotros nadie para intentar comprender por qué nos ha sido dada una condición de contorno antes que otra. Sin embargo, como hemos visto, la limitación que esta hipótesis ha impuesto en la «aguja del Creador» no es menos extraordinaria ni menos precisa que toda la notable y delicada coreografía organizada que constituyen las leyes dinámicas que hemos llegado a comprender a través de las ecuaciones de Newton, Maxwell, Einstein, Schrödinger, Dirac y otros. Aunque pueda parecer que la segunda ley de la termodinámica tiene un carácter vago y estadístico, ella surge de una ligadura geométrica de la más suprema precisión. Me parece poco razonable desesperar de obtener una comprensión científica de las ligaduras que son operativas en la «condición de contorno» que era el big bang cuando la aproximación científica ha demostrado ser tan valiosa para la comprensión de las ecuaciones dinámicas. Para mi modo de pensar, lo primero forma parte de la ciencia tanto como lo segundo, aunque una parte de la ciencia que por ahora no entendemos adecuadamente.

La historia de la ciencia nos ha enseñado lo valiosa que ha sido la idea según la cual las *ecuaciones dinámicas* de la física (leyes de Newton, ecuaciones de Maxwell, etc.) han sido separadas de las llamadas *condiciones de contorno* —condiciones que hay que imponer para que las soluciones apropiadas de estas ecuaciones sean seleccionadas de entre el laberinto de las soluciones inapropiadas. Históricamente han sido las ecuaciones dinámicas las que han resultado tener formas sim-

ples. Los movimientos de las partículas satisfacen leyes sencillas pero, muy a menudo, las *configuraciones reales* de las partículas que resultan en el universo no parecen serlo. A veces tales disposiciones parecen sencillas a primera vista —como es el caso de las órbitas elípticas del movimiento planetario, como las concibió Kepler— pero luego se encuentra que su simplicidad es una consecuencia de las leyes dinámicas. La comprensión más profunda ha derivado siempre de las leyes dinámicas, y tales disposiciones simples tienden también a ser simples aproximaciones a otras mucho más complicadas, tales como los movimientos planetarios perturbados (no-muy-elípticos) que se observan realmente, siendo éstos explicados a partir de las ecuaciones dinámicas de Newton. Las condiciones de contorno sirven para «arrancar» el sistema en cuestión y las ecuaciones dinámicas lo llevan en adelante. Una de las realizaciones más importantes de la ciencia física es que podemos separar el comportamiento dinámico de la cuestión de la disposición de los contenidos reales del universo.

He dicho que esta separación en ecuaciones dinámicas y condiciones de contorno ha sido históricamente de vital importancia. El hecho de que sea posible hacer esta separación es una propiedad del tipo *particular* de ecuaciones (ecuaciones diferenciales) que parecen darse siempre en la física. Pero no creo que ésta sea una división que vaya a permanecer. En mi opinión, cuando finalmente lleguemos a comprender las leyes, o principios, que gobiernan *realmente* el comportamiento de nuestro universo —en lugar de esas maravillosas aproximaciones que hemos llegado a comprender, y que constituyen nuestras teorías SOBERBIAS hasta la fecha— encontraremos que esta distinción entre ecuaciones dinámicas y condiciones de contorno desaparecerá. En su lugar, habrá sólo un maravilloso esquema consistente y general. Por supuesto que al decir esto estoy expresando una opinión muy personal; muchos otros podrían no estar de acuerdo con ella. Pero es un punto de vista como éste el que tengo vagamente en la cabeza al tratar de explorar las implicaciones de alguna teoría desconocida de la gravitación cuántica. (Este punto de vista afectará también a algunas de las consideraciones más especulativas del capítulo final.)

¿Cómo explorar las implicaciones de una teoría desconocida? Quizá las cosas no sean en absoluto tan desesperadas como pueda parecer. ¡Consistencia es la clave! En primer lugar, estoy pidiendo al lector que acepte que nuestra teoría supuesta —que llamaré GQC («gravitación cuántica correcta») —proporcionará una explicación de la hipótesis de curvatura de Weyl (HCW). Esto significa que las singula-

tidades *iniciales* deben estar limitadas de modo que $WEYL = 0$ en el futuro inmediato de la singularidad. Esta ligadura existirá como consecuencia de las leyes de la GQC, y por lo tanto debe aplicarse a *cualquier* «singularidad inicial», no sólo a la singularidad concreta que conocemos como «big bang». No estoy diciendo que sea necesario que *haya* singularidades iniciales en nuestro universo real además del big bang, sino que la idea es que, *si* las hubiera, cualquiera de ellas tendría que estar limitada por la HCW. Una singularidad inicial sería una de la cual, en principio, podrían proceder las partículas. Este es el comportamiento contrario al de las singularidades de los agujeros negros, que son singularidades *finales* —en las que pueden caer las partículas.

Un tipo posible de singularidad distinta de la del big bang sería la singularidad en un *agujero blanco* que, como recordamos en el Capítulo 7, es el inverso temporal de un agujero negro (volver a la Fig. 7.14). Pero hemos visto que las singularidades en el interior de los agujeros negros satisfacen $WEYL \rightarrow \infty$, de modo que para un agujero blanco también debemos tener $WEYL \rightarrow \infty$. Pero la singularidad es ahora una singularidad *inicial*, para la que la HCW requiere $WEYL = 0$. En consecuencia, ¡la HCW *descarta* la ocurrencia de agujeros blancos en nuestro universo! (Afortunadamente esto no sólo es deseable sobre bases termodinámicas —pues los agujeros blancos desobedecerían descaradamente la segunda ley de la termodinámica— sino que ¡también es consistente con las observaciones! De cuando en cuando, diversos astrofísicos han postulado la existencia de agujeros blancos para intentar explicar ciertos fenómenos, pero esto plantea siempre más problemas de los que resuelve.) Nótese que no estoy calificando al propio big bang como «agujero blanco». Un agujero blanco poseería una singularidad inicial *localizada* que no podría satisfacer $WEYL = 0$; pero el big bang omni-comprensivo *puede* tener $WEYL = 0$ y la HCW le permite existir con tal de que tenga esta ligadura.

Existe otro tipo de posibilidad para una «singularidad inicial», a saber: el mismo punto de *explosión de un agujero negro* que finalmente ha *desaparecido* ¡tras los (pongamos por caso) 10^{64} años de evaporación de Hawking! (p. 428; ver también p. 449 a continuación) Hay muchas especulaciones sobre la naturaleza exacta de este supuesto (y argumentado muy plausiblemente) fenómeno. Pienso que es probable que no haya aquí conflicto con la HCW. Semejante explosión (localizada) podría ser en efecto instantánea y simétrica, y no veo conflicto con la hipótesis $WEYL = 0$. (En cualquier caso, supo-

niendo que no existan mini-agujeros negros (cf. p. 428) es probable que la primera de tales explosiones no tenga lugar antes de que el universo alcance un tiempo de existencia ¡ 10^{54} veces mayor que el tiempo T que lleva existiendo ahora! Para hacernos una idea de lo que significa $10^{54} \times T$, pensemos que si T se comprimiera hasta el más corto intervalo que puede ser medido —el más pequeño de los tiempos de desintegración de cualquier partícula inestable— entonces nuestra edad *real* del universo presente, en esta escala, se quedaría corta respecto al $10^{54} \times T$ ¡en un factor de un billón!

Algunos siguen una línea diferente de la que estoy proponiendo. Ellos argumentarán que la GQC no debería tener asimetría temporal sino que, de hecho, permitiría *dos* tipos de estructura para las singularidades, uno de los cuales exige $\text{WEYL} = 0$, y el otro que permite $\text{WEYL} \rightarrow \infty$. Habría una singularidad del primer tipo en nuestro universo, y nuestra percepción de la dirección del tiempo (debido a la segunda ley subsiguiente) es tal que sitúa esta singularidad en lo que llamamos el «pasado» más que en lo que llamamos el «futuro». Sin embargo, me parece que este razonamiento no es adecuado tal como está. No explica por qué no hay *otras* singularidades iniciales del tipo $\text{WEYL} \rightarrow \infty$ (ni otras del tipo $\text{WEYL} = 0$). ¿Por qué, según esta idea, el universo no está plagado de agujeros blancos? Puesto que presumiblemente *está* plagado de agujeros *negros* necesitamos una explicación de por qué no existen los blancos *.

Otro argumento que se invoca a veces en este contexto es el llamado *principio antrópico* (cf. Barrow y Tipler, 1986). Según este razonamiento, el universo particular en que nosotros mismos nos observamos habitar está seleccionado de entre todos los universos *posibles* por el hecho de que ¡es necesario que *nosotros* (o, al menos, alguna especie de criatura sintiente) estemos presentes para observarlo! (Discutiré otra vez el principio antrópico en el Capítulo 10.) Se afirma, utilizando este argumento, que los seres inteligentes sólo podrían habitar un universo con un tipo de big bang muy especial —y, por lo tanto, algo semejante a la HCW podría ser una consecuencia de este principio. Sin embargo, el argumento nunca puede aproximar-

* Alguien podría argumentar (correctamente) que las observaciones no son suficientemente claras, ni mucho menos, para apoyar mi afirmación de que existen agujeros negros pero no agujeros blancos en el Universo. Sin embargo, mi argumento es básicamente teórico. Los agujeros negros están de acuerdo con la segunda ley de la termodinámica, ¡pero los agujeros blancos no lo están! (Por supuesto, se podría *postular* sencillamente la segunda ley y la ausencia de agujeros blancos, pero aquí estamos intentando penetrar más profundamente que esto en los orígenes de la segunda ley.)

se a la cifra requerida de $10^{10^{123}}$, para la «particularidad» del big bang a que llegábamos en el Capítulo 7 (cf. p. 344). Mediante un cálculo aproximado se puede ver que todo el Sistema Solar junto con todos sus habitantes pudo ser creado sencillamente a partir de colisiones aleatorias de partículas de un modo más «barato» que éste, a saber: con una «improbabilidad» (medida en términos de volumen del espacio de fases) de «sólo» una parte en mucho menos de $10^{10^{60}}$. Esto es todo lo que el principio antrópico puede hacer por nosotros, y aún nos quedamos enormemente cortos respecto a la cifra requerida. Además, como sucedía con el punto de vista discutido inmediatamente antes, este principio antrópico no ofrece explicación para la ausencia de agujeros blancos.

Asimetría temporal en la reducción del vector de estado

Parece que tenemos que aceptar la conclusión de que la GQC debe ser una teoría con asimetría temporal, en donde la HCW (o algo muy parecido) sea una de las consecuencias de la teoría. ¿Cómo es posible obtener una teoría con asimetría temporal a partir de dos ingredientes con simetría temporal: la teoría cuántica y la relatividad general? Existen, según parece, algunas posibilidades técnicas imaginables para conseguirlo, ninguna de las cuales ha sido explorada en profundidad (cfr. Ashtekar *et al.*). Sin embargo, deseo examinar una línea diferente. He indicado que la teoría cuántica es «tiempo-simétrica», pero en realidad esto sólo se aplica a la parte **U** de la teoría (ecuación de Schrödinger, etc.). En mis discusiones de la simetría temporal de las leyes físicas, al principio del Capítulo 7, he dejado fuera deliberadamente la parte **R** (colapso de la función de onda). Parece predominar la opinión de que también **R** debería tener simetría temporal. Quizá esta opinión sea debida en parte a una hostilidad a considerar **R** como un «proceso» real independiente de **U**, de modo que la simetría temporal de **U** debería implicar la simetría temporal también para **R**. Deseo argumentar que esto *no es así*: **R** es tiempo-asimétrica al menos si consideramos simplemente que «**R**» significa el procedimiento que adoptan realmente los físicos cuando calculan probabilidades en mecánica cuántica.

Permítaseme primero recordar al lector el procedimiento que se aplica en mecánica cuántica denominado reducción del vector de estado (**R**) (recuérdese la Fig. 6.23). En la Figura 8.1 he indicado esquemáticamente la extraña manera en que se considera que evolucio-

na el vector de estado $|\psi\rangle$ en mecánica cuántica. En su mayor parte esta evolución se considera que procede según la evolución *unitaria* U (ecuación de Schrödinger), pero en instantes diversos, cuando se estima que ha tenido lugar una «observación» (o «medida»), se adopta el procedimiento R y el vector de estado $|\psi\rangle$ *salta* a otro vector de estado, por ejemplo $|\chi\rangle$, donde $|\chi\rangle$ es una entre dos o más posibilidades alternativas $|\chi\rangle$, $|\varphi\rangle$, $|\theta\rangle$,... que vienen determinadas por la naturaleza de la observación concreta O que se está llevando a cabo. Ahora bien, la probabilidad p de saltar $|\psi\rangle$ a $|\chi\rangle$ viene dada por la cantidad en que se reduce la longitud al cuadrado $|\psi|^2$ de $|\psi\rangle$ cuando se proyecta $|\psi\rangle$ sobre la dirección de $|\chi\rangle$ (en el espacio de Hilbert). (Matemáticamente esto es lo mismo que la cantidad en que se reduciría $|\chi|^2$, si se proyectara $|\chi\rangle$ en la dirección de $|\psi\rangle$.) Tal como está, este procedimiento tiene asimetría temporal puesto que inmediatamente *después* de que se haya hecho la observación O , el vector de estado es uno del *conjunto dado* $|\chi\rangle$, $|\varphi\rangle$, $|\theta\rangle$,... de posibilidades alternativas *determinado por* O , mientras que inmediatamente *antes* de O , el vector de estado era $|\psi\rangle$, que *no* tiene por qué ser una de estas alternativas dadas. Sin embargo, esta asimetría es sólo aparente y puede remediarse adoptando un punto de vista diferente sobre la evolución del vector de estado. Consideremos una evolución mecano-cuántica *invertida en el tiempo*. Esta excéntrica descripción se ilustra en la Figura 8.2. Supongamos ahora que el estado es $|\chi\rangle$ inmediatamente *antes* de O , en lugar de inmediatamente después, y supongamos que la evolución unitaria se aplica *hacia atrás en el tiempo* hasta el instante de la observación *previa* O' . Supongamos que este estado evolucionado hacia atrás se transforma en $|\chi'\rangle$ (inmediatamente en el futuro de la observación O'). En la descripción normal evolucionada hacia adelante de la Figura 8.1 teníamos algún otro estado $|\psi'\rangle$ en el

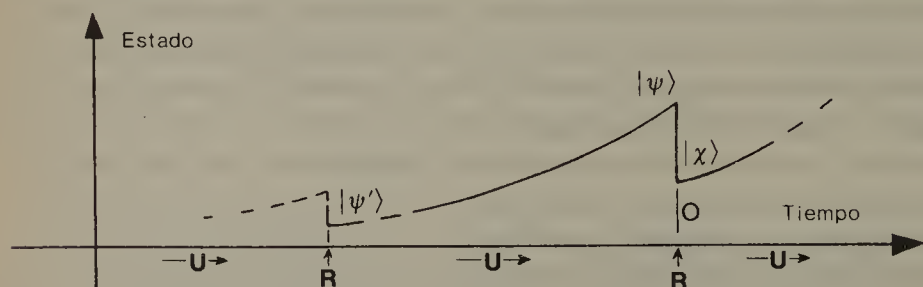


Fig. 8.1. Evolución temporal de un vector de estado: evolución unitaria suave U (según la ecuación de Schrödinger) interrumpida por la reducción discontinua R del vector de estado.

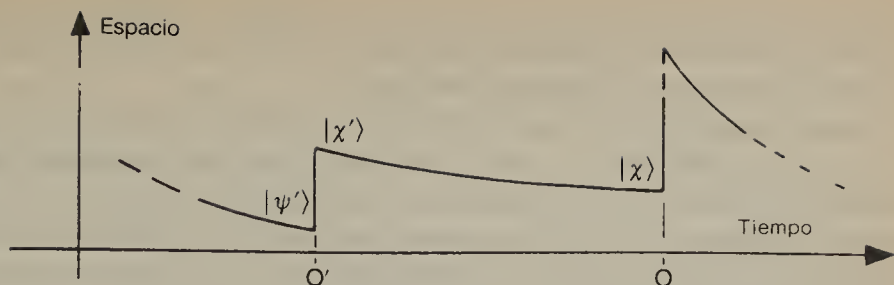


Fig. 8.2. Una imagen más excéntrica de la evolución del vector de estado, en la que se utiliza una descripción invertida en el tiempo. La probabilidad calculada que relaciona la observación en O con la observación en O' sería la misma que en la Fig. 8.1., pero ¿a que se refiere ahora este valor calculado?

futuro inmediato de O' (el resultado de la observación de O , de donde $|\psi'\rangle$ evolucionaría hacia delante hasta $|\psi\rangle$ en O en la descripción normal). Ahora, en nuestra descripción *invertida*, el vector de estado $|\psi'\rangle$ también tiene un papel: representa el estado del sistema en el *pasado* inmediato de O' . El vector de estado $|\psi'\rangle$ es el estado que se observaba realmente en O' , de modo que en nuestro punto de vista con evolución hacia atrás consideramos que el estado $|\psi'\rangle$ es el «resultado», en el sentido *inverso en el tiempo*, de la observación en O' . El cálculo de la probabilidad cuántica p' que relaciona el resultado de la observación en O' con la observación en O , viene dado ahora por la cantidad en que decrece $|\chi'|^2$ al proyectar $|\chi'\rangle$ sobre la dirección de $|\psi'\rangle$ (siendo esta la misma cantidad en la que decrece $|\psi'|^2$ cuando se proyecta $|\psi'\rangle$ sobre la dirección de $|\chi'\rangle$). Es una propiedad fundamental de la operación de U el que efectivamente este es exactamente el mismo valor que teníamos antes ³.

Podría *parecer*, por lo tanto, que hemos establecido que la *teoría cuántica posee simetría temporal*, incluso cuando tenemos en cuenta el proceso discontinuo descrito por la reducción del vector de estado R , además de la evolución unitaria ordinaria U . Sin embargo, *no es así*. Lo que describe la probabilidad cuántica p —de cualquier forma que se calcule— es la probabilidad de encontrar el resultado (a saber $|\chi\rangle$) en O *dado* el resultado (a saber $|\psi'\rangle$) en O' . Esta no es necesariamente la misma que la probabilidad del resultado en O' *dado* el resultado en O . La última sería realmente la que estaría obteniendo nuestra mecánica cuántica invertida en el tiempo. Es curioso cuántos físicos parecen suponer tácitamente que estas dos probabilidades son iguales. (Yo mismo he sido culpable de esta presunción, cf. Penrose,

1979b, p. 584.) Sin embargo, es probable que estas dos probabilidades sean completamente diferentes, de hecho, ¡y sólo la primera está dada correctamente por la mecánica cuántica!

Veamos esto es un caso concreto muy simple. Supongamos que tenemos una lámpara L y una fotocélula (es decir, un detector de fotones) F. Entre L y F tenemos un espejo semirreflectante E que está inclinado un cierto ángulo, pongamos que 45° , respecto a la recta que une L y F (ver Fig. 8.3). Supongamos que la lámpara emite un fotón de cuando en cuando, de un modo aleatorio, y que la lámpara está construida (se podrían utilizar espejos parabólicos) de modo que estos fotones se dirigen exactamente hacia F. Cuandoquiera que la fotocélula reciba un fotón registrará este hecho, y suponemos que tiene una fiabilidad del 100 por cien. Podemos suponer también que siempre que se emita un fotón, este hecho se *registra* en L, de nuevo con una fiabilidad del 100 por cien. (No hay conflicto con los principios de la mecánica cuántica en ninguno de estos requisitos ideales, aunque podría haber dificultades para conseguir en la práctica esta eficiencia.)

El espejo semi-reflectante E es tal que refleja exactamente la mitad de los fotones que le llegan y transmite la otra mitad. La función de onda del fotón incide en el espejo y se desdobra en dos. Existe una amplitud $1/\sqrt{2}$ para la parte reflejada de la onda y $1/\sqrt{2}$ para la parte transmitida. Ambas partes deben considerarse «coexistentes» (en la descripción normal hacia delante en el tiempo) hasta el instante en que se estima que se ha realizado una «observación». En este punto, estas alternativas coexistentes se resuelven por si mismas en

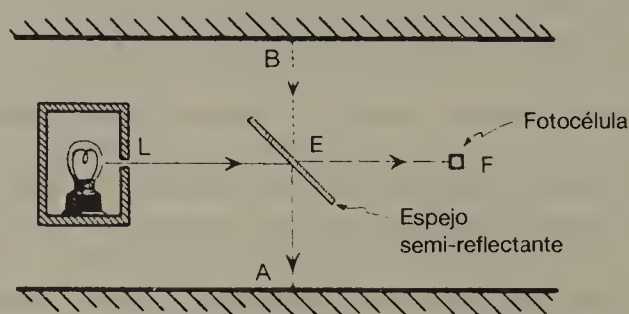


Fig. 8.3. Irreversibilidad temporal de **R** en un experimento cuántico sencillo. La probabilidad de que la fotocélula detecte un fotón, *dado que* la fuente ha emitido uno, es exactamente $\frac{1}{2}$; pero la probabilidad de que la fuente haya emitido un fotón, *dado que* la fotocélula detecta uno *no*, es ciertamente $\frac{1}{2}$.

alternativas *reales* —una o la otra— con probabilidades dadas por los cuadrados de (los módulos de) estas amplitudes, a saber $(1/\sqrt{2})^2 = 1/2$, en cada caso. Cuando se ha hecho la observación, resulta que las probabilidades de que el fotón sea reflejado o transmitido han sido ambas realmente de $1/2$.

Veamos como se aplica esto a nuestro experimento real. Supongamos que se registra que L ha emitido un fotón. La función de onda del fotón se desdobla en el espejo y llega a F con amplitud $1/\sqrt{2}$, de modo que la fotocélula registra o no, con una probabilidad $1/2$ para cada caso. La otra parte de la función de onda llega a un punto A en la *pared del laboratorio* (ver Fig. 8.3), de nuevo con amplitud $1/\sqrt{2}$. Si F *no* registra, entonces debe considerarse que el fotón ha incidido en la pared en A, pues si hubiéramos colocado otra fotocélula en A entonces esta nueva fotocélula registraría siempre que no lo hiciera la fotocélula en F —suponiendo que L ha registrado realmente la emisión de un fotón— y no registraría siempre que registrara F. En este sentido no es necesario colocar una fotocélula en A. Podemos inferir lo que la fotocélula en A *habría* hecho, si hubiera estado allí, simplemente mirando en L y F.

Estaría claro cómo procede el cálculo mecano-cuántico. Planteamos la pregunta:

«Dado que L registra, ¿cuál es la probabilidad de que registre F?»

Para responderla notemos que existe una amplitud $1/\sqrt{2}$ para que el fotón recorra el camino LEF y una amplitud $1/\sqrt{2}$ para que recorra el camino LEA. Elevando al cuadrado encontramos las probabilidades respectivas $1/2$ y $1/2$ de llegar a F y de llegar a A. La respuesta mecano-cuántica a nuestra pregunta es por lo tanto

«un medio».

Ésta es en realidad la respuesta que obtendríamos experimentalmente.

Podríamos utilizar exactamente igual el excéntrico procedimiento con el tiempo invertido para obtener la misma respuesta. Supongamos que notamos que F ha registrado. Consideremos una función de onda con el tiempo invertido para el fotón, suponiendo que el fotón llega finalmente a F. A medida que seguimos la pista hacia atrás en el tiempo, el fotón retrocede desde F hasta que alcanza el espejo E. En este punto la función de onda se bifurca y existe una amplitud $1/\sqrt{2}$ de que llegue a la lámpara L, y una amplitud $1/\sqrt{2}$ de que sea reflejada en E hasta llegar a *otro* punto en la pared del laboratorio, a saber, B en la Figura 8.3. Elevando al cuadrado obtenemos de nuevo

el valor $1/2$ para las dos probabilidades. No obstante, debemos tener cuidado al señalar a qué preguntas responden estas respuestas. Existen las dos preguntas, «Dado que L registra, ¿cuál es la probabilidad de que F registre?», igual que antes, y la pregunta más excéntrica, «Dado que el fotón es lanzado desde la pared en B, ¿cuál es la probabilidad de que F registre?»

Podemos considerar que, en cierto sentido, ambas respuestas son experimentalmente «correctas», ¡aunque la segunda (lanzamiento desde la pared) sería una inferencia más que el resultado de una serie *real* de experimentos! Sin embargo, ninguna de estas preguntas es *la inversa en el tiempo* de la pregunta que plateábamos antes. Ésta sería:

«Dado que F registra, ¿cuál es la probabilidad de que L registre?»

Notemos que la respuesta experimental *correcta* a esta pregunta no es en absoluto «un medio» sino

«uno»

Si la fotocélula registra realmente entonces ¡es virtualmente cierto que el fotón procede de la *lámpara* y no de la pared del laboratorio! En el caso de nuestra pregunta invertida en el tiempo, el cálculo mecano-cuántico nos ha dado ¡una respuesta totalmente errónea!

La consecuencia de esto es que las reglas para la parte **R** de la mecánica cuántica no pueden utilizarse simplemente para estas preguntas con inversión temporal. Si queremos calcular la probabilidad de un estado *pasado* sobre la base de un estado *futuro* conocido, obtendremos respuestas completamente erróneas si tratamos de adoptar el procedimiento estándar **R** consistente en tomar la amplitud mecano-cuántica y elevar su módulo al cuadrado. Este procedimiento sólo funciona para calcular las probabilidades de estados *futuros* sobre la base de estados *pasados* —¡y funciona de manera soberbia! Me parece evidente que, sobre esta base, el procedimiento **R** *no puede tener simetría temporal* (y, dicho sea de paso, no puede haber por consiguiente una deducción a partir del procedimiento con simetría temporal **U**).

Mucha gente podría adoptar la postura de que la razón de esta discrepancia con la simetría temporal es que de algún modo la segunda ley de la termodinámica se ha colado en el argumento, introduciendo una asimetría temporal añadida que no está descrita por el procedimiento de tomar el cuadrado de la amplitud. De hecho, parece ser cierto que cualquier dispositivo de medida físico capaz de llevar a cabo el procedimiento **R** debe involucrar una «irreversibilidad ter-

modinámica» —de modo que la entropía incrementa siempre que tiene lugar una medida. Creo que es probable que la segunda ley *esté* involucrada de un modo esencial en el proceso de medida. Además, no parece de mucho sentido físico tratar de invertir el tiempo en la operación *completa* de un experimento mecano-cuántico como el experimento (idealizado) descrito arriba, incluyendo el registro de todas las medidas implicadas. No he tratado la cuestión de hasta dónde podemos ir realmente con la inversión temporal de un experimento; he considerado sólo la aplicabilidad de ese curioso procedimiento mecano-cuántico que obtiene probabilidades correctas elevando al cuadrado los módulos de las amplitudes. Es un hecho sorprendente que este simple procedimiento pueda aplicarse en la dirección del futuro sin que sea necesario invocar ningún otro conocimiento de un sistema. De hecho, forma parte de la teoría el que *no podamos* influir en estas probabilidades: ¡las probabilidades de la teoría cuántica son completamente *estocástica*! Sin embargo, si intentamos aplicar estos procedimientos en la dirección del pasado (es decir, para retrodicción más que para predicción) entonces nos equivocamos completamente. Pueden invocarse cualquier número de excusas, circunstancias atenuantes u otros factores para explicar *por qué* el procedimiento de tomar el cuadrado de las amplitudes no se aplica correctamente en la dirección del pasado, pero sigue en pie el hecho de que no lo hace. ¡Estas excusas son sencillamente innecesarias en la dirección del futuro! El procedimiento **R**, *tal como se utiliza realmente*, no es simétrico respecto al tiempo.

La caja de Hawking: ¿una conexión con la hipótesis de curvatura de Weyl?

Está bien, estará pensando sin duda el lector, pero ¿qué tiene todo esto que ver con la HCW o la GQC? Es cierto que la *segunda ley*, como hoy opera, puede ser parte perfectamente de la operación de **R**, pero ¿dónde hay algún papel destacable para las singularidades del espacio-tiempo o la gravitación cuántica en estas continuas ocurrencias «cotidianas» de reducción del vector de estado? Para abordar esta cuestión deseo describir un espectacular «experimento mental» propuesto originalmente por Stephen Hawking, si bien el propósito con el que se expone no forma parte de lo que Hawking pretendió originalmente.

Imaginemos una caja cerrada de proporciones monstruosas. Se

considera que sus paredes son totalmente reflectantes e impermeables a cualquier influencia. Ningún objeto material puede atravesarlas, ni tampoco ninguna señal electromagnética, o neutrino, o ninguna otra cosa. Todo debe ser reflejado, ya incida desde dentro o desde fuera. Incluso está prohibido que la atraviesen los efectos de la gravitación. No existe ninguna substancia real de la que puedan construirse estas paredes. Nadie podría *realizar* realmente el «experimento» que voy a describir. (¡Ni nadie querría hacerlo, como vamos a ver!) Éste no es el punto. En un experimento mental se trata de descubrir principios generales a partir de la simple consideración mental de experimentos que se *podrían* realizar. Se ignoran las dificultades tecnológicas siempre que estas dificultades no se deriven de los principios generales bajo consideración. (Recordemos la discusión del gato de Schrödinger en el Capítulo 6.) En nuestro caso, las dificultades para construir las paredes de nuestra caja deben considerarse, para este propósito, como puramente «tecnológicas», de modo que estas dificultades serán ignoradas.

En el interior de la caja hay una gran cantidad de substancia material de algún tipo. No importa mucho cuál sea esta substancia. Sólo nos interesa su masa total M , que deberá ser muy grande, y el gran volumen V de la caja que la contiene. ¿Qué vamos a hacer con nuestra caja construida con tanto gasto y con su contenido de tan nulo interés? El experimento va a ser lo más aburrido que se pueda imaginar. Vamos a dejarla aislada ¡para siempre!

La cuestión que nos interesa es el destino final del contenido de la caja. Según la segunda ley de la termodinámica su entropía deberá crecer. La entropía aumentará hasta que se alcance su valor máximo, de modo que el material habrá llegado al «equilibrio térmico». Nada sucedería a partir de entonces, si no fuera por las «fluctuaciones» en las que se alcanzan temporalmente desviaciones (relativamente) breves respecto al equilibrio térmico. En nuestra situación, suponemos que M es suficientemente grande, y V es algo apropiado (*muy* grande, pero no demasiado grande), de modo que cuando se alcance el «equilibrio térmico» la mayor parte del material habrá colapsado en un *agujero negro*, con sólo un poco de materia y radiación rodeándolo —que constituye un llamado «baño térmico» (¡muy frío!) en el que está inmerso el agujero negro. Para ser precisos, podríamos escoger M como la masa del Sistema Solar y V como el tamaño de la Vía Láctea. Entonces la temperatura del «baño» sería sólo de ¡unos 10^{-7} grados por encima del cero absoluto!

Para comprender más claramente la naturaleza de este equilibrio

y estas fluctuaciones recordemos el concepto de *espacio de fases* que ya encontramos en los Capítulos 5 y 7, especialmente en relación con la definición de entropía. La Figura 8.4 da una descripción esquemática de todo el espacio de fases **P** del contenido de la caja de Hawking. Como recordamos, un espacio de fases es un espacio de muchas dimensiones, cada uno de cuyos puntos representa todo un posible estado global del sistema bajo consideración; en nuestro caso, el contenido de la caja. Así, cada punto de **P** codifica las posiciones y momentos de todas las partículas presentes en la caja, además de toda la información necesaria sobre la *geometría espacio-temporal* dentro de la caja. La subregión **B** (de **P**) a la derecha de la Figura 8.4 representa la totalidad de los estados en los que hay un *agujero negro* dentro de la caja (incluyendo todos los casos en los que hay más de un agujero negro), mientras que la subregión **A** de la izquierda representa la totalidad de los estados libres de agujeros negros. Debemos suponer que cada una de las regiones **A** y **B** está además subdividida en compartimentos más pequeños de acuerdo con la división de «grano grueso» que es necesaria para la definición precisa de la entropía (cf. Fig. 7.3, p. 389), pero estos detalles no nos interesan aquí. Todo lo que tenemos que señalar en esta etapa es que el mayor de estos compartimentos —que representa el equilibrio térmico, con presencia de un agujero negro— es la porción mayor de **B**, mientras la mayor porción de **A** (algo más pequeña) es el compartimento que representa lo que *aparenta* ser equilibrio térmico, salvo el hecho de que no hay agujero negro presente.

Recordemos que en cualquier espacio de fases hay un campo de

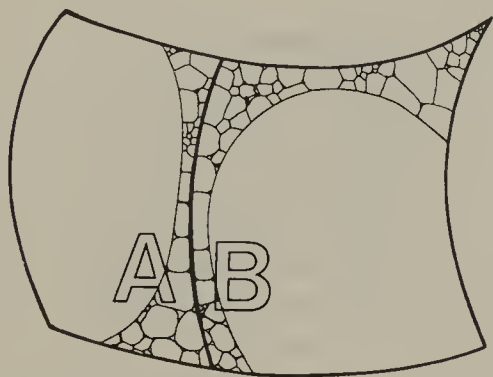


Fig. 8.4. El espacio de fases **P** de la caja de Hawking. La región **A** corresponde a las situaciones en la que *no* existe agujero negro en la caja, y **B** a las situaciones en las que *sí* existe un agujero negro (o más de uno) en la caja.

flechas (campo vectorial) que representa la evolución temporal del sistema físico (ver Capítulo 5, p. 177; también; Fig. 5.11). Así, para ver que sucederá a continuación con nuestro sistema seguimos simplemente a lo largo de las flechas en **P** (ver Fig. 8.5). Algunas de estas flechas pasan de la región **A** a la región **B**. Esto ocurre cuando se forma por primera vez un agujero negro por el colapso gravitatorio de materia. ¿Existen flechas que pasen de la región **B** a la región **A**? Sí que existen, pero sólo si tenemos en cuenta el fenómeno de la *evaporación de Hawking* a que aludimos antes (pp. 428, 439). Según la teoría estrictamente *clásica* de la relatividad general, los agujeros negros sólo pueden engullir cosas; no pueden emitir cosas. Sin embargo, teniendo en cuenta los efectos mecano-cuánticos, Hawking (1975) pudo demostrar que, después de todo, los agujeros negros deberían poder emitir cosas en el nivel cuántico, de acuerdo con el proceso de *radiación de Hawking*. (Esto ocurre por vía del proceso cuántico de «creación de pares virtuales», por el que continuamente se están creando partículas y antipartículas a partir del vacío —por un breve instante— normalmente para aniquilarse entre sí inmediatamente después sin dejar huella. Cuando está presente un agujero negro, sin embargo, puede «engullir» una de las partículas del par antes de que tenga tiempo de ocurrir la aniquilación, y su compañera puede escapar del agujero. Estas partículas que escapan constituyen la radiación de Hawking.) En el curso normal de las cosas, esta radiación de Hawking es realmente pequeñísima. Pero en el estado de equilibrio térmico la cantidad de energía que pierde el agujero negro por

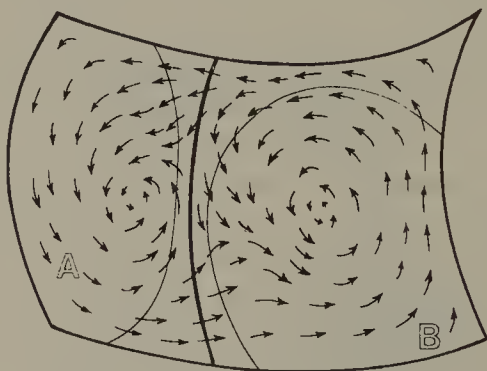


Fig. 8.5. El «flujo hamiltoniano» del contenido de la caja de Hawking (compárese con Fig. 5.11). Las líneas de flujo que cruzan de **A** a **B** representan el colapso en un agujero negro, y las que cruzan de **B** a **A** representan la desaparición de un agujero negro por evaporación.

radiación compensa exactamente la energía que gana tragando otras «partículas térmicas» que se están moviendo por el «baño térmico» en el que se encuentra el propio agujero negro. En ocasiones, debido a una «fluctuación», el agujero podría emitir algo más o engullir un poco menos y, en consecuencia, perder energía. Al perder energía pierde masa (por la ecuación de Einstein $E = mc^2$) y, según las reglas que gobiernan la radiación de Hawking, se hace un poco más caliente. Muy de tarde en tarde, cuando la fluctuación es suficientemente grande, es incluso posible que el agujero negro entre en una situación incontrolada en la que se va haciendo cada vez más caliente, perdiendo cada vez más energía a medida que esto sucede, y haciéndose cada vez más pequeño hasta que finalmente (es de suponer) ¡desaparece completamente en una violenta explosión! Cuando esto sucede (y suponiendo que no haya otros agujeros en la caja) tenemos la situación en la que, en nuestro espacio de fases **P**, pasamos de la región **B** a la región **A**, de modo que ¡realmente *hay* flechas de **B** a **A**!

En este punto, haré un comentario sobre lo que significa una «fluctuación». Recordemos los compartimentos de grano-grueso que consideramos en el último capítulo. Los puntos del espacio de fases que pertenecen a un mismo compartimento deben considerarse «indistinguibles» (macroscópicamente) unos de otros. La entropía incrementa debido a que al seguir las flechas tendemos a entrar en compartimentos cada vez más enormes a medida que avanza el tiempo. Finalmente, el punto en el espacio de fases se pierde en el compartimento más enorme de todos, a saber, el que corresponde al equilibrio térmico (entropía máxima). Sin embargo, esto sólo será verdadero hasta cierto punto. Si esperamos lo suficiente, el punto en el espacio de fases podrá encontrar *eventualmente* un compartimento más pequeño y en consecuencia la entropía descenderá. Normalmente esto no durará mucho (hablando en términos relativos) y la entropía pronto volverá a crecer cuando el punto del espacio de fases entre de nuevo en el compartimento más grande. Esto es una *fluctuación*, con su momentánea disminución de entropía. Normalmente el descenso de la entropía no es muy grande, pero muy de tarde en tarde ocurrirá una *enorme* fluctuación y la entropía podrá descender de forma substancial —y quizá permanecer con un bajo valor durante un intervalo de tiempo considerable.

Éste es el tipo de cosas que necesitamos para ir de la región **B** a la región **A** vía el proceso de evaporación de Hawking. Se necesita una fluctuación muy grande debido a que debe atravesarse un compartimento muy pequeño en el que las flechas pasan de **B** a **A**. Aná-

logamente, cuando nuestro punto del espacio de fases yace en el interior del compartimento más grande de **A** (que representa el estado de equilibrio térmico sin agujeros negros), deberá pasar mucho tiempo antes de que tenga lugar un colapso gravitatorio y el punto se mueva hasta **B**. Se necesita de nuevo una gran fluctuación. (¡La radiación térmica no está dispuesta a sufrir un colapso gravitatorio!)

¿Es *mayor* el número de flechas que llevan de **A** a **B** que el de flechas que llevan de **B** a **A**, o es el *mismo* en ambos casos? Esto será un punto importante para nosotros. Para plantear la cuestión de otra forma: ¿es «más fácil» para la Naturaleza producir un agujero negro mediante el colapso gravitatorio de partículas térmicas, o lo es deshacerse de un agujero negro por radiación de Hawking, o ambas cosas son igualmente «difíciles»? Estrictamente hablando no es el «número» de flechas lo que nos interesa sino la velocidad del flujo del volumen en el espacio de fases. Piénsese en el espacio de fases lleno de algún tipo de fluido incompresible (¡de alta dimensión!). Las flechas representan el flujo de este fluido. Recordemos el *teorema de Liouville* que se describió en el Capítulo 5 (p. 235). El teorema de Liouville afirma que el flujo conserva el volumen en el espacio de fases, ¡que es lo mismo que decir que nuestro fluido en el espacio de fases es realmente incompresible! El teorema de Liouville parece decirnos que el flujo de **A** a **B** debe ser *igual* al flujo de **B** a **A** ya que, al ser incompresible el «fluido» del espacio de fases, no puede acumularse en una u otra parte. ¡Parecería así que debe tener exactamente la misma «dificultad» construir un agujero negro a partir de la radiación térmica que destruirlo!

Esta fue en efecto la propia conclusión de Hawking, aunque él llegó a esta idea basándose en consideraciones algo diferentes. El principal argumento de Hawking era que toda la física básica implicada en el problema tiene *simetría temporal* (relatividad general, termodinámica, los procedimientos unitarios estándar de la teoría cuántica), de modo que si hiciéramos marchar el reloj hacia atrás obtendríamos la misma respuesta que si marchara hacia delante. Esto equivale simplemente a invertir las direcciones de todas las flechas en **P**. Entonces se seguiría también de este razonamiento que debe haber exactamente tantas flechas de **A** a **B** como de **B** a **A** *con tal de que* se verifique que el inverso temporal de la región **B** sea de nuevo la región **B** (y, de modo equivalente, que el inverso temporal de **A** sea de nuevo **A**). Esta condición equivale a la famosa sugerencia de Hawking de que los agujeros negros y sus inversos temporales, a saber, los agujeros blancos, ¡son en realidad físicamente idénticos! Su razo-

namiento era que en una física con simetría temporal el estado de equilibrio térmico también debería tener simetría temporal. No quiero entrar aquí en una discusión detallada de esta sorprendente posibilidad. La idea de Hawking era que la radiación mecano-cuántica de Hawking podía considerarse de algún modo como la inversa temporal del «engullimiento» clásico de material por el agujero negro. Su hipótesis, aunque ingeniosa, implica diversas dificultades teóricas y yo personalmente no creo que pueda funcionar.

En cualquier caso, esta hipótesis no es compatible con las ideas que estoy desarrollando aquí. He argumentado que, mientras que los agujeros negros deberían existir, ¡los agujeros blancos están *prohibidos* por la *hipótesis de curvatura de Weyl*! HCW introduce en la discusión una *asimetría temporal* que no era considerada por Hawking. Debería señalarse que puesto que los agujeros negros y sus singularidades en el espacio-tiempo son una parte importante de la discusión de lo que sucede en el interior de la caja de Hawking, la física desconocida que debe gobernar el comportamiento de tales singularidades está ciertamente involucrada. Hawking adopta la postura de que esta física desconocida debería ser una teoría de la gravitación cuántica *con simetría temporal*, ¡mientras que yo estoy defendiendo que es la GQC con *asimetría temporal*! Estoy defendiendo que una de las implicaciones principales de la GQC debería ser la HCW (y, por consiguiente, la segunda ley de la termodinámica en la forma que la conocemos), de modo que trataremos de descubrir las implicaciones de la HCW para nuestro problema actual.

Veamos cómo afecta la inclusión de la HCW a la discusión del flujo de nuestro «fluido incompresible» en **P**. En el espacio-tiempo, el efecto de la singularidad de un agujero negro es el de absorber y destruir toda la materia que incida en él. Lo que es más importante para nuestros propósitos presentes, ¡*destruye información*! El efecto de esto, en **P**, es que algunas líneas de flujo convergerán (ver Fig. 8.6). Dos estados que antes eran diferentes pueden llegar a ser el mismo tan pronto como se destruya la información que hacía distinción entre ellos. Cuando convergen las líneas de flujo en **P** tenemos una *violación* efectiva del teorema de Liouville. Nuestro «fluido» ya no es incompresible, sino que ¡se está *aniquilando continuamente* dentro de la región **B**!

Parece que ahora estamos en dificultades. Si nuestro «fluido» se está destruyendo continuamente en la región **B**, entonces debe haber más líneas de flujo de **A** a **B** que de **B** a **A** —de modo que, después de todo, ¡es «más fácil» crear un agujero negro que destruirlo! Esto



Fig. 8.6. En la región **B** las líneas de flujo deben converger debido a la pérdida de información en las singularidades de los agujeros negros. ¿Se compensa esto con una creación de líneas de flujo debido al procedimiento cuántico **R** (principalmente en la región **A**)?

podría tener sentido si no fuera por el hecho de que ahora sale más «fluido» de la región **A** del que entra. No hay agujeros negros en región **A** —y los agujeros blancos están prohibidos por la HCW— de modo que ciertamente el teorema de Liouville ¡debería seguir siendo perfectamente válido en la región **A**! Sin embargo, parece ahora que necesitamos algún medio de «crear fluido» en la región **A** para enmascarar la pérdida en la región **B**. ¿Qué mecanismo puede haber para incrementar el número de líneas de flujo? Lo que parece que necesitamos es que a veces un mismo estado pueda tener más de un posible sucesor (es decir, líneas de flujo que se bifurcan). Este tipo de incertidumbre en la evolución futura de un sistema físico lleva el «aroma» de la parte **R** de la teoría cuántica. ¿Sería posible que **R** sea, en cierto sentido, «la otra cara de la moneda» de la HCW? Mientras que la HCW sirve para que converjan líneas de flujo en **B**, el procedimiento mecano-cuántico **R** provoca que las líneas de flujo se bifurquen. Estoy afirmando, de hecho, que es un proceso *objetivo* mecano-cuántico de reducción del vector de estado (**R**) el que provoca que las líneas de flujo se bifurquen, y de este modo ¡compensa exactamente la fusión de líneas de flujo debida a la HCW! (Fig. 8.6)

Para que tenga lugar dicha bifurcación necesitamos que **R** tenga asimetría temporal, como ya hemos visto: recuérdese nuestro experimento anterior con la lámpara, la fotocélula y el espejo semi-reflektante. Cuando la lámpara emite un fotón hay dos alternativas (igualmente probables) para el resultado final: o bien el fotón llega a la fotocélula y la fotocélula registra, o bien el fotón llega a la pared en **A** y la fotocélula no registra. En el espacio de fases para este expe-

rimento tenemos una línea de flujo que representa la emisión del fotón y esta línea se bifurca en dos: una que describe la situación en la que la fotocélula se dispara, y la otra que refleja la situación en que no lo hace. Ésta parece ser una auténtica bifurcación debido a que sólo hay un estado de partida permitido y hay dos resultados posibles. El otro estado de partida que debiéramos haber tenido que considerar era la posibilidad de que el fotón fuera lanzado desde la pared del laboratorio en **B**, en cuyo caso habría dos estados de partida y dos resultados. Pero este estado de partida alternativo ha sido excluido sobre la base de la inconsistencia con la segunda ley de la termodinámica —es decir, desde el punto de vista aquí expresado, excluido finalmente por la HCW cuando se rastrea la evolución hacia el pasado.

Reiteraré que el punto de vista que estoy exponiendo no es realmente «convencional» —aunque no tengo completamente claro que es lo que diría un físico «convencional» para resolver las cuestiones que aquí se plantean. (¡Sospecho que no muchos de ellos han dedicado mucha atención a estos problemas!) Ciertamente he oído varias opiniones diferentes. Por ejemplo, se ha sugerido de tarde en tarde, por algunos físicos, que la radiación de Hawking nunca podría dar lugar a que un agujero negro desaparezca *completamente*, sino que debería quedar siempre alguna pequeña «pepita». (Por lo tanto, en esta perspectiva, ¡*no* hay flechas que vayan de **B** a **A**!) Esto apenas afecta a mi argumento (y, en realidad, lo refuerza). Podrían evitarse mis conclusiones, no obstante, postulando que el volumen total del espacio de fases **P** es *infinito*, pero esto está en contra de ciertas ideas bastante básicas sobre la entropía de los agujeros negros y sobre la naturaleza del espacio de fases de un sistema (cuántico) acotado; y otras formas que he oído de evitar técnicamente mis conclusiones no me parecen más satisfactorias. Una objeción mucho más seria es la de que las idealizaciones implicadas en la construcción real de una caja de Hawking son demasiado grandes y se transgreden ciertas cuestiones de principio al suponer que se puede construir. Yo mismo no estoy muy seguro de ello, aunque estoy inclinado a creer que las idealizaciones necesarias pueden ser realmente admitidas.

Finalmente, hay un punto importante que quiero tratar. Empecé la discusión suponiendo que teníamos un espacio de fases *clásico* —y el teorema de Liouville se aplica a la física clásica. Pero entonces necesitaba ser considerado el fenómeno cuántico de la radiación de Hawking. (Y también se necesitaba la teoría cuántica para la *dimensionalidad finita* tanto como el volumen finito de **P**.) Como vimos en

el Capítulo 6, la versión cuántica del espacio de fases es el *espacio de Hilbert*, de modo que presumiblemente deberíamos haber utilizado a lo largo de toda nuestra discusión el espacio de Hilbert antes que el espacio de fases. En el espacio de Hilbert *existe* un teorema análogo al de Liouville; surge de lo que se llama la naturaleza «unitaria» de la evolución temporal U . Quizá todo mi argumento podría expresarse completamente en términos de espacio de Hilbert, en lugar del espacio de fases clásico, pero es difícil ver cómo discutir de esta forma el fenómeno clásico implícito en la geometría del espacio-tiempo de los agujeros negros. Mi opinión es que para la teoría *correcta* no serían apropiados ni el espacio de Hilbert ni el espacio de fases clásico, sino que tendríamos que utilizar algún tipo de espacio matemático hasta ahora no descubierto que fuera intermedio entre los dos. Por consiguiente, mi argumento debería considerarse sólo en un nivel heurístico, y es simplemente *sugerente* más que concluyente. De todas formas, creo que proporciona un fuerte motivo para pensar que la HCW y R están profundamente relacionados y que, en consecuencia, R *debe ser en realidad un efecto de gravitación cuántica*.

Para reiterar mis conclusiones: estoy desarrollando la hipótesis de que la reducción mecano-cuántica del vector de estado es en realidad la otra cara de la moneda de la HCW. Según esta idea, las dos implicaciones principales de nuestra deseada teoría de la «gravitación cuántica correcta» (GQC) serán la HCW y R . El efecto de la HCW es la *confluencia* de líneas de flujo en el espacio de fases, mientras que el efecto de R es una *ramificación* de las líneas de flujo exactamente compensadora. Ambos procesos están íntimamente asociados a la segunda ley de la termodinámica.

Nótese que la confluencia de líneas de flujo tiene lugar siempre dentro de la región B , mientras que la ramificación de líneas de flujo puede tener lugar en A o en B . Recordemos que A representa la *ausencia* de agujeros negros, así que la reducción del vector de estado puede tener lugar cuando no hay agujeros negros. Evidentemente no es necesario tener un agujero negro en el laboratorio para que R sea efectivo (como sucede en nuestro experimento con el fotón recién considerado.) Aquí sólo estamos interesados con un balance global entre posibles cosas que *podrían* suceder en una situación. Según la idea que estoy expresando, es simplemente la *posibilidad* de que pudieran formarse agujeros negros en alguna etapa (y en consecuencia, se destruya información) la que debe ser compensada por la falta de determinismo en la teoría cuántica.

¿Cuándo se reduce el vector de estado?

Supóngase que aceptamos, sobre la base de los argumentos precedentes, que la reducción del vector de estado pudiera ser de alguna forma un fenómeno gravitatorio en última instancia. ¿Pueden hacerse más explícitas las relaciones entre **R** y la gravitación? Sobre la base de esta idea, ¿cuándo tendría lugar *realmente* el colapso del vector de estado?

Señalaré primero que incluso en las aproximaciones más «convencionales» a la teoría de la gravitación cuántica existen algunas dificultades técnicas serias en conciliar los principios de la relatividad general con las reglas de la teoría cuántica. Estas reglas (principalmente la forma de reinterpretar el momento como derivada respecto a la posición, en la expresión para la ecuación de Schrödinger, cf. p. 363) no se ajustan en absoluto a las ideas de la geometría del espacio-tiempo curvo. Mi propio punto de vista es que tan pronto como se introduce una cantidad «significativa» de curvatura espacio-temporal, las reglas de la superposición lineal cuántica deben fallar. Es en este punto donde las superposiciones de amplitudes complejas de estados potenciales alternativos quedan reemplazadas por alternativas reales con pesos probabilísticos— y una de estas alternativas tiene lugar *realmente*.

¿Qué entiendo por una cantidad «significativa» de curvatura? Entiendo que se ha alcanzado el nivel en el que la medida de la curvatura introducida tiene aproximadamente la escala de *un gravitón*⁵ o más. (Recuérdese que, según las reglas de la teoría cuántica, el campo electromagnético está «cuantizado» en unidades individuales, llamadas «fotones». Cuando se descompone el campo en sus frecuencias individuales, la parte de frecuencia ν puede llegar sólo en números enteros de fotones, cada uno de energía $h\nu$. Reglas análogas se aplicarán presumiblemente al campo gravitatorio.) Un gravitón sería la unidad más pequeña de curvatura que estaría permitida por la teoría cuántica. La idea es que, en cuanto se alcanza este nivel, las reglas ordinarias de superposición lineal, de acuerdo con el procedimiento **U**, se modificarán cuando se apliquen a gravitones, y se establecerá algún tipo de «inestabilidad no-lineal» con asimetría temporal. En lugar de tener para siempre superposiciones lineales complejas de «alternativas» coexistentes, una de las alternativas vence en esta etapa y el sistema «cae» en una alternativa u otra. Quizá la elección de alternativa se hace por simple azar, o quizá hay algo más profundo que subyace a esta elección. Pero ahora, una u otra se ha hecho *realidad*. El procedimiento **R** ha tenido efecto.

Nótese que, según esta idea, el procedimiento **R** ocurre espontáneamente de una manera totalmente objetiva, independiente de cualquier intervención humana. La idea es que el nivel de «un gravitón» estaría cómodamente situado entre el «nivel cuántico» de átomos, moléculas, etc., en donde son válidas las reglas lineales (**U**) de la teoría cuántica ordinaria, y el «nivel clásico» de nuestras experiencias cotidianas. ¿Cuál es la magnitud de este nivel de un gravitón? Debe subrayarse que no se trata realmente de una cuestión de *tamaño* físico: es más una cuestión de distribución de masa y energía. Hemos visto que los efectos de interferencia cuántica pueden ocurrir a distancias muy grandes siempre que no haya mucha energía implicada. (Recuérdese la autointerferencia del fotón descrita en la p. 254 y los experimentos EPR de Clauser y Aspect, p. 286.) La escala de masas características de la gravitación cuántica es la que se conoce como la *masa de Planck*

$$m_p = 10^{-5} \text{ gramos}$$

(aproximadamente). Esto parece ser bastante mayor de lo que nos gustaría, ya que objetos de masa mucho *menor* que ésta, como motas de polvo, pueden percibirse directamente comportándose de forma clásica. (La masa m_p es un poco más pequeña que la de una pulga.) Sin embargo, no creo que el criterio de un-gravitón tuviera que aplicarse de un modo tan crudo como éste. Trataré de ser un poco más explícito, pero en el momento de escribir este libro sigue habiendo demasiadas oscuridades y ambigüedades como para que este criterio se aplique exactamente.

Consideremos en primer lugar una manera muy directa de poder observar una partícula: utilizando una *cámara de niebla* de Wilson. En este caso tenemos una cámara llena de vapor que está precisamente en el punto de condensación en gotas. Cuando una partícula cargada que se mueve rápidamente entra en esta cámara, habiendo sido producida, pongamos por caso, por la desintegración de un átomo radiactivo situado fuera de la cámara, su paso a través del vapor provoca que algunos átomos próximos a su camino se ionicen (es decir, se carguen eléctricamente debido a que algunos electrones son desalojados). Estos átomos ionizados actúan como centros en los que el vapor se condensa en pequeñas gotas. De esta forma, tenemos una traza de gotas que el experimentador puede observar directamente (Fig. 8.7).

Ahora bien, ¿cuál es la descripción mecano-cuántica de esto? En el instante en que se desintegra nuestro átomo radiactivo, emite una

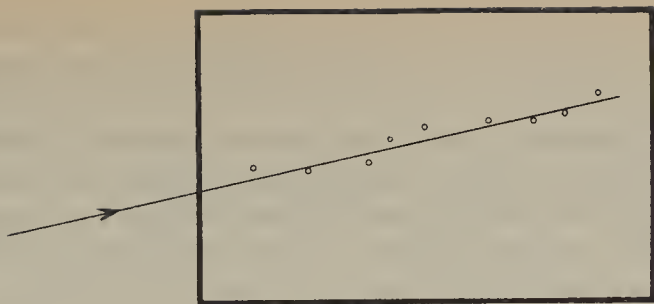


Fig. 8.7. Una partícula cargada que entra en una cámara de niebla de Wilson y provoca la condensación de una cadena de gotas.

partícula. Pero existen muchas direcciones posibles en las que podría viajar dicha partícula emitida. Habrá una amplitud para esta dirección, una amplitud para aquella dirección, y una amplitud para cualquier otra dirección, todas ellas ocurriendo simultáneamente en una superposición lineal cuántica. La totalidad de todas estas alternativas superpuestas constituye una onda esférica que emana del átomo desintegrado: la *función de onda* de la partícula emitida. A medida que cada posible traza de partícula entra en la cámara de niebla queda asociada a una cadena de átomos ionizados, cada uno de los cuales comienza a actuar como un centro de condensación para el vapor. Todas estas diferentes cadenas de átomos ionizados posibles deben coexistir también en una superposición lineal cuántica, de modo que ahora tenemos una superposición lineal de un gran número de *diferentes* cadenas de gotas condensadas. En cierta etapa, esta superposición lineal cuántica compleja se transforma en una colección de alternativas *reales* con pesos probabilistas reales que, según el procedimiento **R**, son los cuadrados de los módulos de las amplitudes de probabilidad. Solamente *una* de estas alternativas se realiza en el mundo físico real de la experiencia, y esta alternativa particular es la que observa el experimentador. Según el punto de vista que estoy proponiendo, esta etapa ocurre en cuanto la diferencia entre los campos gravitatorios de las diversas alternativas alcance el nivel de un-gravitón.

¿Cuándo sucede esto? Según un cálculo muy grosero ⁶, si hubiera habido simplemente *una* gota esférica completamente uniforme, la etapa de un gravitón se alcanzaría cuando la gota crezca hasta aproximadamente una centésima de m_p , que es una diezmillonésima de gramo. Existen muchas incertidumbres en este cálculo (incluyendo algunas dificultades de principio) y el tamaño es un poco grande para que nos quedemos tranquilos, pero el resultado no es totalmente irra-

zonable. Hay que esperar que más adelante se obtendrán algunos resultados más precisos y que será posible tratar una cadena entera de gotas y no sólo una simple gota. Puede que también haya importantes diferencias cuando se tenga en cuenta el hecho de que las gotas se componen de un número muy grande de átomos diminutos, en lugar de ser totalmente uniformes. Además, el propio criterio de «un-gravitón» necesita hacerse mucho más preciso matemáticamente.

En la situación anterior he considerado que podría haber una observación real de un proceso cuántico (la desintegración de un átomo radiactivo) en el que los efectos cuánticos se han amplificado hasta el punto en que las diferentes alternativas cuánticas producen diferentes, y directamente observables, alternativas macroscópicas. Mi idea sería que **R** puede tener lugar objetivamente incluso cuando no está *manifiestamente* presente tal amplificación. Supongamos que, en lugar de entrar en una cámara de niebla, nuestra partícula entra simplemente en una gran caja de gas (o fluido) de tal densidad que es prácticamente seguro que chocará, o perturbará de alguna otra manera, con un gran número de átomos del gas. Consideremos sólo dos de las alternativas para la partícula como parte de la superposición lineal compleja inicial: podría no entrar en absoluto en la caja, o podría entrar a lo largo de un camino particular y rebotar en algún átomo del gas. En el segundo caso, dicho átomo de gas saldrá despedido a gran velocidad de una forma en que no lo hubiera hecho si la partícula no hubiera incidido sobre él, y a continuación chocará y él mismo rebotará en algún otro átomo más. Cada uno de los dos átomos se moverán entonces de formas en que no lo hubieran hecho en otro caso, y pronto habrá una cascada de movimiento de átomos en el gas que no habría sucedido si la partícula no hubiera entrado inicialmente en la caja (Fig. 8.8). En este segundo caso no transcurriría mucho tiempo hasta que prácticamente todos los átomos en el gas hubieran sido perturbado por este movimiento.

Pensemos ahora en cómo deberíamos describir esto mecano-cuánticamente. Inicialmente sólo tenemos la partícula original cuyas diferentes posiciones debe considerarse que ocurren en una superposición lineal compleja —como parte de la función de onda de la partícula. Pero tras un corto intervalo, todos los átomos del gas deben estar involucrados. Consideremos la superposición lineal compleja de dos caminos que hubiera podido tomar la partícula, uno entrando en la caja y el otro no. La mecánica cuántica estándar insiste en que extendamos esta superposición a todos los átomos del gas: debemos

superponer dos estados tales que todos los átomos de gas en uno de los estados están desplazados respecto a sus posiciones en el otro estado. Consideremos ahora la *diferencia* entre los campos gravitatorios de la totalidad de los átomos individuales. Incluso aunque la distribución *global* del gas sea virtualmente la misma en los dos estados a superponer (y los campos gravitatorios globales pudieran ser prácticamente idénticos), si *restamos* un campo de otro obtenemos un campo *diferencia* (fuertemente oscilante) —que podría ser perfectamente «significativo» en el sentido en que lo entiendo aquí— y puede que se supere fácilmente el nivel un-gravitón para este campo *diferencia*. En cuanto se alcanza este nivel tiene lugar la reducción del vector de estado: en el estado *real* del sistema, *o bien* la partícula ha entrado en la caja de gas, *o bien* no lo ha hecho. La superposición

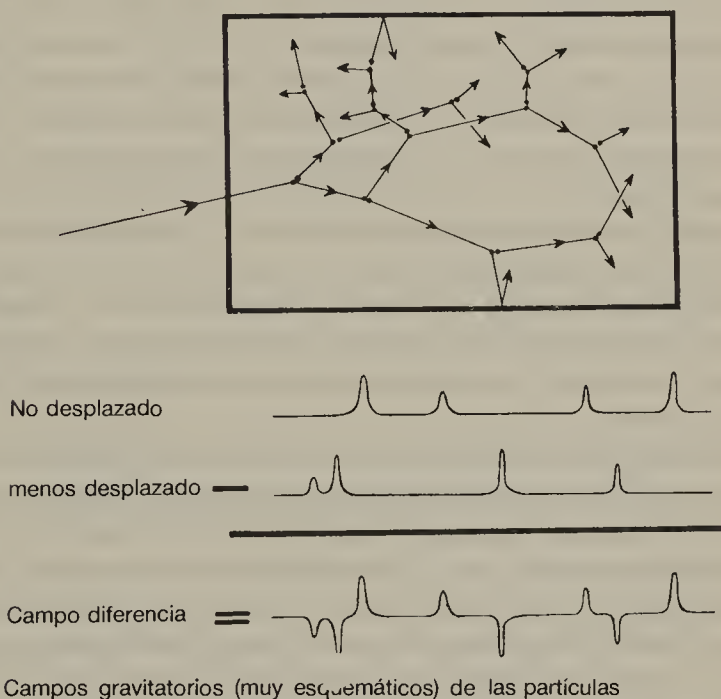


Fig. 8.8. Si una partícula entra en una gran caja llena de algún gas no pasará mucho tiempo hasta que prácticamente todos los átomos del gas hayan sido perturbados. Una superposición lineal cuántica de la partícula entrante, y de la partícula no entrante, implicaría la superposición lineal de dos geometrías espacio-temporales diferentes que describen los campos gravitatorios de las dos configuraciones de las partículas del gas. ¿Cuándo se alcanzará el nivel de un-gravitón para la *diferencia* entre estas geometrías?

lineal compleja se ha reducido a alternativas estadísticamente ponderadas, y sólomente *una* de estas tiene lugar realmente.

En el ejemplo previo consideré una cámara de niebla como una manera de proporcionar una observación mecano-cuántica. Considero muy probable que se puedan tratar otros tipos de observaciones semejantes (placas fotográficas, cámaras de chispas, etc.) utilizando el «criterio de un-gravitón» mediante una aproximación semejante a la que he indicado antes para la caja de gas. Queda mucho trabajo por hacer en ver como podría aplicarse en detalle este procedimiento.

Por ahora éste es sólo el germen de una idea para la que creo muy necesaria una nueva teoría ⁷. Pienso que cualquier esquema completamente satisfactorio tendría que incorporar algunas ideas radicalmente nuevas acerca de la naturaleza de la geometría del espacio-tiempo, incluyendo probablemente una descripción esencialmente no-local ⁸. Una de las razones más imperativas para creer esto procede de los experimentos tipo EPR (cf. pp. 353, 361), en los que una «observación» (en este caso, el registro de una fotocélula) en un extremo de una habitación puede efectuar la reducción *simultánea* del vector de estado en el otro extremo. La construcción de una teoría completamente objetiva de la reducción del vector de estado que sea consistente con el espíritu de la relatividad es un profundo reto, ya que «simultaneidad» es un concepto extraño a la relatividad por ser dependiente del movimiento del observador. Mi opinión es que nuestra imagen actual de la realidad física, especialmente en relación con la naturaleza del *tiempo*, va a sufrir una gran conmoción —incluso mayor, quizá, que la que ya han supuesto hasta ahora la relatividad y la mecánica cuántica.

Debemos volver a nuestra pregunta original. ¿Cómo se relaciona todo esto con la física que gobierna las acciones de nuestros cerebros? ¿Qué tendría que ver con nuestros pensamientos y nuestros sentimientos? Para intentar algún tipo de respuesta será necesario en primer lugar examinar algo de cómo están contruidos realmente nuestros cerebros. Volveré después a la que creo que es la pregunta, fundamental: ¿qué tipo de acción *física* nueva está probablemente implícita cuando pensamos o percibimos conscientemente?

9. Cerebros reales y modelos de cerebro

¿Cómo son realmente los cerebros?

En el interior de nuestra cabezas hay una magnífica estructura que controla nuestras acciones y de algún modo da lugar a una consciencia del mundo que nos rodea. Pero, como Alang Turing dijo en cierta ocasión ¹, ¿no hay nada más parecido a un puchero de potaje! Es difícil ver cómo un objeto de apariencia tan poco prometedora pueda lograr los milagros de que le sabemos capaz. Sin embargo, un examen más próximo comienza a revelar que el cerebro tiene una estructura mucho más intrincada y una sofisticada organización (Fig. 9.1). La parte superior, con más circunvoluciones (y la más parecida al potaje), es el *cerebro* propiamente dicho. Está claramente dividido por la mitad en los *hemisferios cerebrales* izquierdo y derecho; y, de una manera bastante menos tajante, en una zona delantera con el lóbulo frontal y una zona trasera con otros tres lóbulos: el parietal, el temporal y el occipital. Debajo y en la parte de atrás hay una porción bastante más pequeña y algo esférica (parecida quizá a dos ovillos de lana): el *cerebelo*. Más en el interior, y parcialmente ocultas bajo el cerebelo, hay varias estructuras curiosas y de apariencia complicada: el puente y la médula (incluyendo la formación reticular, una región que nos interesará más adelante) que constituyen el tronco cerebral, tálamo, hipotálamo, hipocampo, cuerpo calloso, y muchas otras construcciones extrañas y de nombres singulares.

El cerebro * propiamente dicho es la parte de la que los seres

* A lo largo de este capítulo habrá que tener en cuenta la distinción más precisa entre cerebro global, o encéfalo, y cerebro propiamente dicho, que es una parte del anterior. [Nota del traductor.]

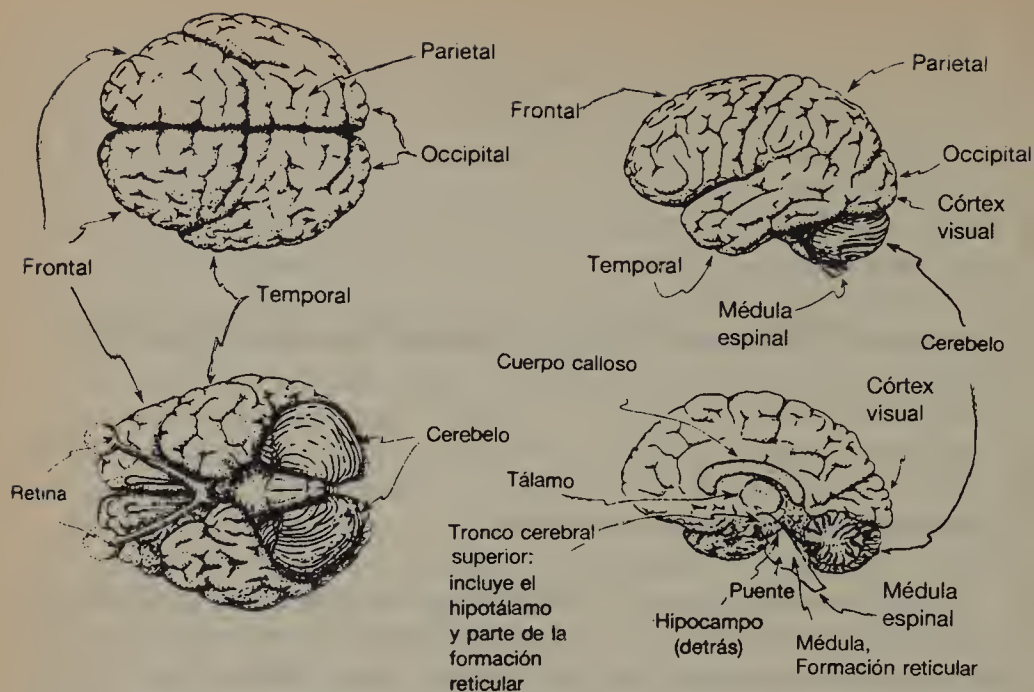


Fig. 9.1. El cerebro humano: vistas superior, lateral, inferior y sección central.

humanos se sienten más orgullosos —pues no sólo es la parte más grande del cerebro humano sino que es también mayor, en proporción al encéfalo en conjunto, en el *hombre* que en los otros animales. (También el *cerebelo* es mayor en el hombre que en la mayoría de los animales.) El cerebro y el cerebelo tienen capas superficiales externas relativamente delgadas de *substancia gris* y regiones internas mayores de *substancia blanca*. Estas regiones de substancia gris se denominan respectivamente *corteza cerebral* y *corteza cerebelar*. La substancia gris es en donde parece que se ejecutan los diversos tipos de tareas computacionales, mientras que la substancia blanca consiste en largas fibras nerviosas que transportan señales de una parte del cerebro a otra.

Diversas partes de la corteza cerebral están asociadas con funciones específicas. El *córtez visual* es una región en el interior del lóbulo occipital, justo en la parte trasera del cerebro, que está relacionada con la recepción e interpretación de la visión. Es curioso que la Naturaleza escogiera esta región para reinterpretar las señales procedentes de los ojos que, al menos en el hombre, están situados ¡justo en la parte frontal de la cabeza! Pero la Naturaleza se comporta de forma

aún más curiosa que ésta. Es el hemisferio cerebral *derecho* el que está relacionado casi exclusivamente con el lado *izquierdo* del cuerpo, mientras que el hemisferio cerebral izquierdo está relacionado con el lado derecho del cuerpo —¡de modo que prácticamente todos los nervios deben cruzar de un lado a otro cuando entran o salen del cerebro! En el caso del córtex visual, no se trata simplemente de que el lado derecho esté asociado al ojo izquierdo sino que está asociado con *el lado izquierdo del campo visual de ambos ojos*. Análogamente, el córtex visual izquierdo está asociado con el lado derecho del campo visual de ambos ojos. Esto significa que los nervios que salen del lado derecho de la retina de cada ojo deben ir al córtex visual derecho (recuérdese que la imagen en la retina está invertida) y que los nervios que salen del lado izquierdo de la retina de cada ojo deben ir al córtex visual izquierdo. (Ver Fig. 9.2) De este modo se forma un mapa muy bien definido del lado izquierdo del campo visual en el córtex visual derecho y se forma otro mapa del lado derecho del campo visual en el córtex visual izquierdo.

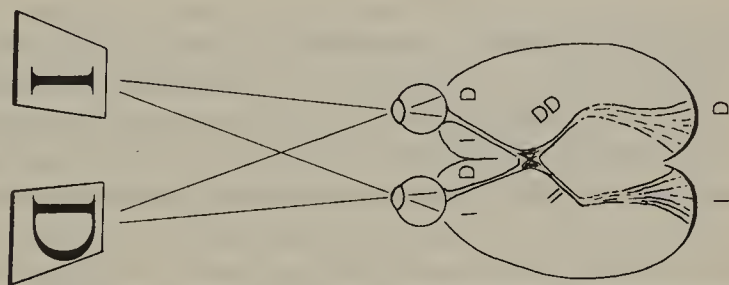


Fig. 9.2. El lado izquierdo del campo visual de ambos ojos se proyecta en el córtex visual derecho, y el lado derecho del campo visual se proyecta en el córtex visual izquierdo. (Vista inferior: nótese que las imágenes en la retina están invertidas.

Las señales de los oídos también cruzan al lado contrario del cerebro de esta curiosa manera. El córtex auditivo derecho (parte del lóbulo temporal derecho) procesa principalmente el sonido recibido desde la izquierda, y el córtex auditivo izquierdo, en general, los sonidos que proceden de la derecha. El olfato parece una excepción a las reglas generales. El córtex olfativo derecho, situado en la parte frontal del cerebro (en el lóbulo frontal —lo que es excepcional para un área sensorial), está relacionado principalmente con la ventana derecha de la nariz y el izquierdo, con la ventana izquierda.

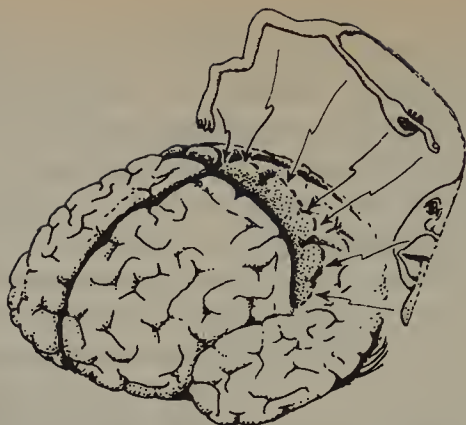


Fig. 9.3. El «homúnculo somatosensorial» ilustra gráficamente las porciones del cerebro —inmediatamente detrás de la división entre los lóbulos frontal y parietal— que están más directamente interesadas en el sentido del tacto para las diversas partes del cuerpo.

Las sensaciones del *tacto* tienen que ver con la región del lóbulo parietal denominada *córtex somatosensorial*. Esta región está exactamente detrás de la división entre los lóbulos frontal y parietal. Existe una correspondencia muy concreta entre las diversas partes de la superficie del cuerpo y las regiones del córtex somatosensorial. Esta correspondencia se ilustra a veces gráficamente en términos de lo que se denomina el «homúnculo somatosensorial», que es una figura humana distorsionada que se representa yaciendo a lo largo del córtex somatosensorial como en la Figura 9.3. El córtex somatosensorial derecho trata las sensaciones del lado izquierdo del cuerpo, y el izquierdo, las del derecho. Existe una región análoga en el lóbulo *frontal*, situada justo *delante* de la división entre el lóbulo frontal y el parietal, conocida como *córtex motor*. Este está relacionada con la activación del *movimiento* de las diferentes partes del cuerpo y de nuevo existe una correspondencia muy específica entre los diversos músculos del cuerpo y las regiones del córtex motor. Ahora tenemos un «homúnculo motor» para representar esta correspondencia, como se muestra en la Figura 9.4. El córtex motor derecho controla el lado izquierdo del cuerpo, y el córtex motor izquierdo controla el lado derecho.

Las regiones de la corteza cerebral recién mencionadas (los córtex visual, auditivo, olfativo, somatosensorial y motor) se llaman *primarias* puesto que son las más directamente relacionadas con el input y el output del cerebro. Próximas a estas regiones primarias están las

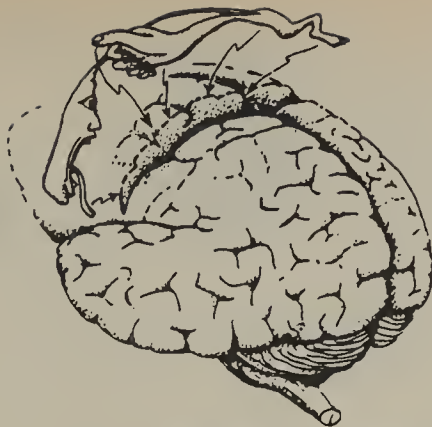


Fig. 9.4. El «homúnculo motor» ilustra las porciones del cerebro —inmediatamente delante de la división entre los lóbulos frontal y parietal— que más directamente activan los movimientos de las diversas partes del cuerpo.

regiones *secundarias* de la corteza cerebral, que están relacionadas con un nivel de abstracción más sutil y complejo. (Ver Fig. 9.5.) La información sensorial recibida por los córtex visual, auditivo y somatosensorial se procesa en las regiones secundarias asociadas, y la región motora secundaria está relacionada con los planes concebidos del movimiento que son traducidos por el córtex motor primario en instrucciones más específicas para los movimientos musculares reales. (Dejemos aparte el córtex olfativo en nuestras consideraciones, ya que se comporta de modo diferente y parece que se conoce más bien poco sobre él.) Las restantes regiones de la corteza cerebral se denominan *terciarias* (o córtex *de asociación*). Es en estas regiones terciarias en donde se lleva a cabo fundamentalmente la actividad más abstracta y sofisticada del cerebro. Aquí es —en conjunción, en alguna medida, con la periferia— donde se mezcla y analiza la información de las diversas regiones sensoriales de una manera muy compleja, donde reside la memoria, se construyen las imágenes del mundo externo, se conciben y evalúan planes generales, y se entiende o se formula el habla.

El habla es particularmente interesante ya que se suele considerar como algo muy específico de la inteligencia humana. Es curioso que (al menos en la inmensa mayoría de las personas diestras y en la mayor parte de las personas zurdas) los principales centros del *habla* están precisamente en el *lado izquierdo* del cerebro. Las áreas esen-

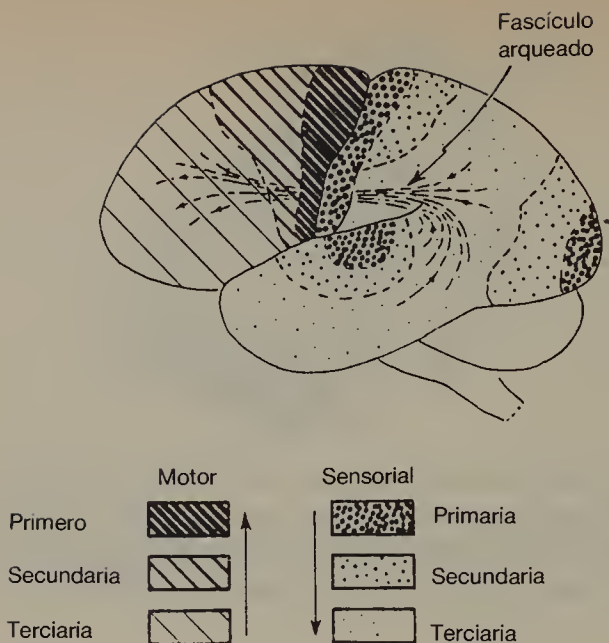


Fig. 9.5. La acción del cerebro en líneas muy generales. Los datos externos de los sentidos entran en las regiones sensoriales primarias, son procesados con sucesivos grados de sofisticación en la regiones sensoriales secundarias y terciarias, transferidos a la región motora terciaria y, finalmente, son transformados en instrucciones específicas para el movimiento en las regiones motoras primarias.

ciales son el *área de Broca*, una región en la parte inferior trasera del lóbulo frontal, y otra llamada *área de Wernicke*, dentro y también alrededor de la parte superior trasera del lóbulo temporal (ver Fig. 9.6). El área de Broca está relacionada con la formulación de enunciados, y el área de Wernicke con la comprensión del lenguaje. Las lesiones en el área de Broca dificultan el habla pero dejan intacta la comprensión, mientras que con el área de Wernicke lesionada el habla es fluida pero con poco contenido. Un haz nervioso llamado *fascículo arqueado* conecta las dos áreas. Cuando éste se lesiona no queda impedida la comprensión y el habla sigue siendo fluida, pero la comprensión no puede ser expresada verbalmente.

Podemos formarnos ahora una imagen muy grosera de lo que hace el cerebro. El *input* del cerebro procede de las señales visuales, auditivas, olfativas, táctiles y otras que se registran inicialmente en las porciones *primarias* de (principalmente) los lóbulos *posteriores* (parie-

tal, temporal y occipital). El *output* del cerebro, en forma de activación de movimientos corporales, se produce principalmente en las porciones primarias de los lóbulos *frontales* del cerebro. Entre ambos tiene lugar algún tipo de procesamiento. De un modo general, existe un movimiento en la actividad cerebral que comienza en las porciones primarias de los lóbulos posteriores, se desplaza a las porciones secundarias a medida que se van analizando los datos de entrada, y continúa hacia las porciones terciarias de los lóbulos posteriores a medida que estos datos se van comprendiendo completamente (v.g. como sucede con la comprensión del habla en el área de Wernicke). El fascículo arqueado —el haz de fibras nerviosas antes mencionado, pero ahora en ambos lados del cerebro— lleva entonces esta información procesada al lóbulo frontal, en cuyas regiones terciarias se formulan planes generales de actuación (v.g. como la formulación del habla en el área de Broca). Estos planes generales de actuación se traducen en concepciones más concretas sobre movimientos corporales en el córtex motor primario desde donde finalmente se envían las señales a los diversos grupos de músculos en el cuerpo (y a menudo a varios a la vez).

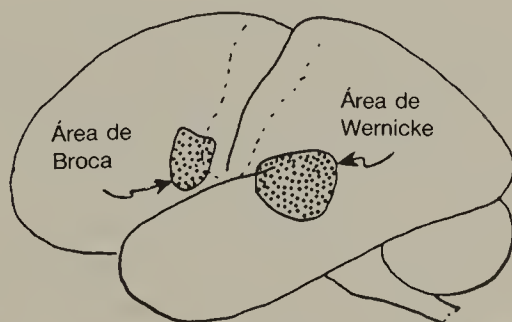


Fig. 9.6. Normalmente sólo en el lado izquierdo: el área de Wernicke está relacionada con la comprensión y el área de Broca lo está con la formulación del habla.

La imagen de un soberbio dispositivo de computación parece presentarse ante nosotros. Los defensores de la IA fuerte (cf. Capítulo 1, etc.) mantendrán que aquí tenemos un ejemplo supremo de un computador algorítmico —una máquina de Turing en efecto— en donde hay un input (como la cinta input a la izquierda de una máquina de Turing) y un output (como la cinta output a la derecha de la

máquina), y entre medias se realizan todo tipo de computaciones complicadas. Por supuesto, la actividad del cerebro puede llevarse a cabo también independientemente de cualquier input sensorial. Esto sucede cuando simplemente pensamos, calculamos o meditamos sobre recuerdos del pasado. Para los defensores de la IA fuerte estas actividades del cerebro serían simplemente actividad algorítmica adicional, y ellos podrían sugerir que el fenómeno de la «consciencia» aparece allí donde semejante actividad interna alcanza un nivel de sofisticación suficiente.

Sin embargo, no debemos conformarnos con estas rápidas explicaciones. La imagen general de la actividad del cerebro presentada arriba es solamente una imagen bastante tosca. En primer lugar, incluso la recepción de la visión no es tan sencilla como la he presentado. Parece haber varias regiones diferentes (aunque más pequeñas) de la corteza en donde se hacen mapas del campo visual, aparentemente con otros diversos propósitos. (Nuestra *consciencia* de la visión parece diferir con respecto a ellos.) Parece que hay también otras regiones sensoriales y motoras diseminadas por la corteza cerebral (por ejemplo, los movimientos del ojo pueden ser activados por varios puntos en los lóbulos *posteriores*).

En mis descripciones anteriores no he mencionado siquiera el papel de las partes del encéfalo distintas del cerebro propiamente dicho. ¿Cuál es, por ejemplo, el papel del *cerebelo*? Aparentemente es el responsable de una precisa coordinación y control del cuerpo —su ritmo, equilibrio y delicadeza de movimientos. Imaginemos el arte fluido de un bailarín, la fácil precisión de un jugador de tenis profesional, el rapidísimo control de un piloto de carreras, y los movimientos seguros de las manos de un pintor o un intérprete musical; imaginemos también los gráciles saltos de una gacela y el andar sigiloso de un gato. Tal precisión no sería posible sin el cerebelo, y todos los movimientos se volverían titubeantes y torpes. Parece que cuando se está aprendiendo una nueva habilidad, ya sea andar o conducir un coche, se debe planear inicialmente cada acción en detalle y es el cerebro quien controla; pero cuando ya ha sido dominada esta habilidad —y se ha convertido en una «segunda naturaleza»— es el cerebelo el que asume el mando. Además, resulta una experiencia familiar que si *pensamos* nuestras acciones en una habilidad que ya se domina entonces podemos perder momentáneamente el control. La *reflexión* parece implicar la re-introducción del control cerebral y, aunque así se introduce una consiguiente flexibilidad en la actividad, se pierde la acción cerebelar precisa y fluida. Tales descripciones están sin duda

muy simplificadas pero dan una impresión razonable del papel del cerebelo *.

Resultaba también confuso, en mis anteriores descripciones de la acción del cerebro, el dejar de lado cualquier información sobre las demás partes del encéfalo. Por ejemplo, el *hipocampo* juega un papel vital al asentar la memoria a largo plazo (permanente), siendo almacenada la memoria real en alguna parte de la corteza cerebral —probablemente en varios lugares a la vez. El cerebro puede retener también imágenes a *corto* plazo de otras maneras; y puede mantenerlas durante algunos minutos o incluso horas (quizá «conservándolas en la mente»). Pero para poder recordar estas imágenes después de que hayamos dejado de prestarles atención es necesario que queden asentadas de forma permanente, y para esto es esencial el hipocampo. (Las lesiones en el hipocampo provocan una temible condición en la que no se retienen nuevos recuerdos una vez que han dejado la atención del sujeto.) El *cuerpo calloso* es la región a través de la que se comunican entre sí los hemisferios cerebrales izquierdo y derecho. (Veremos más adelante algunas de las sorprendentes consecuencias de tener el cuerpo calloso seccionado.) El *hipotálamo* es la sede de las emociones —placer, rabia, miedo, desesperación, hambre— y sirve de medio a las manifestaciones tanto mentales como físicas de las emociones. Hay un flujo continuo de señales entre el hipotálamo y las diferentes partes del cerebro. El *tálamo* actúa como un importante centro de procesamiento y estación repetidora, y transmite muchos de los impulsos nerviosos desde el mundo externo a la corteza cerebral. La *formación reticular* es responsable del estado general de alerta o consciencia implicado en el encéfalo como un todo o en las distintas partes del encéfalo. Hay numerosos caminos para los nervios que conectan estas, y muchas otras, áreas de importancia vital.

La descripción anterior da sólo una muestra de las partes más importantes del encéfalo. Concluiré la sección diciendo algo más sobre la organización del encéfalo en general. Sus diferentes partes se clasifican en tres regiones que, tomadas en orden según se alejan de la médula espinal, se llaman *cerebro posterior* (o rombencéfalo), el *cerebro medio* (o mesencéfalo) y el *cerebro anterior* (o prosencéfalo). En el desarrollo temprano del embrión se encuentran estas tres regiones, en este orden, como tres abultamientos al final de la médula

* Curiosamente, el comportamiento «cruzado» del cerebro no se aplica al cerebelo, de modo que la mitad derecha del cerebelo controla principalmente el lado *derecho* del cuerpo, y la mitad izquierda controla el lado *izquierdo* del cuerpo.

espinal. De la que está situada en la parte más extrema, el cerebro anterior en fase de desarrollo, brotan dos yemas, una a cada lado, que se convertirán en los hemisferios cerebrales. El cerebro anterior completamente desarrollado incluye muchas partes importantes del encéfalo —no sólo el cerebro propiamente dicho, sino también el cuerpo calloso, tálamo, hipotálamo, hipocampo, y muchas otras partes. El cerebelo es parte del cerebro posterior. La formación reticular tiene una parte en el cerebro medio y otra parte en el cerebro posterior. El cerebro anterior es el «más reciente» en el sentido del desarrollo evolutivo, y el cerebro posterior es el más «antiguo».

Espero que este breve cuadro, aunque impreciso en diversos aspectos, dará al lector alguna idea de cómo es el cerebro humano y qué hace de manera general. Hasta aquí apenas he tocado el tema central de la *consciencia*. Abordaremos este tema a continuación.

¿Dónde está la sede de la consciencia?

Se han expresado muchas opiniones diferentes con respecto a la relación entre el estado del cerebro y el fenómeno de la consciencia. Es notable el escaso consenso de opinión para un fenómeno de tan obvia importancia. Es evidente, sin embargo, que no todas las partes del cerebro están involucradas por igual en su manifestación. Por ejemplo, como se apuntó antes, el cerebelo parece ser mucho más «autómata» que el cerebro propiamente dicho. Las acciones bajo control cerebelar parecen tener lugar casi de forma «autónoma» sin que tengamos que «reflexionar» sobre ellas. Mientras que podemos decidir conscientemente el andar de un lugar a otro, no tenemos consciencia a menudo de los elaborados planes de movimientos musculares detallados que serán necesarios para el movimiento controlado. Lo mismo puede decirse de las acciones reflejas inconscientes, como la de retirar la mano de una estufa caliente, que podría estar mediada no por el cerebro en general sino por la parte superior de la médula espinal. A partir de esto, podemos estar inclinados, al menos, a inferir que es probable que el fenómeno de la consciencia tenga más que ver con la acción del cerebro propiamente dicho que con la del cerebelo o la médula espinal.

Por otra parte, está lejos de ser evidente que la actividad del cerebro deba *siempre* incidir sobre nuestra consciencia. Por ejemplo, como he descrito antes, en la acción normal de caminar, en la que *no* somos conscientes de la actividad detallada de nuestros músculos

y miembros —al ser el control de esta actividad principalmente cerebral (asistido por otras partes del cerebro y la médula espinal)— parece que *también* las regiones motoras primarias deberían estar involucradas. Además, lo mismo sería válido para las regiones sensoriales primarias: podríamos no tener consciencia, en ese momento, de las variaciones de presión en las plantas de nuestros pies cuando caminamos, pero las regiones correspondientes de nuestro córtex somatosensorial estarían siendo activadas continuamente.

De hecho, el distinguido neurocirujano canadiense Wilder Penfield (quien, en los años 40 y 50, fue responsable de gran parte del mapa detallado de las regiones sensorial y motora del cerebro humano) ha argumentado que la consciencia *no* está simplemente asociada a la actividad cerebral. Él sugirió, basado en sus experiencias al realizar numerosas operaciones cerebrales en sujetos conscientes, que cierta región a la que denominó *tronco cerebral superior*, consistente principalmente en el tálamo y el cerebro medio (cf. Penfield y Jasper, 1947) —aunque él tenía en mente sobre todo la formación reticular— debería considerarse, en cierto sentido, como la «sede de la consciencia». El tronco cerebral superior está en comunicación con el cerebro, y Penfield argumentaba que la «atención consciente» o «consciencia de acción voluntaria» aparecería siempre que esta región del tronco cerebral estuviera en comunicación directa con la región apropiada de la corteza cerebral, es decir, la región particular asociada con cualesquiera sensaciones específicas, pensamientos, recuerdos o acciones que sean percibidas o evocadas conscientemente en el momento. Subrayó que aunque él pudiera, por ejemplo, estimular la región del córtex motor del sujeto que provoca el movimiento del brazo derecho (y el brazo derecho se moviera realmente), esto no provocaría que el sujeto *quisiera* mover el brazo derecho. (De hecho, el sujeto podría incluso reaccionar con el brazo izquierdo y detener el movimiento del brazo derecho —¡como en la bien conocida interpretación cinematográfica que Peter Sellers hacía del Dr. Strangelove!) * Penfield sugería que el *deseo* del movimiento podría tener más que ver con el tálamo que con la corteza cerebral. Su idea era que la consciencia es una manifestación de la actividad del tronco cerebral superior pero, puesto que se necesita además que haya algo que sea *consciente de*, no es sólo el tronco cerebral el que está implicado sino también alguna región de la corteza cerebral que esté en ese momento en co-

* La película, dirigida por Stanley Kubrick, se tituló en España «¿Teléfono rojo? Volamos hacia Moscú?» [Nota del traductor.]

municación con el tronco cerebral superior y cuya actividad representa el sujeto (impresión sensorial o recuerdo) o el objeto (acción voluntaria) de dicha consciencia.

Otros neurofisiólogos han argumentado también que la formación reticular, en concreto, podría considerarse la «sede» de la consciencia, si realmente existe tal sede. Después de todo, la formación reticular es responsable del estado general de alerta del cerebro (Moruzzi y Magoun, 1949). Si se lesiona, el resultado es la inconsciencia. Siempre que el cerebro está en un estado consciente de vigilia, la formación reticular está activa; de lo contrario no lo está. Parece haber así una clara asociación entre la actividad de la formación reticular y el estado de una persona que normalmente denominamos «consciente». Sin embargo, la cuestión se complica por el hecho de que en el estado de ensueño, en el que realmente se tiene «consciencia» (en el sentido de tener consciencia del propio sueño), las partes normalmente activas de la formación reticular *no* parecen estar activas. Una cosa que también preocupa a algunas personas, a propósito de asignar semejante honroso status a la formación reticular, es que ésta es, en términos evolutivos, una parte muy *antigua* del cerebro. Si todo lo que se necesita para ser consciente es una formación reticular activa, ¡entonces las ranas, los lagartos e incluso los bacalaos son conscientes!

Personalmente no considero que este último argumento tenga mucha fuerza. ¿Qué evidencia tenemos de que los lagartos o los bacalaos *no* posean alguna forma de consciencia de bajo nivel? ¿Qué derecho tenemos a afirmar, como harían algunos, que los seres humanos son los únicos habitantes de nuestro planeta dotados de una capacidad real de tener «consciencia»? ¿Somos las únicas cosas, entre las criaturas de la Tierra, para quienes es posible «ser»? Lo dudo. Aunque las ranas y los lagartos, y especialmente los bacalaos, no me inspiran una gran convicción cuando los observo, me resulta muy fuerte la impresión de «presencia consciente» cuando miro a un perro o un gato o, especialmente, cuando en el zoológico me mira un simio o un mono. No pido que ellos sientan como yo, ni siquiera que haya mucha sofisticación en su modo de sentir. No pido que sean «autoconscientes» en ningún sentido fuerte (aunque yo admitiría que puede estar presente un elemento de autoconsciencia) *. ¡Todo lo que pido simplemente es que a veces *sientan*! Como en el estado de ensueño,

* Hay alguna evidencia convincente de que al menos los chimpancés son capaces de auto-consciencia, como parecen demostrar los experimentos en los que se permite a los chimpancés que jueguen con espejos cf. Oakley (1985) capítulos 4 y 5.

aceptaría que hay presente alguna forma de consciencia, aunque presumiblemente de un nivel muy bajo. Si partes de la formación reticular son, en cierto sentido, las únicas responsables de la consciencia, entonces deberían estar activas, aunque en un bajo nivel, en el estado de ensueño.

Otro punto de vista (O'Keefe, 1985) afirma que es la acción del *hipocampo* la que más tiene que ver con el estado consciente. Como señalé antes, el hipocampo es crucial para el asentamiento de la memoria a largo plazo. Podría alegarse que el asentamiento de recuerdos permanentes está asociado con la consciencia y, si esto es correcto, el hipocampo jugaría realmente un papel central en el fenómeno del conocimiento consciente.

Otros sostendrán que es la propia corteza cerebral la responsable de la consciencia. Puesto que el cerebro propiamente dicho es el orgullo del hombre (¡aunque los cerebros de los delfines son mayores!) y puesto que las actividades mentales más estrechamente asociadas con la inteligencia parecen ser llevadas a cabo por este cerebro, entonces ¡es ciertamente allí donde reside el alma del hombre! Ésta será presumiblemente la conclusión del punto de vista de la IA fuerte, por ejemplo. Si la «consciencia» es simplemente una característica de la *complejidad* de un algoritmo —o quizá de su «profundidad» o cierto «nivel de sutileza»— entonces, según la idea de la IA fuerte, los algoritmos complicados que ejecuta la corteza cerebral confirmarían a esta región como la más firme candidata a la capacidad de manifestar consciencia.

Muchos filósofos y psicólogos parecen aceptar la idea de que la consciencia humana está muy ligada al *lenguaje* humano. Por consiguiente, es sólo en virtud de nuestras capacidades lingüísticas por lo que podemos alcanzar una sutileza de pensamiento que es la impronta misma de nuestra humanidad —y la expresión de nuestras propias almas. Es el lenguaje, según este punto de vista, el que nos distingue de los otros animales, y nos proporciona así una excusa para privarles de su libertad y sacrificarlos cuando sentimos que surge dicha necesidad. Es el lenguaje el que nos permite filosofar y describir cómo sentimos, de modo que podamos convencer a los demás de que *nosotros* tenemos consciencia del mundo exterior y también tenemos consciencia de nosotros mismos. Desde este punto de vista, nuestro lenguaje se considera como ingrediente clave de nuestra posesión de consciencia.

Ahora bien, debemos recordar que nuestros centros del lenguaje están (en la inmensa mayoría de las personas) solamente en los lados

izquierdos de nuestros cerebros (áreas de Broca y de Wernicke). El punto de vista recién expresado parecería implicar que la consciencia es algo que está asociado solamente con la corteza cerebral izquierda ¡y no con la derecha! De hecho, ésta resulta ser la opinión de varios neurofisiólogos (en particular, John Eccles, 1973) aunque para mí, como profano, me parece realmente una idea muy extraña por las razones que voy a explicar.

Experimentos de escisión cerebral

En relación con esto mencionaré un importante conjunto de observaciones concernientes a sujetos humanos (y animales) a los que se ha seccionado completamente el cuerpo calloso, de modo que los hemisferios izquierdo y derecho de la corteza cerebral no pueden comunicar entre sí. En el caso de los humanos ², el seccionamiento del cuerpo calloso fue realizado como operación terapéutica al descubrirse que éste era un tratamiento efectivo para una forma de epilepsia particularmente grave que sufrían estos sujetos. Numerosos tests psicológicos les fueron suministrados por Roger Sperry y sus colaboradores algún tiempo después de que hubieran sufrido estas operaciones. Los sujetos estaban colocados de tal forma que se les presentaban estímulos completamente separados para los campos izquierdo y derecho de visión, de modo que el hemisferio izquierdo sólo recibía información visual de lo que se mostraba en el lado derecho, y el hemisferio derecho, sólo de lo del lado izquierdo. Si se le mostraba brevemente una foto de un lápiz en el lado derecho y una foto de una copa en el izquierdo, el sujeto diría (oralmente) «Eso es un *lápiz*», ya que era el lápiz, y no la copa, lo percibido por el único lado del cerebro aparentemente capaz de hablar. Sin embargo, el lado izquierdo sería capaz de seleccionar un plato, antes que un trozo de papel, como el objeto apropiado para asociar con la copa. El lado izquierdo estaría bajo el control del hemisferio derecho y, aunque incapaz de hablar, este hemisferio derecho ejecutaría ciertas acciones bastante complejas y decididamente humanas. De hecho, se ha sugerido que el *pensamiento geométrico* (especialmente en tres dimensiones), y también la música, pueden ser llevados a cabo normalmente en el hemisferio *derecho* principalmente, para compensar las capacidades verbales y analíticas del izquierdo. El cerebro derecho puede comprender los nombres comunes o las frases elementales, y puede llevar a cabo operaciones aritméticas muy sencillas.

Lo más sorprendente de estos sujetos con cerebro escindido es que los dos lados parecen comportarse como individuos prácticamente independientes, cada uno de los cuales puede comunicar por separado con el experimentador —aunque la comunicación es más difícil, y a un nivel más primitivo, con el hemisferio derecho que con el izquierdo, debido a la falta de capacidad verbal del derecho. Una mitad del cerebro del sujeto puede comunicar con la otra de manera simple, v.g. observando el movimiento del brazo controlado por el otro lado, o quizá oyendo sonidos reveladores (como el repiqueteo en un plato). Pero incluso esta comunicación primitiva entre los dos lados puede eliminarse en condiciones de laboratorio cuidadosamente controladas. Pese a todo, todavía pueden pasar de un lado a otro vagos sentimientos emocionales debido, presumiblemente, a que estructuras que no están seccionadas, tales como el hipotálamo, siguen estando en comunicación con ambos lados.

Resulta tentador plantear la cuestión: ¿Tenemos dos individuos conscientes separados que habitan en el mismo cuerpo? Esta cuestión ha sido objeto de mucha controversia. Algunos mantendrán que la respuesta debe ser ciertamente «sí», mientras otros afirmarán que ningún lado debe considerarse como un individuo por propio derecho. Algunos estarán de acuerdo en que el hecho de que puede haber sentimientos emocionales comunes a los dos lados pone en evidencia que todavía hay un solo individuo simple involucrado. Pero otro punto de vista es que sólo el hemisferio del lado *izquierdo* representa un individuo consciente, y el del lado derecho es un autómatas. Parece que esta idea es la que aceptan quienes sostienen que el lenguaje es un ingrediente esencial de la consciencia. De hecho, sólo el hemisferio izquierdo puede responder convicentemente «sí» a la pregunta oral «¿Es usted consciente?» El hemisferio derecho, como un perro, un gato o un chimpancé, podría ser puesto en dificultades incluso para descifrar las palabras que constituyen la pregunta, y puede ser completamente incapaz de expresar verbalmente su respuesta.

Pero el tema no puede despacharse tan a la ligera. En un experimento más reciente de gran interés, Donald Wilson y sus colaboradores (Wilson *et al.*, 1977; Gazzaniga, LeDoux y Wilson, 1977) examinaron un sujeto con cerebro escindido, a quien denominaban «P.S.». Después de la operación de seccionamiento sólo el hemisferio izquierdo podía hablar pero *ambos* hemisferios podían comprender el habla; posteriormente, el hemisferio derecho ¡aprendía también a hablar! Evidentemente *ambos* hemisferios eran conscientes. Además, parecían ser conscientes *por separado*, ya que tenían diferentes gustos y

deseos. Por ejemplo, el hemisferio izquierdo describía que su deseo era ser un dibujante y el derecho, ¡un piloto de carreras!

Por mi parte, simplemente no puedo creer la afirmación común de que el lenguaje humano ordinario es necesario para el pensamiento o para la consciencia. (Trataré de dar algunas de mis razones en el próximo capítulo.) Por lo tanto, me pongo de lado de quienes creen, en general, que las dos mitades de un sujeto con cerebro escindido pueden ser conscientes independientemente. El ejemplo de P.S. sugiere con fuerza que, al menos en este caso particular, ambas mitades pueden serlo realmente. En mi propia opinión la única diferencia real, a este respecto, entre P.S. y los demás es que ¡su consciencia del lado derecho puede convencer realmente a los demás de su existencia!

Si aceptamos que P.S. tiene realmente dos mentes independientes, entonces se nos presenta una curiosa situación. Presumiblemente *antes* de la operación cada sujeto con cerebro escindido poseía una sola consciencia; ¡pero después hay dos! De algún modo, la simple consciencia original se ha *bifurcado*. Podemos recordar al hipotético viajero del Capítulo 1 (p. 53) que se sometía voluntariamente a la máquina teleportadora, y que (de forma imprevista) se despertaba para descubrir que su yo supuestamente «real» había llegado a Venus. En aquel caso, la bifurcación de su consciencia parecía proporcionar una paradoja. En efecto, podemos preguntar: «¿Qué ruta siguió “realmente” su corriente de consciencia?» Si *usted* fuera el viajero, ¿cuál de ellos terminaría siendo «usted»? La máquina teleportadora podía despacharse como cosa de ciencia ficción, pero en el caso de P.S. parece que tenemos algo aparentemente análogo pero que *¡ha sucedido realmente!* ¿Cuál de las consciencias de P.S. «es» el P.S. de antes de la operación? Sin duda muchos filósofos desdeñarían la pregunta como sin significado. Pero no parece haber modo operacional de decidir la cuestión. Cada hemisferio compartirá recuerdos de una existencia consciente antes de la operación y, sin duda, ambos afirmarán ser esa persona. Esto puede ser curioso pero no es en sí mismo una paradoja. De todas formas, aún nos sigue quedando algo de enigmático.

El enigma sería aún más exacerbado si las dos consciencias pudieran de algún modo ser unidas de nuevo con posterioridad. El reempalmar los nervios individuales seccionados del cuerpo calloso parece imposible con la tecnología actual, pero podríamos concebir algo menos drástico que el seccionamiento real de las fibras nerviosas como situación inicial. Quizá estas fibras podrían ser congeladas temporalmente, o paralizadas con alguna droga. No tengo noticia de que se haya llevado a cabo un experimento semejante, pero supongo que

podría ser técnicamente factible en un plazo no muy largo. Presumiblemente, después de que el cuerpo calloso se haya reactivado sólo una consciencia resultaría. ¡Imagínese que esta consciencia es usted! ¿Cómo se sentiría después de haber sido dos personas separadas con distintos «yos» en algún momento en el pasado?

Ceguera cortical

Los experimentos de escisión cerebral parecen indicar al menos que no debe haber necesariamente una única «sede de la consciencia». Pero hay otros experimentos que parecen sugerir que algunas partes de la corteza cerebral están más asociadas a la consciencia que otras. Uno de éstos tiene que ver con el fenómeno de la *ceguera cortical*. Lesiones en una región de corteza visual pueden provocar ceguera en el campo de visión correspondiente. Si se mantiene un objeto en dicha región del campo visual, entonces el objeto no será percibido. Se tiene ceguera con respecto a esa región de visión.

Sin embargo, algunos curiosos descubrimientos (cf. Weiskrantz, 1987) indican que las cosas no son tan simples como eso. A un paciente, denominado «D.B.» hubo que extirparle parte de su córtex visual y esto le provocó la incapacidad de ver nada en una cierta región del campo visual. Sin embargo, cuando se colocaba algo en esta región y se le pedía a D.B. que *imaginara* qué cosa era (normalmente una marca como una cruz o un círculo, o un segmento rectilíneo inclinado un cierto ángulo), él descubrió que podía hacerlo ¡con casi un 100 por cien de exactitud! La precisión de estas «conjeturas» resultó una sorpresa para el propio D.B., que seguía manteniendo que no podía percibir nada en dicha región *.

Las imágenes recibidas en la retina son procesadas también en regiones del cerebro *distintas* de la corteza visual, y una de las regiones más oscuras involucradas está en la parte inferior del lóbulo temporal. Parece que D.B. podía estar basando sus «conjeturas» en información obtenida por esta región temporal inferior. Nada era directamente percibido *conscientemente* mediante la activación de estas regiones, pero la información estaba allí, para ser revelada sólo en la

* Algo en cierto modo complementario de la ceguera cortical es una situación conocida como «ceguera negada», según la cual un sujeto que de hecho es totalmente ciego insiste en que es capaz de ver bastante bien, ¡dando la impresión de ser *visualmente* consciente de los entornos *inferidos*! (Véase Churchland, 1984, p. 143.)

exactitud de las «conjeturas» de D.B. De hecho, después de algún entrenamiento, D.B. fue capaz de obtener una cantidad limitada de consciencia real respecto a estas regiones.

Todo esto parece probar que algunas áreas de la corteza cerebral (v.g. el córtex visual) están más asociadas con la percepción consciente que lo están otras áreas, pero que, con entrenamiento, algunas de estas otras áreas pueden aparentemente ser sacadas a la luz al alcance de la consciencia directa.

Procesamiento de información en el córtex visual

El *córtex visual* es, más que cualquier otra parte del cerebro, la región de la que mejor se conoce como trata la información que recibe, y se han propuesto varios modelos para dar cuenta de esta acción. En realidad, algún procesamiento de información visual tiene lugar en la misma retina *antes* de que se alcance el córtex visual. (¡La retina se considera realmente como una parte del cerebro!) Uno de los primeros experimentos que apuntaba cómo se lleva a cabo el procesamiento en el córtex visual fue el que les valió el Premio Nobel de 1981 a David Hubel y Torsten Wiesel. En sus experimentos fueron capaces de demostrar que ciertas células del córtex visual del gato respondían a líneas en el campo visual que tenían un *ángulo de inclinación particular*. Otras células próximas respondían a líneas con diferentes ángulos de inclinación. A menudo no importaba qué era lo que formaba este ángulo. Podía ser una línea que señalaba la frontera entre oscuridad y luz o, por el contrario, entre luz y oscuridad, o simplemente una línea oscura sobre un fondo claro. Sólo la característica del «ángulo de inclinación» había sido abstraída por las células concretas que se estaban examinando. Pero otras células respondían a colores concretos, o a las diferencias entre lo que percibe cada ojo de modo que se pueda obtener la percepción de profundidad. A medida que nos alejamos de las regiones de recepción primaria encontramos células que son sensibles a aspectos cada vez más sutiles de nuestra percepción de lo que vemos. Por ejemplo, la imagen de un triángulo blanco completo es percibida cuando miramos el dibujo de la Figura 9.7; pero las líneas que forman el propio triángulo no están en realidad presentes en la figura sino que son *inferidas*. ¡Se han encontrado, en efecto, células en el córtex visual (en lo que se llama córtex visual secundario) que pueden registrar las posiciones de estas líneas inferidas!



Fig. 9.7. ¿Puede ver un triángulo blanco superpuesto a otro triángulo que está sostenido por un anillo? (El contorno del triángulo blanco no está dibujado pero aún así existen células en el cerebro que responden a estas líneas invisibles aunque percibidas.)

Hubo diversas afirmaciones en la literatura ⁴, a comienzos de los años 70, del descubrimiento de una célula en el córtex visual del mono que respondía sólo cuando se registraba en la retina la imagen de un *rostro*. Basada en esta información se formuló la «hipótesis de la célula de la abuela», según la cual habría ciertas células en el cerebro que responderían ¡sólo cuando la abuela del sujeto entraba en la habitación! De hecho, hay descubrimientos recientes que indican que ciertas células responden solamente a palabras concretas. ¿Quizá esto vaya en el camino de la verificación de la hipótesis de la célula de la abuela?

Evidentemente hay mucho que aprender sobre el procesamiento detallado que lleva a cabo el cerebro. Se conoce muy poco, por el momento, sobre cómo llevan a cabo sus tareas los centros superiores del cerebro. Dejemos ahora esta cuestión y volvamos nuestra atención a las células reales del cerebro que le capacitan para conseguir estas notables hazañas.

¿Cómo funcionan las señales nerviosas?

Todo el procesamiento que se hace en el cerebro (y también en la médula espinal y en la retina) se logra mediante las notablemente versátiles células del cuerpo denominadas *neuronas*⁵. Tratemos de ver cómo es una neurona. En la Figura 9.8 he dado una imagen de una neurona. Existe un bulbo central, quizá algo parecido a una estrella, a menudo de una forma más bien similar a la de un rábano, llamado *soma*, que contiene el núcleo de la célula. Partiendo de un extremo del soma hay una larga fibra nerviosa —a veces realmente muy larga (con frecuencia llega a tener varios centímetros de longitud, en un ser humano), si se tiene en cuenta que nos estamos refiriendo a una simple célula microscópica— conocida como *axón*. El axón es el «cable» a lo largo del cual se transmite la señal *output* de la célula. El axón tiene varias bifurcaciones de las que pueden salir muchas ramas más pequeñas. Al final de cada una de estas fibras nerviosas resultantes se encuentra un pequeño *botón sináptico*. En el otro extremo del soma, y a menudo ramificándose en todas direcciones a partir de él, existen *dendritas* arboriformes a lo largo de las cuales los datos *input* llegan al soma. (En ocasiones también hay botones sinápticos en las dendritas, dando lo que se llaman sinapsis *dendrodendríticas* entre dendritas. No las tendré en cuenta en mi discusión ya que la complicación que introducen no es esencial.)

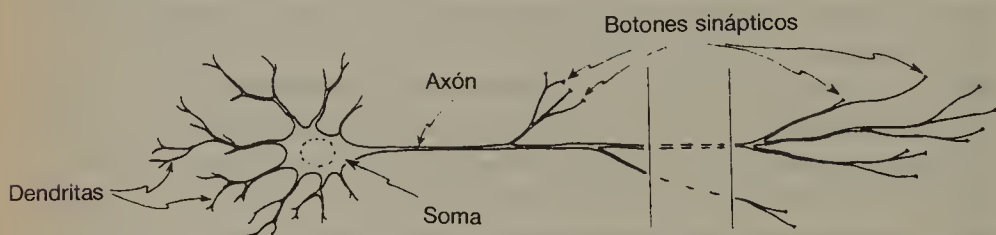


Fig. 9.8. Una neurona (con frecuencia es relativamente mucho más larga de lo que se indica). Hay una gran variedad de apariencias distintas para los diferentes tipos de neuronas.

La célula entera, al ser una unidad autocontenida, tiene una membrana celular que rodea al soma, axón, botones sinápticos, dendritas y todo lo demás. Para que la señales pasen de una neurona a otra es necesario que de algún modo «salten la barrera» entre ellas. Esto se hace en una unión conocida como *sinapsis* en donde un botón sináptico

tico de una neurona se une a un punto de otra neurona, ya sea en el propio soma de la neurona o en una de sus dendritas (ver Fig. 9.9). En realidad hay un espacio de separación muy estrecho entre el botón sináptico y el soma o dendrita a la que se une, llamado *espacio sináptico* (ver Fig. 9.10). La señal se propaga de una neurona a la siguiente a través de esta separación.

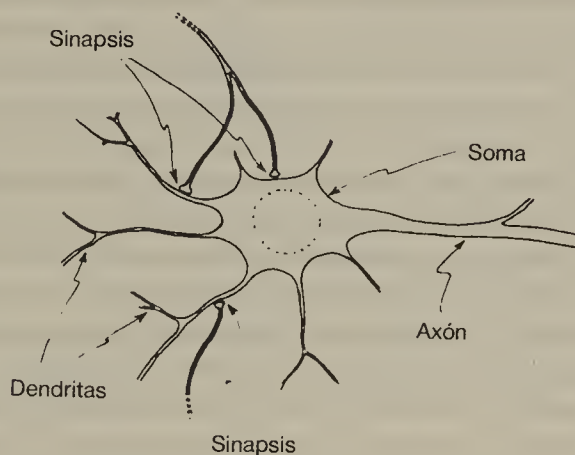


Fig. 9.9. Sinapsis: las uniones entre una neurona y la próxima.

¿Qué forma adoptan las señales cuando se propagan a lo largo de las fibras nerviosas y a través de los espacios sinápticos? ¿Qué es lo que provoca que la próxima neurona emita una señal? Para un profano como yo, los procedimientos que ha adoptado de hecho la Naturaleza parecen extraordinarios —¡y enormemente fascinantes! Podríamos haber pensado que las señales serían precisamente como corrientes eléctricas viajando por cables, pero es mucho más complicado que eso.

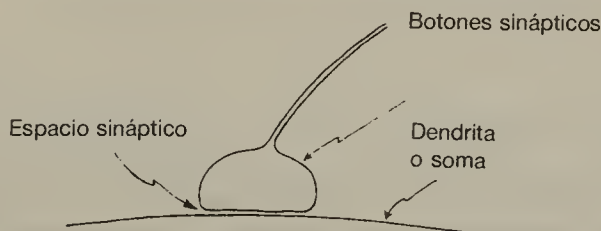


Fig. 9.10. Una sinapsis con más detalle. Hay una estrecha separación a través de la que fluyen los neurotransmisores químicos.

Una fibra nerviosa consiste básicamente en un tubo cilíndrico que contiene una solución mezcla de sal ordinaria (cloruro sódico) y cloruro potásico, principalmente de este último, de modo que hay iones de sodio, potasio y cloro dentro del tubo (Fig. 9.11). Estos iones también están presentes fuera, aunque en proporciones diferentes, de modo que en el exterior hay más iones sodio que iones potasio. En el estado de reposo del nervio existe una carga neta negativa en el interior del tubo (es decir, más iones cloro que iones sodio y potasio juntos —recuérdese que los iones sodio y potasio están cargados positivamente, y los iones cloro cargados negativamente) y una carga neta positiva en el exterior (es decir, más iones sodio y potasio que iones cloro). La membrana celular que constituye la superficie del cilindro es algo «porosa», de modo que los iones tienden a migrar a través de ella y neutralizar la diferencia de carga. Para compensar esto y mantener el exceso de carga negativa en el interior, existe una «bomba metabólica» que bombea iones sodio hacia fuera muy lentamente a través de la membrana circundante. Esto sirve también, en parte, para mantener el exceso de potasio sobre sodio en el interior. Existe otra bomba metabólica que (a una escala algo menor) bombea iones potasio desde el exterior, y que contribuye en consecuencia al exceso de potasio en el interior (aunque trabaja contra el mantenimiento del desequilibrio de carga).

Una *señal* a lo largo de la fibra consiste en una región en la que este desequilibrio de carga está *invertido* (es decir, ahora es positiva en el interior y negativa en el exterior) moviéndose a lo largo de la

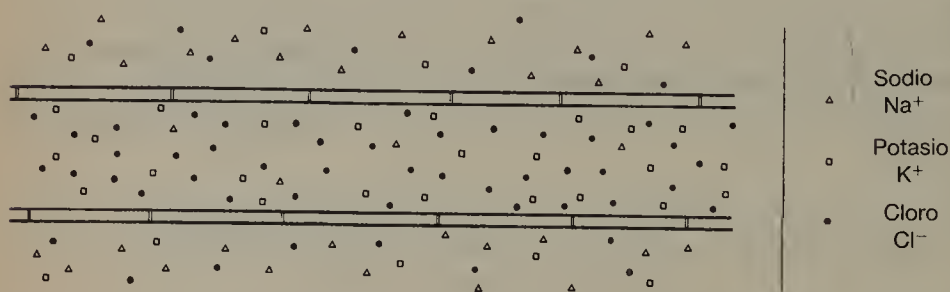


Fig. 9.11. Representación esquemática de una fibra nerviosa. En el estado de reposo existe en el interior un exceso de iones cloro sobre iones sodio y potasio juntos, dando lugar a una carga negativa; en el exterior sucede lo contrario, dando lugar a una carga positiva. El balance sodio/potasio es también diferente en el interior y en el exterior, con más potasio en el interior y más sodio en el exterior.

fibra (Fig. 9.12). Imaginémonos nosotros mismos situados en la fibra nerviosa en un punto por delante de semejante región de inversión de carga. A medida que la región se aproxima, su campo eléctrico provoca que se abran pequeñas «puertas», llamadas *compuertas de sodio*, en la membrana celular; esto permite que los iones de sodio fluyan a través de ellas en dirección contraria (por una combinación de fuerzas eléctricas y presiones debidas a las diferencias de concentración, es decir, «ósmosis») desde el exterior hacia el interior. Esto da como resultado que la carga se vuelve positiva en el interior y negativa en el exterior. Cuando ha sucedido esto, la región con inversión de carga que constituye la señal nos ha alcanzado. Esto provoca ahora que se abra otro conjunto de pequeñas «puertas» (*compuertas de potasio*) que permiten que los iones potasio fluyan desde el interior y comience a restaurarse así el exceso de carga negativa en el interior. ¡La señal ha pasado! Finalmente, a medida que la señal se aleja de nuevo, la lenta pero incesante acción de las bombas empuja otra vez hacia afuera a los iones sodio y hacia adentro a los iones potasio. Esto restaura el estado de reposo de la fibra nerviosa y la deja lista para otra señal.

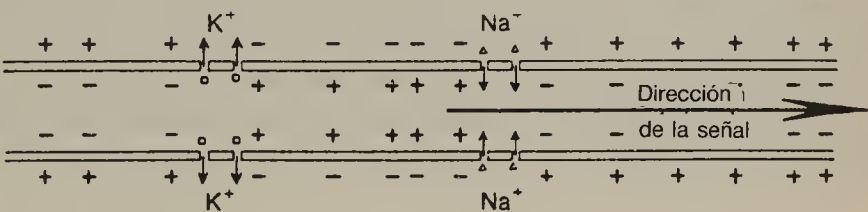


Fig. 9.12. Una señal nerviosa es una región con inversión de carga que viaja a lo largo de la fibra. En su parte delantera las compuertas de sodio se abren para permitir el paso del sodio hacia el interior; y en su parte trasera, las compuertas de potasio se abren para permitir que el potasio salga hacia afuera. Las bombas metabólicas actúan para restaurar el *status quo*.

Nótese que la señal consiste simplemente en una región con inversión de carga que se mueve a lo largo de la fibra. El *material* real (es decir, los iones) apenas se mueven: ¡sólo cruza la membrana celular en una u otra dirección!

Este mecanismo curiosamente exótico parece trabajar muy eficientemente. Es universalmente utilizado, tanto por vertebrados como por invertebrados. Pero los vertebrados perfeccionaron una innovación adicional, a saber, el tener la fibra nerviosa rodeada de una funda

aislante de una sustancia blanquecina grasa llamada *mielina*. (Es esta funda de mielina la que da su color a la «sustancia blanca» del cerebro.) Este aislamiento posibilita que la señales nerviosas viajen sin pérdidas (entre «estaciones repetidoras») a una velocidad muy respetable —hasta 120 metros por segundo.

Cuando una señal llega a un botón sináptico, éste desprende una sustancia química conocida como un neurotransmisor. Esta sustancia atraviesa el espacio sináptico hasta la otra neurona —ya sea en un punto de una de sus dendritas o en el propio soma. Ahora bien, algunas células tienen botones sinápticos que desprenden un neurotransmisor químico con tendencia a *excitar* el «disparo» del soma de la siguiente neurona, es decir, a iniciar una nueva señal a lo largo de su axón. Estas sinapsis se llaman *excitatorias*. Otras tienden a *desanimar* a la siguiente neurona de que se dispare y se llaman *inhibitorias*. El efecto total de las sinapsis excitatorias que están activas en cualquier momento se suma, y de ello se resta el total de las sinapsis inhibitorias activas; si el resultado neto alcanza cierto umbral crítico, la siguiente neurona es inducida al disparo. (Las excitatorias dan lugar a una *diferencia de potencial* eléctrico positiva entre el interior y el exterior de la neurona siguiente y las inhibitorias dan lugar a una diferencia de potencial negativa. Estas diferencias de potencial se suman de la forma apropiada. La neurona se disparará cuando esta diferencia de potencial alcance un nivel crítico en el axón asociado, de modo que el potasio no pueda salir con suficiente rapidez para restaurar el equilibrio.)

Modelos de computador

Una característica importante de la transmisión nerviosa es que las señales son (en su mayoría) fenómenos del tipo «todo-o-nada». La intensidad de la señal no varía: o está o no está. Esto da a la acción del sistema nervioso un aspecto parecido al de un computador. En realidad hay un gran número de semejanzas entre la acción de un gran número de neuronas interconectadas y el funcionamiento interno de un computador digital, con sus cables conductores de corriente y sus puertas lógicas (en un momento veremos más sobre esto). No sería difícil, en principio, hacer una simulación por ordenador de la acción de un sistema semejante de neuronas. Surge naturalmente una pregunta: ¿no significa esto que, cualquiera que pueda ser el cableado

del cerebro, siempre podrá ser modelado por la acción de un computador?

Para hacer más clara esta comparación diré lo que es realmente una *puerta lógica*. En un computador tenemos también una situación «todo-o-nada» en donde o hay un pulso de corriente en un cable conductor o no lo hay, siendo siempre la misma la intensidad del pulso cuando *está* presente. Puesto que todo está cronometrado de manera muy exacta, la *ausencia* de un pulso será una señal definida, y será algo que pueda ser «notado» por el computador. De hecho, cuando utilizamos el término «puerta lógica» estamos considerando implícitamente que la presencia o ausencia de un pulso significa «verdadero» o «falso» respectivamente. En realidad, esto no tiene nada que ver con la verdad o falsedad real: sólo se utiliza normalmente para hacer comprensible una terminología. Escribamos también el dígito «1» para «verdadero» (presencia de pulso) y «0» para «falso» (ausencia de pulso) y, como en el Capítulo 4, utilicemos «&» para «y» (que es el enunciado de que ambos son «verdaderos», es decir, la respuesta es 1 si y sólo si ambos argumentos son 1), «v» para «o» (que «significa» que uno u otro o ambos son «verdaderos», es decir, 0 si y sólo si ambos argumentos son 0), « $\Rightarrow$ » para «implica» (es decir, $A \Rightarrow B$ significa que si **A** es verdadero entonces **B** es verdadero, que es equivalente a «o **A** es falso o **B** es verdadero»), « $\Leftrightarrow$ » para «si y solo si» (ambos «verdaderos» o ambos «falsos»), « $\sim$ » para «no» («verdadero» si «falso»; «falso» si «verdadero»). Podemos describir la acción de las diversas operaciones lógicas en términos de las llamadas «tablas de verdad»:

$$\begin{array}{ll} \mathbf{A \& B:} & \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} & \mathbf{A \vee B:} & \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \\ \mathbf{A \Rightarrow B:} & \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} & \mathbf{A \Leftrightarrow B:} & \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array}$$

donde, en cada caso, **A** etiqueta las filas (esto es, **A** = 0 da la primera fila y **A** = 1, la segunda) y análogamente **B** etiqueta las columnas. Por ejemplo, si **A** = 0 y **B** = 1, lo que da la entrada superior derecha de cada tabla, entonces en la *tercera* tabla tenemos 1 para el valor de $A \Rightarrow B$. (Un ejemplo verbal de esto en términos de lógica *real*: la afirmación «si estoy dormido entonces soy feliz» se verifica ciertamente —de manera trivial— en el caso particular en que resulte estar a

la vez despierto y feliz.) Finalmente, la puerta lógica «no», sencillamente tiene el efecto:

$$\sim 0 = 1 \quad \text{y} \quad \sim 1 = 0.$$

Éstos son los tipos básicos de puertas lógicas. Existen algunas otras, pero todas estas pueden construirse a partir de las mencionadas ⁶.

Ahora bien, ¿podemos en principio construir un computador a base de conexiones de *neuronas*? Indicaré que, incluso sólo con las consideraciones muy primitivas sobre el disparo de las neuronas que acabamos de discutir, esto es realmente posible. Veamos cómo puede ser posible en principio construir puertas lógicas a base de conexiones de neuronas. Necesitamos una nueva manera de codificar los dígitos, puesto que la ausencia de una señal no desencadena nada. Consideremos (de forma arbitraria) que un pulso *doble* denota 1 (o «verdadero») y un pulso *simple* denota 0 (o «falso») y consideremos un esquema simplificado en el que el umbral de disparo de una neurona sea siempre igual a *dos* pulsos excitatorios simultáneos. Es fácil construir una puerta «y» (es decir, «&»). Como se muestra en la Figura 9.13, podemos considerar que las dos fibras nerviosas de entrada

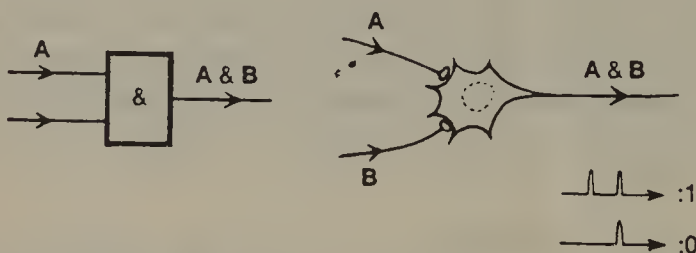


Fig. 9.13 Una puerta «y». En el «modelo de neurona» de la derecha, se considera que la neurona se «dispara» sólo cuando el input alcanza una intensidad doble de la de un pulso simple

terminan en el único par de botones sinápticos en la neurona de salida. (Si ambos son pulsos dobles, entonces tanto el primer pulso como el segundo alcanzarán el umbral exigido de dos-pulsos, mientras que si uno cualquiera de ellos es un pulso simple, entonces sólo uno del par alcanzará este umbral. Estoy suponiendo que los pulsos están minuciosamente cronometrados y que, para más concreción, en el caso de un pulso doble es el *primero* del par el que determina el cronometraje.) La construcción de una puerta «no» (es decir, «~»)

es mucho más complicada, y una manera de hacerlo se da en la Figura 9.14. Aquí la señal de entrada procede a lo largo de un axón que se divide en dos ramas. Una rama toma una ruta indirecta, de longitud tal como para retardar la señal en un intervalo exactamente igual al intervalo de tiempo entre los dos pulsos de un pulso doble; luego ambas ramas se bifurcan una vez más, con una rama de cada bifurcación terminando en una neurona inhibitoria, pero de forma que la que viene de la rama retrasada se desdobla primero para dar una ruta directa y otra indirecta. La salida de dicha neurona sería *nada*, en el caso de un pulso simple de entrada, y un *pulso doble* (en posición retardada) en el caso de un pulso doble de entrada. El axón que conduce esta salida se desdobla en tres ramas, todas ellas terminando en botones sinápticos inhibitorios en una neurona final excitatoria. Cada una de las dos ramas restantes del axón originalmente dividido se divide en dos una vez más y las cuatro ramas terminan también en esta neurona final pero ahora con botones sinápticos excitatorios. El lector puede ocuparse en verificar que esta neurona excitatoria final proporciona la requerida salida «no» (es decir, pulso doble si la en-

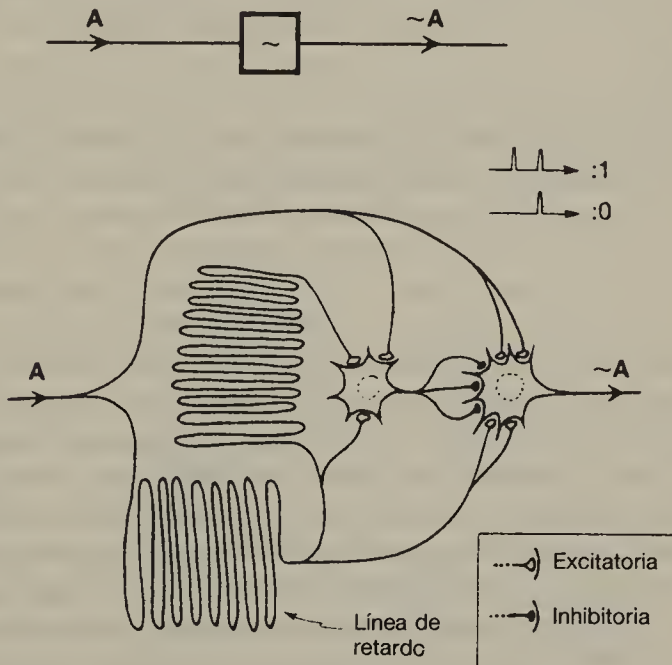


Fig. 9.14. Una puerta «no». En el «modelo de neurona» se necesita de nuevo un input de una intensidad doble (al menos) para que una neurona se dispare

trada es uno simple y un pulso simple si la entrada es uno doble). (¡Este esquema parece absurdamente complicado, pero es lo mejor que he podido imaginar!) El lector o lectora puede entretenerse buscando construcciones «neurales» para las otras puertas lógicas citadas.

Por supuesto, estos ejemplos explícitos no deben ser considerados como modelos serios de lo que es el cerebro en detalle. Sólo estoy tratando de indicar que existe una equivalencia lógica esencial entre el modelo de disparo de neurona que he dado arriba y la construcción de un computador electrónico. Es fácil ver que un computador puede simular cualquier modelo semejante de interconexiones neuronales; a su vez, las construcciones detalladas dadas arriba dan una indicación del hecho de que, recíprocamente, los sistemas de neuronas pueden simular un computador —y por lo tanto podrían actuar como una máquina (universal) de Turing. Aunque la discusión de las máquinas de Turing que se dio en el Capítulo 2 no utilizaba puertas lógicas, y de hecho necesitamos mucho más que sólo puertas lógicas si tenemos que simular una máquina de Turing general, no existen nuevas cuestiones de principio implicados en su realización —*siempre que nos esté permitido aproximar la cinta infinita* de una máquina de Turing por un banco de neuronas grande pero finito. ¡Esto parecería ser un argumento de que los cerebros y los computadores son esencialmente equivalentes!

Pero antes de dar un salto demasiado apresurado a esta conclusión, consideramos diversas diferencias entre la acción cerebral y la acción de los computadores actuales que posiblemente puedan ser de importancia. En primer lugar, he simplificado bastante mi descripción del disparo de una neurona como un fenómeno todo-o-nada. Eso se refiere a un pulso simple que viaja a lo largo de un axón pero, en realidad, cuando una neurona se «dispara» emite toda una secuencia de tales pulsos en rápida sucesión. Incluso cuando no está activada, una neurona emite pulsos pero sólo a un ritmo lento. Cuando se dispara es la *frecuencia* de estos pulsos sucesivos la que aumenta enormemente. Existe también un aspecto probabilístico en el disparo de una neurona. El mismo estímulo no siempre produce el mismo resultado. Además, la acción cerebral no tiene exactamente el cronometraje preciso que se necesita para las corrientes de un computador electrónico; y habría que señalar que la acción de las neuronas —a un ritmo máximo de unas 2.000 veces por segundo— es mucho más lenta que la de los circuitos electrónicos más rápidos, en un factor de aproximadamente 10^{-6} . También, a diferencia del cableado muy preciso de un computador electrónico, parece que hubiera mucha alea-

toriedad y redundancia en la forma detallada en que están conectadas realmente las neuronas —aunque sabemos ahora que existe una precisión mucho mayor en el modo en que está cableado el cerebro (en el nacimiento) de lo que se había pensado unos cincuenta años atrás.

Casi todo lo anterior parecería ser una desventaja para el cerebro comparado con un computador. Pero hay otros factores a favor del cerebro. Con las puertas lógicas hay muy pocos cables de entrada y de salida (digamos tres o cuatro como mucho) mientras que las neuronas pueden tener un número enorme de sinapsis. (Como un ejemplo extremo, las neuronas del cerebelo conocidas como células de Purkinje tienen unos 80.000 terminales sinápticos excitatorios.) También, el número total de neuronas del cerebro todavía supera con mucho al número de transistores incluso en el mayor computador actual —¡probablemente 10^{11} para el cerebro y «sólo» unas 10^9 para el computador! Pero, por supuesto, es muy probable que la cifra para el computador vaya aumentando en el futuro. Además, el alto número de células cerebrales aparece generalmente debido al inmenso número de pequeñas *células granulosas* que se encuentran en el cerebelo —alrededor de treinta mil millones (3×10^{10}). Si creemos que es simplemente la enormidad del número de neuronas la que nos permite tener experiencias conscientes, mientras que los actuales computadores no parecen tenerlas, entonces necesitamos encontrar alguna explicación adicional de por qué la acción del *cerebelo* parece ser completamente *inconsciente* mientras que la consciencia puede estar asociada con el *cerebro* propiamente dicho que tiene sólo el doble de neuronas (unas 7×10^{10}) con una densidad mucho más pequeña.

Plasticidad cerebral

Hay algunos otros puntos de diferencia entre la acción cerebral y la acción de un computador que me parecen ser de mucha más importancia que los mencionados hasta ahora, y que tienen que ver con un fenómeno conocido como *plasticidad cerebral*. No es realmente legítimo considerar el cerebro simplemente como una colección *fija* de cables neuronales. Las interconexiones entre neuronas no son fijas, como lo serían en el modelo de computador anterior, sino que están cambiando continuamente. No quiero decir que cambien las localizaciones de los axones o las dendritas; gran parte de su complejo «cableado» está establecido en líneas generales desde el nacimiento. Me estoy refiriendo a las uniones sinápticas en las que tiene lugar real-

mente la comunicación entre neuronas diferentes. A menudo éstas ocurren en lugares llamados *espinas dendríticas*, que son minúsculas protuberancias en las dendritas con las cuales pueden tomar contacto los botones sinápticos (ver Fig. 9.15). Aquí, «contacto» ya *no* significa simplemente tocar sino dejar una estrecha separación (espacio sináptico) de la distancia justa —alrededor de cuarenta milésimas de milímetro. Bajo ciertas condiciones, estas espinas dendríticas pueden contraerse y romper el contacto, o pueden (ellas u otras nuevas) crecer y hacer nuevos contactos. Por lo tanto, si pensamos que las conexiones de neuronas en el cerebro constituyen en efecto un computador, entonces ¡es un computador capaz de cambiar continuamente!

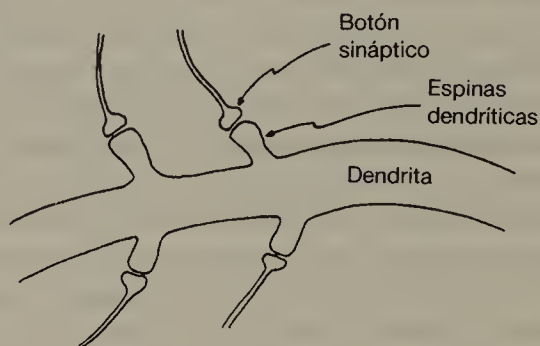


Fig. 9.15. Uniones sinápticas que incluyen espinas dendríticas. La efectividad de la unión queda inmediatamente afectada por el crecimiento o contracción de la espina

Según una de las principales teorías sobre cómo se establece la memoria a largo plazo, son tales cambios en las conexiones sinápticas los que proporcionan los medios para almacenar la información necesaria. Si esto es así, vemos entonces que la plasticidad cerebral no es ya una complicación accidental sino que es una característica *esencial* de la actividad del cerebro.

¿Cuál es el mecanismo subyacente a estos cambios continuos? ¿Con qué rapidez se efectúan estos cambios? La respuesta a la segunda pregunta parece controvertida, pero hay al menos una escuela de pensamiento que mantiene que tales cambios sólo pueden ocurrir en unos pocos segundos. Eso habría que esperar si tales cambios son los responsables del almacenamiento de memorias permanentes, puesto que tales memorias pueden asentarse en cuestión de segundos (cf. Kandel, 1976). Esto tendrá significativas consecuencias para nosotros más

adelante. Volveré a esta importante cuestión en el próximo capítulo.

¿Qué hay de los mecanismos subyacentes a la plasticidad cerebral? Una ingeniosa teoría (debida a Donald Hebb, 1954) propone la existencia de ciertas sinapsis (ahora llamadas «sinapsis de Hebb») con la propiedad siguiente: una sinapsis de Hebb entre una neurona A y una neurona B será reforzada cuando el disparo de A vaya seguido del disparo de B, y debilitada si no lo hace. Esto es independiente de si la propia sinapsis de Hebb está o no involucrada de modo significativo en causar el disparo de B. Ello da lugar a cierta forma de «aprendizaje». Se han propuesto varios modelos matemáticos para tratar de simular una actividad de aprendizaje de solución de problemas basada en este tipo de teoría. Éstos se conocen como *redes neurales*. Parece que tales modelos son realmente capaces de algún tipo de aprendizaje rudimentario, pero hasta ahora están lejos de ser modelos realistas del cerebro. En cualquier caso, parece probable que los mecanismos que controlan los cambios en las conexiones sinápticas sean bastante más complicados que los considerados. Evidentemente se requiere una mejor comprensión.

En relación con esto existe otro aspecto de la liberación de neurotransmisores por los botones sinápticos. A veces ésta no tiene lugar en los espacio sinápticos sino que los neurotransmisores entran en el fluido intercelular general, quizá para influir a otras neuronas a gran distancia. Muchas sustancias neuroquímicas diferentes parecen emitirse de esta forma —y existen teorías de la memoria diferentes de la que he citado, que dependen de las posibles variedades distintas de tales sustancias químicas que puedan estar involucradas. Ciertamente el estado del cerebro puede ser influenciado de un modo general por la presencia de sustancias químicas que son producidas por otras partes del cerebro (v.g. como sucede con las hormonas). El problema general de la neuroquímica es complicado, y es difícil ver cómo proporcionar una simulación por computador detallada y fiable de todo lo que pueda ser relevante.

Computadores paralelos y la «unicidad» de la consciencia

Mucha gente parece ser de la opinión de que el desarrollo de los computadores *paralelos* tiene la clave para construir una máquina con las capacidades de un cerebro humano. Consideraremos a continuación esta idea muy popular actualmente. Un computador en paralelo, como opuesto a un computador en serie, lleva a cabo simultáneamen-

te un gran número de cálculos independientes, y los resultados de estas operaciones básicamente autónomas son combinados intermitentemente para dar contribuciones al cálculo global. La motivación para este tipo de arquitectura de computador proviene básicamente de un intento de imitar el modo de operar del sistema nervioso, puesto que diferentes partes del cerebro parecen llevar a cabo funciones de cálculo separadas e independientes (v.g. el procesado de información visual en el córtex visual).

Hay que hacer aquí dos puntualizaciones. La primera es que no hay diferencia *en principio* entre un computador en serie y uno en paralelo. Ambos son de hecho *máquinas de Turing* (cf. Capítulo 2, p. 78). Sólo puede haber diferencias en eficiencia, o rapidez, del cálculo como un todo. Hay algunos tipos de cálculo para los que una organización en paralelo es realmente más eficiente, pero no es éste el caso siempre ni mucho menos. El segundo punto es que, al menos en mi opinión, es muy poco probable que la computación en paralelo clásica tenga la clave de cómo se está desarrollando nuestro pensamiento *consciente*. Una característica distintiva del pensamiento consciente (al menos cuando estamos en un estado psicológico normal, ¡y no sometidos a una operación de «escisión cerebral»!) es su «unicidad» —en oposición a las muchas actividades independientes que transcurren al mismo tiempo.

Expresiones tales como «¿Cómo puede esperar usted que piense más de una cosa cada vez?» son un lugar común. ¿Es *siquiera* posible mantener cosas separadas en la consciencia simultáneamente? Quizá se *puedan* mantener unas pocas cosas a la vez, pero esto se parece más a un continuo ir y venir entre los varios temas sobre los que se está pensando simultánea, consciente e independientemente. Si fuéramos a pensar conscientemente sobre dos cosas totalmente independientes sería más probable que tuviésemos dos *consciencias separadas*, incluso aunque fuera sólo por un corto tiempo, mientras que parece que experimentamos (al menos en una persona normal) *una sola* consciencia que puede ser vagamente consciente de cierto número de cosas, pero que en un instante dado se concentra en *una sola* cosa concreta.

Por supuesto, lo que queremos decir aquí con «una sola cosa» no está completamente claro. En el próximo capítulo encontraremos algunos ejemplos muy notables de «pensamientos simples» en las inspiraciones de Poincaré y Mozart. Pero no tenemos que ir tan lejos para reconocer que aquello de lo que una persona puede ser consciente en un instante dado puede ser implícitamente muy complicado.

Imaginemos que vamos a decidir lo que queremos cenar, por ejemplo. Podrá haber mucha información implícita en una idea consciente semejante, y una descripción verbal completa sería bastante larga.

Esta «unicidad» de la percepción consciente me parece que está en fuerte oposición con la imagen de un computador paralelo. Por el contrario, esta imagen podría ser más apropiada como un modelo de la acción *inconsciente* del cerebro. Se pueden realizar simultáneamente y de forma más o menos automática varios movimientos independientes diferentes —andar, abrochar un botón, respirar, o incluso hablar— ¡sin tener una consciencia absoluta de *ninguno* de ellos!

Por el contrario, me parece concebible que pudiera existir alguna relación entre esta «unicidad» de la consciencia y el *paralelismo cuántico*. Recuérdese que, según la teoría cuántica, está permitido que coexistan en una superposición alternativas diferentes en el nivel cuántico. Por lo tanto, un *estado cuántico simple* podría consistir en principio en un gran número de actividades diferentes, todas ellas ocurriendo simultáneamente. Esto es lo que quiere decir el paralelismo cuántico, y en un momento consideraremos la idea teórica de un «computador cuántico» en el que podría utilizarse en principio este paralelismo para ejecutar un gran número de cálculos simultáneos. Si un «estado mental» consciente pudiera ser en algún modo semejante a un estado cuántico, entonces alguna forma de «unicidad» o globalidad de pensamiento parecería más apropiada de lo que lo sería en el caso de un computador paralelo ordinario. Hay algunos aspectos atractivos en esta idea a los que volveré en el próximo capítulo. Pero antes de que podamos mantener seriamente esta idea debemos plantear la cuestión de si es probable que los efectos cuánticos tengan una mínima relevancia en la actividad del cerebro.

¿Hay un papel para la mecánica cuántica en la actividad cerebral?

Las discusiones anteriores sobre la actividad cerebral han sido totalmente clásicas, excepto en el momento en que ha sido necesario apelar a fenómenos físicos cuyas causas implícitas subyacentes son parcialmente mecano-cuánticas (v.g. iones, con su carga eléctrica unidad, compuertas de sodio y de potasio, los potenciales químicos definidos que determinan el carácter sí/no de las señales nerviosas, la química de los neurotransmisores). ¿Hay algún papel más definido para un control genuinamente mecano-cuántico en algún lugar clave?

Esto parecería necesario si la discusión del final del capítulo anterior fuera a tener auténtica relevancia.

Existe, de hecho, al menos un lugar evidente en donde la acción en el nivel mecano-cuántico simple puede tener importancia para la actividad neural, y éste está en la *retina*. (¡Recuérdese que técnicamente la retina es parte del cerebro!) Experimentos con sapos han mostrado que, en condiciones adecuadas, *un sólo fotón* que incida en la retina adaptada a la oscuridad puede ser suficiente para desencadenar una señal nerviosa macroscópica (Baylor, Lamb y Yau, 1979). Lo mismo parece ser cierto para el hombre (Hecht, Shlaer y Pirenne, 1941) pero en este caso hay presente un mecanismo adicional que suprime tales señales débiles de modo que no hagan confusa la imagen percibida con demasiado «ruido» visual. Se necesita una señal combinada de unos *siete* fotones para que un sujeto humano adaptado a la oscuridad pueda hacerse consciente realmente de su llegada. De todas formas, parece que hay presentes en la retina humana células con sensibilidad a un solo fotón.

Puesto que en el cuerpo humano *existen* neuronas que pueden ser disparadas por sucesos cuánticos simples, ¿no es razonable preguntar si podrían encontrarse células de este tipo en algún lugar de las partes principales del cerebro humano? Por lo que yo conozco no existe evidencia de esto. Todos los tipos de célula que se han examinado requieren que se alcance un umbral, y se necesita un número muy elevado de cuantos para que se dispare la célula. Podríamos especular, no obstante, que en algún lugar profundo del cerebro se encontrarán células con sensibilidad a un solo cuanto. Si se probara que éste es el caso, entonces la mecánica cuántica estaría involucrada de manera significativa en la actividad cerebral.

Incluso esto no parece todavía muy *útil* mecano-cuánticamente, puesto que el cuanto se está utilizando simplemente como medio de desencadenar una señal. No se ha obtenido ningún efecto característico de interferencia cuántica. Parece que, como mucho, todo lo que saquemos de esto será una incertidumbre sobre si una neurona se disparará o no, y es difícil ver cómo esto sería de mucha ayuda para nosotros.

Sin embargo, algunos de los temas implicados aquí no son así de simples. Volvamos a considerar la retina. Supongamos que llega un fotón a la retina, habiendo sido reflejado previamente en un espejo semi-reflectante. Su estado implica una superposición lineal compleja de su impacto en una célula de la retina y su no impacto en una célula de la retina, saliendo en lugar de ello, por ejemplo, por la ventana

hacia el espacio (cf. Fig. 6.17, p. 322). Cuando llega el instante en el que *pudiera* haber golpeado la retina, y en tanto que la regla lineal U de la teoría cuántica siga siendo válida (es decir, evolución determinista del vector de estado de acuerdo con la ecuación de Schrödinger, cf. p. 317), tendremos una superposición lineal compleja de una señal nerviosa y una ausencia de señal nerviosa. Cuando ésta incida sobre la consciencia del sujeto sólo se percibirá *una* de estas alternativas teniendo lugar efectivamente, y se deberá haber realizado el otro procedimiento mecano-cuántico R (reducción del vector de estado, cf. p. 317). (Al decir esto estoy dejando de lado la idea de los muchos universos, cf. p. 372, ¡que tiene su multitud de problemas propios!) En línea con las consideraciones apuntadas al final del capítulo anterior deberíamos preguntar si el paso de la señal perturba la cantidad de materia suficiente para que se alcance el criterio de *un-gravitón*. Aunque es cierto que en la retina se consigue una impresionante amplificación para convertir la energía del fotón en un movimiento de masa en la señal real —quizá en un factor que llegue a 10^{20} en masa movida— esta masa queda aún muy lejos de la masa de Planck m_p en un factor muy grande (digamos del orden de 10^8). Sin embargo, una señal nerviosa crea un *campo eléctrico* variable detectable en sus proximidades (un campo toroidal, con el nervio como eje, moviéndose a lo largo del nervio). Este campo podría perturbar significativamente el *entorno* y el criterio de un-gravitón podría alcanzarse fácilmente en dicho entorno. Por lo tanto, según el punto de vista que he estado exponiendo, el procedimiento R podría haberse efectuado ya antes de que percibamos o no el destello de luz, según sea el caso. Desde este punto de vista ¡nuestra consciencia no es necesaria para reducir el vector de estado!

Computadores cuánticos

Si *especulamos* con que dichas neuronas sensibles a un solo cuanto están jugando un papel importante en lo más profundo del cerebro, podemos preguntar qué efectos tendrán. Discutiré primero el concepto de Deutsch de *computador cuántico* (cf. también Capítulo 4, p. 194) y luego plantearé si podemos considerarlo relevante para nuestras consideraciones sobre este punto.

Como se indicó antes, la idea básica consiste en hacer uso del paralelismo cuántico según el cual debe considerarse que dos cosas

muy diferentes tienen lugar simultáneamente en una superposición lineal cuántica —como el fotón que simultáneamente se refleja y atraviesa el espejo semi-reflectante o quizá atraviesa cada una de las dos rendijas. Para un computador cuántico estas dos cosas diferentes superpuestas serían, en su lugar, dos *computaciones* diferentes. No nos suponemos interesados en obtener las respuestas a *ambos* cómputos, sino en algo que utilice parte de la información extraída del par superpuesto. Finalmente, se haría una «observación» adecuada sobre el par de cómputos, cuando ambos estuvieran acabados, para obtener la respuesta requerida'. De esta forma, el dispositivo podría ahorrar tiempo ¡ejecutando los dos cómputos simultáneamente! Hasta aquí no parece que haya una ganancia significativa al proceder de esta manera, ya que presumiblemente sería mucho más directo utilizar un par de computadores clásicos distintos en paralelo (o un solo computador paralelo clásico) que un computador cuántico. Sin embargo, la ganancia real para un computador cuántico llegaría cuando se necesitase un número *muy grande* de cálculos paralelos —quizá un número indefinidamente grande— cuyas respuestas no nos interesasen individualmente sino sólo alguna combinación apropiada de todos los resultados.

En concreto, la construcción detallada de un computador cuántico implicaría una versión cuántica de una puerta lógica, en donde el output sería el resultado de cierta «operación unitaria» aplicada al input —un caso de la acción de U — y toda la operación del computador estaría llevando a cabo hasta el final un proceso U , hasta que un «acto de observación» final haría entrar en acción a R .

Según el análisis de Deutsch, los computadores cuánticos no pueden utilizarse para ejecutar operaciones no-algorítmicas (es decir, cosas más allá del poder de una máquina de Turing), pero pueden, en ciertas situaciones muy artificiales, conseguir una velocidad mayor, en el sentido de la *teoría de la complejidad* (ver pp. 187-188), que la de una máquina de Turing estándar. Hasta aquí estos resultados son un poco decepcionantes para una idea tan sorprendente, pero todavía estamos en el principio.

¿Qué relación podría tener esto con la acción de un cerebro que contenga un número significativo de neuronas sensibles a un solo cuanto? El mayor problema con la analogía sería que los efectos cuánticos se perderían rápidamente en el «ruido» —el cerebro es un objeto demasiado «caliente» para preservar la coherencia cuántica (es decir, el comportamiento usualmente descrito por la acción continuada de U) durante cualquier intervalo de tiempo apreciable. En mis propios términos, esto significaría que continuamente se estaría cumpliendo

el criterio de un-gravitón, de modo que la operación **R** continuaría todo el tiempo, entremezclada con **U**.

Hasta aquí, esto no parece muy prometedor si esperamos sacar de la mecánica cuántica algo útil para el cerebro. ¡Quizá estemos condenados a ser computadores después de todo! Yo, personalmente, no lo creo. Pero se necesitan consideraciones adicionales si queremos descubrir nuestra vía de escape.

¿Más allá de la teoría cuántica?

Quiero volver a un punto que ha sido un tema subyacente en gran parte de este libro. ¿Nuestra imagen de un mundo gobernado por las reglas de las teorías clásicas y cuánticas, tal como las entendemos actualmente, es verdaderamente adecuada para la descripción de cerebros y mentes? Ciertamente existe un interrogante para cualquier descripción cuántica «ordinaria» de nuestros cerebros, ya que la acción de «observación» se considera un ingrediente esencial de la interpretación válida de la teoría cuántica convencional. ¿Debe considerarse el cerebro «observándose a sí mismo» cuando un pensamiento o percepción emerge a la atención consciente? La teoría convencional nos deja sin ninguna regla clara de cómo la teoría cuántica podría tomar esto en consideración y, en consecuencia, aplicarlo al cerebro como un todo. He intentado formular un criterio para la aparición de **R** que es independiente de la consciencia (el «criterio de un-gravitón») y si se pudiera desarrollar algo parecido a esto en una teoría completa coherente entonces podría emerger una manera de proporcionar una descripción cuántica del cerebro más clara de la que existe actualmente.

Sin embargo, creo que estos problemas fundamentales no surgen sólo en nuestros intentos de describir la acción cerebral. Las acciones de los propios computadores digitales dependen de forma vital de efectos cuánticos —efectos que, en mi opinión, no están totalmente libres de las dificultades inherentes a la teoría cuántica. ¿Cuál es esta dependencia cuántica «vital»? Para comprender el papel de la mecánica cuántica en el cálculo digital debemos preguntar primero cómo podríamos intentar hacer que un objeto totalmente *clásico* se comporte como un computador digital. En el Capítulo 5 considerábamos el «computador de bolas-de-billar» clásico de Fredkin-Toffoli (p. 223); pero señalábamos también que este «dispositivo» teórico depende de

ciertas idealizaciones que dejan de lado un problema de inestabilidad esencial inherente a los sistemas clásicos. Este problema de inestabilidad se describió como una difusión efectiva en el espacio de fases a medida que evoluciona el tiempo (p. 235; Fig. 5.14), lo que conduce a una casi inevitable pérdida de precisión en la operación de un dispositivo clásico. Es la mecánica cuántica la que en última instancia detiene esta degradación de la precisión. En los computadores electrónicos modernos es necesaria la existencia de *estados discretos* (por ejemplo, para codificar los dígitos 0 y 1), de modo que la cuestión de si el computador se encuentra en uno de estos estados o en el otro tiene una respuesta precisa. Esto constituye la propia esencia de la naturaleza «digital» de la operación del computador. Esta discretización depende en última instancia de la mecánica cuántica. (Recordemos la discretización cuántica de los estados de energía, de las frecuencias espectrales, del spin, etc., cf. Capítulo 6.) Incluso las viejas calculadoras mecánicas dependían de la *solidez* de sus diversas partes —y la solidez, también, descansa realmente en la discretización de la teoría cuántica.

Pero la discretización cuántica no es obtenible de la sola acción de **U**. ¡La ecuación de Schrödinger es *peor* aún para impedir la indeseable difusión y «pérdida de precisión» de lo que lo son las ecuaciones de la física clásica! Según **U**, una función de onda de una simple partícula, inicialmente localizada en el espacio, se difundirá asimismo sobre regiones cada vez mayores a medida que pasa el tiempo (p. 318). Los sistemas más complejos también estarían sometidos a veces a esta irrazonable falta de localización (¡recuérdese el gato de Schrödinger!) si no fuera por la acción ocasional de **R**. (Los estados *discretos* de un átomo, por ejemplo, son aquellos con energía, momento y momento angular total definidos. Un estado general que «se difunde» es una superposición de estados discretos semejantes. Es la acción de **R**, en alguna etapa, la que impone que el átomo «esté» realmente en uno de estos estados discretos.)

Creo que ni la mecánica clásica ni la cuántica —la última sin ningún cambio fundamental adicional que hiciera de **R** un proceso «real»— podrán explicar nunca la forma en que *pensamos*. Quizá incluso la acción digital de los computadores necesita una comprensión más profunda de la interrelación entre las acciones de **U** y **R**. Con los computadores sabemos, al menos, que la acción es *algorítmica* (¡porque así los hemos diseñado!), y no tratamos de sacar provecho de ningún supuesto comportamiento *no*-algorítmico en las leyes físicas. Pero yo mantengo que con los cerebros y las mentes la situación

es muy diferente. Se puede alegar plausiblemente que existe un ingrediente *no*-algorítmico esencial en los procesos de pensamiento (consciente). En el próximo capítulo trataré de exponer las razones de mi creencia en tal ingrediente y especular acerca de qué curiosos efectos físicos reales podrían constituir una «consciencia» que influya en la acción del cerebro.

10. ¿Dónde reside la física de la mente?

¿Para qué están las mentes?

Cuando se discute el problema mente-cuerpo se suele centrar la atención en dos temas: «¿Cómo es posible que una consciencia pueda *surgir* realmente a partir de un objeto material (un cerebro)?»; y, recíprocamente; «¿Cómo es posible que una consciencia, mediante la acción de su voluntad, *influya* realmente en el movimiento (que en apariencia está determinado físicamente) de los objetos materiales?». Éstos son los aspectos pasivo y activo del problema mente-cuerpo. Parece que tenemos, en la «mente» (o, más bien, en la «consciencia»), «algo» no-material que, por una parte, está producido por el mundo material y, por otra, puede influenciarlo. Sin embargo, en mis discusiones preliminares en este último capítulo preferiré considerar un problema algo diferente y quizá más científico —que tiene relevancia tanto para el problema pasivo como para el activo— con la esperanza de que nuestros intentos de encontrar una respuesta puedan acercarnos un poco hacia una mejor comprensión de estos viejos enigmas fundamentales de la filosofía. Mi pregunta es: «¿Qué *ventaja selectiva* confiere la consciencia a quienes realmente la poseen?»

Al enunciar la pregunta de esta manera se hacen implícitamente varias suposiciones. En primer lugar, está la creencia de que la consciencia es realmente «algo» *descriptible científicamente*. Se supone también que esta «cosa» realmente «hace algo» —y, además, que lo que hace es útil para la criatura que la posee, de modo que otra cualquiera criatura equivalente, aunque sin consciencia, se comportaría de una forma menos eficaz. En el extremo opuesto se podría pensar que la consciencia es simplemente un concomitante pasivo de la posesión de

un sistema de control suficientemente elaborado, y que *no* «hace» *nada* realmente por sí sola. (Este último sería presumiblemente el punto de vista de los defensores de la IA fuerte, por ejemplo.) Como alternativa, quizá exista algún propósito divino o misterioso para el fenómeno de la consciencia —posiblemente un propósito teleológico aún no revelado a nosotros— y cualquier discusión de este fenómeno en términos simplemente de ideas de selección natural pasaría por alto este «propósito». Algo preferible, para mi modo de pensar, sería una versión más científica de este tipo de argumento, a saber, el *principio antrópico*, que afirma que la naturaleza del universo en el que nosotros mismos nos encontramos está fuertemente condicionada por la exigencia de que deben estar presentes seres sintientes como nosotros para observarlo. (Se aludió brevemente a este principio en el Capítulo 8, p. 440, y volveré a él más tarde.)

Abordaré la mayoría de estos temas a su debido tiempo, pero notemos primero que el término «mente» es quizá un poco equívoco cuando nos referimos al problema «mente-cuerpo». Después de todo hablamos a menudo de «la mente inconsciente». Esto demuestra que no consideramos que los términos «mente» y «consciencia» sean sinónimos. Quizá cuando nos referimos a la mente *inconsciente* tenemos una vaga imagen de «alguien ahí detrás» que actúa entre bastidores pero que normalmente (excepto quizá en sueños, alucinaciones, obsesiones o lapsus freudianos) no tiene incidencia directa sobre lo que percibimos. Quizá la mente inconsciente *tenga* realmente una consciencia autónoma, pero esta consciencia se mantiene normalmente separada de la parte de la mente a la que normalmente nos referimos como «nosotros».

Puede que esto no sea tan rebuscado como parece a primera vista. Existen experimentos que parecen indicar que puede haber algún tipo de «conocimiento» presente aun cuando un paciente esté siendo operado bajo anestesia general —en el sentido de que conversaciones mantenidas en ese momento pueden influir «inconscientemente» más tarde sobre el paciente, y pueden ser después recordadas bajo hipnosis como si realmente hubieran sido «experimentadas» en su momento. Además, algunas sensaciones que parecían haber quedado bloqueadas para la consciencia mediante sugestión hipnótica pueden ser recordadas más tarde bajo nueva hipnosis como si «hubieran sido experimentadas», aunque de algún modo aparcadas en una «vía diferente» (cf. Oakley y Eames, 1985). Estos resultados no están nada claros para mí, aunque no concibo que fuera correcto asignar un «conocimiento» ordinario a la mente inconsciente, y no tengo en verdad

deseos de tratar aquí estas especulaciones. De todas formas, la división entre la mente consciente y la inconsciente es ciertamente un tema sutil y complicado al que tendremos que volver.

Tratemos de ser tan claros como sea posible sobre lo que entendemos por «consciencia» y cuándo creemos que está presente. No creo que sea prudente, en este estadio de comprensión, intentar proponer una *definición* precisa de consciencia, pero podemos fiarnos, en buena medida, de lo que nuestras impresiones subjetivas y sentido común intuitivo nos dicen sobre el significado del término y sobre cuándo es probable que esté presente esta propiedad de consciencia. Yo sé, más o menos, cuando yo mismo estoy consciente, y supongo que los demás experimentan algo análogo a lo que yo experimento. Para estar consciente parece que debo ser consciente *de* algo, quizá de una sensación como dolor o acaloramiento, o de una escena animada o de un sonido musical; o quizá soy consciente de un sentimiento tal como intriga, desesperación o felicidad; o puedo ser consciente del recuerdo de alguna experiencia pasada, o de llegar a una comprensión de lo que algún otro está diciendo, o de una nueva idea propia; o puedo estar intentando conscientemente hablar o ejecutar alguna otra acción como la de levantarme de mi asiento. También puedo «remontarme al pasado» y ser consciente *de* tales intenciones o de mi sentimiento de dolor, o de mi experiencia de un recuerdo o de llegar a una comprensión; o incluso puedo ser simplemente consciente de mi propia consciencia. Puedo estar dormido y ser aún consciente en cierto grado, siempre que esté teniendo algún sueño; o quizá, a medida que me voy despertando, estoy influyendo conscientemente en el rumbo de ese sueño. Estoy dispuesto a creer que la consciencia es una cuestión de grado y no simplemente algo que está o no está. Supondré que la palabra «consciencia» es esencialmente sinónimo de «conocimiento» (aunque quizá «conocimiento» es algo más pasivo que lo que entiendo por «consciencia»), mientras que «mente» y «alma» tienen connotaciones adicionales que son mucho *menos* claramente definibles en el presente. Tendremos suficientes dificultades para entender lo que es la «consciencia» tal como está, así que espero que el lector me perdonará si no me ocupo en esencia de los problemas adicionales de «mente» y «alma»!

Está también la cuestión de qué entendemos por el término «inteligencia». Esto es, después de todo, lo que interesa a la gente de la IA, antes que el tema quizá más difuso de la «consciencia». Alan Turing (1950), en su famoso artículo, no se refería directamente a la «consciencia» sino al «pensamiento», y en el título figuraba la palabra

«inteligencia». En mi propia forma de ver las cosas la cuestión de la inteligencia es subsidiaria de la de la consciencia. Me parece poco concebible que la verdadera inteligencia pudiera estar presente a menos que estuviera acompañada de la consciencia. Por el contrario, si resultara que las gentes de la IA *fuera*n finalmente capaces de simular inteligencia sin que la consciencia esté presente, entonces podría considerarse insatisfactorio definir el término «inteligencia» sin incluir esta inteligencia simulada. En ese caso, el tema de la «inteligencia» no me interesaría realmente para este punto. Estoy interesado principalmente en la «consciencia».

Cuando afirmo mi propia creencia en que la verdadera inteligencia requiere consciencia estoy sugiriendo implícitamente (puesto que yo no creo en la tesis de la IA *fuerte* de que la simple activación de un algoritmo produciría la consciencia) que la inteligencia no puede simularse adecuadamente mediante procedimientos algorítmicos, es decir, mediante un computador, en el sentido en que hoy utilizamos el término. (Ver la discusión del «test de Turing» dada en el Capítulo 1.) En efecto, argumentaré con fuerza dentro de un momento (ver, concretamente, la discusión del pensamiento matemático dada tres secciones más adelante, en p. 516) que debe haber un ingrediente esencialmente *no-algorítmico* en la actuación de la consciencia.

Abordemos a continuación la cuestión de si *existe* una diferencia operacional entre algo que sea consciente y alguna cosa equivalente que no lo sea. ¿Revelaría siempre su presencia la consciencia en algún objeto? Me gustaría pensar que la respuesta a esta pregunta es necesariamente «sí». Sin embargo, mi fe en esto apenas se ve alentada por la falta total de consenso sobre en qué partes del reino animal debe encontrarse la consciencia. Algunos no admiten que pueda poseerla en absoluto algún animal no-humano (y, algunos, ni siquiera que la poseyeran los seres humanos antes de alrededor del año 1000 a.C., cf. Jaynes, 1980), mientras que otros atribuirán consciencia a un insecto, un gusano, ¡o quizá incluso una roca! Por mi parte dudaría que un gusano o un insecto —aunque ciertamente no una roca— tenga mucho, si tiene algo, de esta cualidad; pero los mamíferos, de modo general, me dan la impresión de cierta genuina consciencia. De esta falta de consenso podemos inferir, al menos, que no hay un criterio aceptado generalmente para la manifestación de la consciencia. Pudiera ser todavía que *exista* una impronta de comportamiento consciente, aunque no universalmente reconocida. Aun así, esto sólo podría significar el papel *activo* de la consciencia. Es difícil ver cómo la simple presencia de consciencia, sin su contrapartida activa, pudiera

ser directamente verificada. Esto tiene un apoyo en el terrible hecho de que, durante algún tiempo en los años 40, el curare fue utilizado como «anestésico» en operaciones realizadas a niños —cuando, en realidad, el efecto de esta droga es paralizar la acción de los nervios motores sobre los músculos, de modo que la agonía que *realmente experimentaban* estos infortunados niños no tenía modo de hacerse manifiesta al cirujano en esa época (cf. Dennett, 1978, p. 209).

Volvamos al posible papel activo que *pueda* tener la consciencia. ¿Es necesario que la consciencia deba jugar —y a veces lo hace realmente— un papel activo discernible operacionalmente? Tengo diversas razones para creerlo. Una primera razón la constituye el modo por el que, utilizando el «sentido común», tenemos a menudo la sensación de que percibimos directamente que alguna otra persona *está* realmente consciente? No parece muy probable que *esa* impresión sea falsa *. Mientras que puede darse el caso de que una persona que *está* consciente no lo esté de modo obvio (como los niños drogados con curare), ¡no es probable que una persona que *no* esté consciente parezca estarlo! Por lo tanto debe haber realmente algún modo de comportamiento que es característico de la consciencia (incluso aunque no sea *siempre* manifestado por la consciencia) y al que somos sensibles a través de nuestras «intuiciones de sentido común».

En segundo lugar, consideremos el implacable proceso de la selección natural. Veamos este proceso a la luz del hecho de que, como hemos visto en el último capítulo, no toda la actividad del cerebro es directamente accesible a la consciencia. De hecho, el cerebelo «más antiguo» —con su enorme superioridad en densidad local de neuronas— parece llevar a cabo acciones muy complejas sin que la consciencia esté en modo alguno involucrada directamente. No obstante, la Naturaleza ha decidido que evolucionen seres sintientes como nosotros mismos, en lugar de contentarse con criaturas que podrían estar dirigidos por mecanismos de control completamente inconscientes. Si la consciencia no sirve para ningún propósito evolutivo, ¿por qué la Naturaleza se tomó la molestia de hacer evolucionar cerebros *conscientes* cuando parece que hubieran bastado cerebros «autómatas» no sintientes como los cerebelos?

Además, existe una sencilla «razón definitiva» para creer que la consciencia debe tener *algún* efecto activo, incluso si este efecto *no* tiene una ventaja selectiva. En efecto, ¿por qué los seres como no-

* Al menos con la tecnología de computadores actual (véase la discusión del test de Turing dada en el Capítulo 1).

sotros mismos nos preocupamos a veces —especialmente cuando se investiga en el tema— por cuestiones sobre el «yo»? (Casi podría decir: «¿Por qué están *ustedes* leyendo este capítulo?» o «¿Por qué sentí *yo* en primer lugar un fuerte deseo de escribir un libro sobre este tema?») Es difícil imaginar que un autómatas completamente inconsciente perdiera su tiempo en estos temas. Puesto que, por el contrario, los seres conscientes *parecen* actuar de esta extraña manera de vez en cuando, ellos se están comportando en consecuencia de una forma que es *diferente* de la forma en que lo harían si *no fueran* conscientes —¡así que la consciencia tiene algún efecto activo! Por supuesto, no habría problema en programar deliberadamente un computador para que pareciera que se comportaba de esta forma ridícula (v.g. podría ser programado para que deambulara murmurando: «Oh, ¡Dios mío!, ¿Cuál es el sentido de la vida? «¿Por qué estoy aquí?» «¿Qué diantres es este “yo” que yo siento?»). Pero ¿por qué se molestaría la selección natural en favorecer semejante raza de individuos?, cuando seguramente el implacable mercado libre de la jungla ¡debería haber extirpado hace tiempo tales absurdos inútiles!

Me parece claro que estas meditaciones y murmuraciones a que nos entregamos cuando (quizá temporalmente) nos hacemos filósofos no son cosas que sean seleccionadas *por sí mismas*, sino que son el «equipaje» necesario (desde el punto de vista de la selección natural) que deben llevar los seres que *son* conscientes y cuya consciencia ha sido seleccionada por la selección natural aunque por alguna razón diferente y presumiblemente muy poderosa. Es un equipaje que no es demasiado gravoso y es fácil de llevar (aunque quizá a regañadientes), diría yo, para las indómitas fuerzas de la selección natural. En ocasiones, quizá cuando se dan la paz y la prosperidad de que a veces disfruta nuestra afortunada especie, y no tenemos que estar siempre luchando contra los elementos (o nuestros vecinos) por la supervivencia, los tesoros del contenido de nuestro equipaje pueden empezar a intrigarnos y a ser cuestionados. Es cuando vemos que otros se comportan con esta extraña conducta filosófica cuando nos quedamos *convencidos* de que estamos tratando con individuos, distintos de uno mismo, que también tienen mentes.

¿Qué hace realmente la consciencia?

Aceptemos que la presencia de la consciencia en una criatura supone realmente una ventaja selectiva para la misma. ¿Cuál sería en

concreto esta ventaja? Una idea que he oído expresar es que la consciencia podría ser ventajosa para un predador que trata de imaginar lo que probablemente haría su presa a continuación «poniéndose él mismo en el lugar» de la presa. Imaginarse que él mismo *es* la presa podría darle una ventaja sobre ella.

Es perfectamente posible que haya parte de verdad en esta idea, pero no me siento muy a gusto con ella. En primer lugar, ello supone alguna consciencia pre-existente por parte de la misma presa, puesto que difícilmente sería útil el imaginar que uno «sea» un autómatas, ya que un autómatas —*inconsciente* por definición— ¿no es algo que *se pueda* «ser» conscientemente! En cualquier caso, también podríamos imaginar que un predador autómatas totalmente inconsciente pudiera contener en sí mismo, como parte de su programa, una subrutina que fuera el programa real de su presa autómatas. No me parece que sea *lógicamente* necesario que la consciencia tenga que entrar de ninguna forma en esta interrelación predador-presa.

Por supuesto, es difícil ver cómo el proceso aleatorio de la selección natural podría haber sido suficientemente ingenioso para dar a un predador *autómatas* una copia completa del programa de la presa. ¡Esto sonaría más a *espionaje* que a selección natural! Y un programa *parcial* (en el sentido de un trozo de «cinta» de una máquina de Turing, o algo que se aproxime a una cinta de máquina de Turing) apenas sería de mucha ventaja selectiva para el predador. Parecería necesaria la improbable posesión de toda la cinta, o al menos de toda una parte autocontenida de ella. En consecuencia, y como una posible alternativa a esto, podría haber una verdad parcial en la idea de que algún elemento de consciencia, y no simplemente un programa de computador, pueda inferirse del argumento predador-presa. Pero ello no parece abordar el *verdadero* problema de cuál es realmente la diferencia entre una acción consciente y una «programada».

La idea aludida arriba parece estar relacionada con un punto de vista sobre la consciencia que se oye exponer a menudo, a saber: que un sistema tendrá «consciencia» de alguna cosa si tiene dentro de sí un modelo de esa cosa, y que se hace «*auto-consciente*» cuando tiene dentro de sí un modelo de *sí mismo*. Pero un programa de ordenador que contenga dentro de sí (digamos como subrutina) alguna descripción de otro programa de ordenador no hace al primer programa consciente del segundo; ni ningún aspecto *auto-referencial* de un programa le hace *auto-consciente*. A pesar de las afirmaciones que parecen hacerse con frecuencia, los temas reales concernientes a la consciencia y autoconsciencia apenas se tocan, en mi opinión, en conside-

raciones de este tipo. Una vídeo-cámara no es consciente de las escenas que está registrando; y tampoco una vídeo-cámara que esté dirigida hacia un espejo posee auto-consciencia (Fig. 10.1).

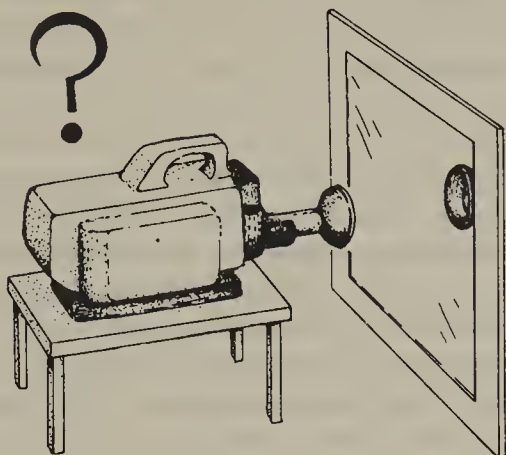


Fig. 10.1. Una vídeo-cámara dirigida hacia un espejo se forma un modelo de ella dentro de sí misma. ¿Le hace esto auto-consciente?

Deseo seguir una línea diferente. Hemos visto que no todas las actividades que llevan a cabo nuestros cerebros están acompañadas de atención consciente (y, en particular, la acción cerebelar no parece que sea consciente). ¿Qué podemos hacer con el pensamiento consciente que no podamos hacer inconscientemente? El problema se hace más escurridizo por el hecho de que cualquier cosa que hagamos parece exigir consciencia inicialmente, pues parece también capaz de ser aprendida y luego ejecutada inconscientemente (quizá por el cerebelo). De algún modo se necesita la consciencia para manejar situaciones en las que tenemos que hacer nuevos juicios, y en las que no se han establecido reglas por adelantado. Es difícil ser muy precisos sobre las distinciones entre los tipos de actividad mental que parecen exigir la consciencia y las que no lo hacen. Quizá, como sostendrían los defensores de la IA fuerte (y otros), nuestra «formación de nuevos juicios» estaría aplicando de nuevo algunas reglas algorítmicas bien definidas, pero unas de un oscuro «alto nivel» de cuyo funcionamiento no somos conscientes. Sin embargo, creo que el tipo de terminología que tendemos a utilizar, que distingue nuestra actividad mental consciente de la inconsciente, es al menos *sugeridora* de una distinción no-algorítmica:

Consciencia necesaria

«sentido común»
«juicio de verdad»

«comprensión»
«valoración artística»

Consciencia no necesaria

«automático»
«seguir reglas
despreocupadamente»
«programado»
«algorítmico»

Quizá estas distinciones no sean siempre muy precisas, debido en particular a que entran muchos factores *inconscientes* en nuestros juicios conscientes: experiencia, intuición, prejuicio, incluso nuestra utilización normal de la lógica. Pero, afirmaré, los propios juicios son las manifestaciones de la actuación de la *consciencia*. Sugiero así que, mientras que las acciones inconscientes del cerebro son las que proceden según procesos algorítmicos, la acción de la consciencia es muy diferente y actúa de una forma que no puede describirse mediante ningún algoritmo.

Resulta irónico que las ideas que estoy exponiendo aquí representan casi una inversión de otras que he oído frecuentemente. A menudo se dice que es la mente *consciente* la que se comporta de la forma «racional» que podemos entender, mientras que es el inconsciente el que es misterioso. La gente que trabaja en IA afirma a menudo que en cuanto podamos comprender conscientemente alguna línea de pensamiento también podremos ver como hacer un computador capaz de seguirla; es de los misteriosos procesos inconscientes de los que no tenemos (¡aún!) idea de como tratar. Mi propia línea argumental ha sido que los procesos inconscientes podrían perfectamente ser algorítmicos, pero en un nivel muy complicado que sería monstruosamente difícil desentrañar en detalle. El pensamiento completamente consciente que puede racionalizarse como algo enteramente lógico puede otra vez (a menudo) formalizarse como algo algorítmico, pero esto se hace a un *nivel completamente diferente*. No estamos pensando ahora en las operaciones internas (disparo de neuronas, etc.) sino en la manipulación de pensamientos globales. A veces, esta manipulación mental tiene un carácter algorítmico (como en la lógica primitiva: los antiguos silogismos griegos tal como fueron formalizados por Aristóteles o la lógica simbólica del matemático George Boole; cf. Gardner, 1958); y a veces no lo tiene (como en el teorema de Gödel y algunos de los ejemplos dados en el Capítulo 4). La *formación de juicios*, que estoy afirmando que es la impronta de la cons-

ciencia, es *ella misma* algo sobre lo que la gente de la IA no tendría ninguna idea de cómo programar en un ordenador.

Se objeta a veces que, después de todo, los *criterios* para estos juicios no son conscientes, de modo que ¿por qué estoy atribuyendo estos juicios a la consciencia? Esto sería no captar el punto esencial de las ideas que estoy tratando de expresar. No estoy pidiendo que comprendamos conscientemente *cómo* formamos nuestras impresiones y juicios conscientes; eso sería caer en la confusión de niveles a la que me acabo de referir. Las *razones* subyacentes a nuestras impresiones conscientes serían cosas no directamente accesibles a la consciencia. Éstas tendrían que considerarse en un nivel físico más profundo que los de los pensamientos reales de que tenemos consciencia. (¡Haré un intento en una hipótesis posterior!) Son las propias impresiones conscientes las que *son* los juicios (no-algorítmicos).

De hecho, en los últimos capítulos ha estado subyacente la cuestión de que parece haber algo *no-algorítmico* en nuestro pensamiento consciente. En particular, una consecuencia del argumento del Capítulo 4 concerniente especialmente al teorema de Gödel era que, al menos en matemáticas, la contemplación consciente puede a veces capacitarnos para comprobar la verdad de un enunciado de un modo en que no podría hacerlo un algoritmo. (Elaboraré este argumento un poco más adelante.) En realidad, ¡los algoritmos por sí mismos *nunca* comprueban la verdad! Sería tan fácil hacer que un algoritmo sólo produjera falsedades como hacer que produjera verdades. Necesitamos *intuiciones externas* para decidir la validez o no de un algoritmo (más adelante volveremos sobre esto). Estoy exponiendo aquí el argumento de que es esta capacidad para distinguir (o «intuir») en circunstancias apropiadas verdad de falsedad (¡y belleza de fealdad!) lo que constituye la impronta de la consciencia.

Aclararé, sin embargo, que no estoy dando a entender ninguna forma de «adivinación» mágica. ¡La consciencia no es de ninguna ayuda al tratar de conjeturar el número de la suerte en una lotería (limpia)! Me estoy refiriendo a los juicios que hacemos continuamente mientras estamos conscientes, reuniendo todos los hechos, impresiones sensoriales y experiencias recordadas que son de relevancia, y sopesando las cosas —incluso formando en ocasiones juicios inspirados. En principio hay información disponible suficiente para que se pueda hacer el juicio relevante, pero el proceso de formular el juicio apropiado, extrayendo lo que se necesita del amasijo de datos, puede ser algo para lo que no existe ningún procedimiento algorítmico evidente —o incluso, en caso de que exista uno, puede que no sea prác-

tico. Quizá tengamos una situación en la que, una vez hecho el juicio, puede haber más de proceso algorítmico (o quizá simplemente uno más fácil) en *verificar* que el juicio es exacto de lo que lo hay en la formación inicial de dicho juicio. Estoy conjeturando que, en tales circunstancias, la consciencia conseguiría su propósito mediante una forma de conjurar los juicios apropiados.

¿Por qué digo que la impronta de la consciencia es una formación de juicios no-algorítmica? La razón procede en parte de mi experiencia como matemático. Sencillamente no confío en mis acciones algorítmicas inconscientes cuando mi consciencia no les ha prestado la atención debida. Con frecuencia no hay nada físico en el algoritmo *como tal* cuando se ha realizado algún cálculo, pero ¿era el algoritmo *correcto* a elegir para el problema a resolver? En un sencillo ejemplo tendríamos que aprender las reglas algorítmicas para multiplicar dos números y también para dividir un número por otro (o podemos preferir recurrir a la ayuda de una calculadora de bolsillo algorítmica), pero ¿como sabemos si, para el problema a resolver, tendríamos que multiplicar o dividir los números? Para ello necesitamos *pensar* y hacer un juicio *consciente*. (Veremos dentro de un momento por qué tales juicios deben ser, al menos algunas veces, no-algorítmicos.) Por supuesto, una vez que hemos tratado de un gran número de problemas similares, la decisión sobre si multiplicar o dividir los números puede convertirse en una segunda naturaleza y puede ser ejecutada algorítmicamente —quizá por el cerebelo. En esta etapa ya no es necesaria la consciencia, y nos llegamos a sentir confiados para permitir a nuestra mente consciente vagar y contemplar otras materias— aunque, de cuando en cuando, podemos tener necesidad de verificar que el algoritmo no se ha desviado de alguna forma (quizá sutil).

Cosas de este mismo tipo están sucediendo continuamente en todos los niveles del pensamiento matemático. A menudo, cuando hacemos matemáticas, nos esforzamos por conseguir nuevos algoritmos, pero el propio intento no parece ser un procedimiento algorítmico. Una vez que se encuentra un algoritmo apropiado el problema está resuelto en cierto sentido. Además, el juicio matemático sobre si algún algoritmo es realmente preciso o apropiado es el tipo de cosas que requiere mucha atención consciente. Algo similar ocurriría en la discusión de los sistemas formales para las matemáticas que se describieron en el Capítulo 4. Podemos empezar con algunos axiomas de los que deben derivarse diversas proposiciones matemáticas. El último procedimiento puede ser realmente algorítmico, pero es necesario que se haga un juicio por un matemático consciente para decidir si los

axiomas son apropiados. El que estos juicios son necesariamente *no* algorítmicos debería hacerse más evidente a partir de la discusión que se hace dos secciones más adelante. Pero antes de llegar a esto consideraremos lo que podría ser un punto de vista más predominante sobre lo que nuestros cerebros están haciendo y cómo han aparecido.

¿Selección natural de algoritmos?

Si suponemos que la acción del cerebro humano, consciente o no, consiste simplemente en la ejecución de algún algoritmo muy complicado, entonces debemos preguntar cómo se formó un algoritmo de eficacia tan extraordinaria. La respuesta estándar, por supuesto, sería que surgió por «selección natural». Por lo que respecta a las criaturas con cerebros evolucionados, aquellos con los algoritmos más eficaces tendrían una mejor tendencia a sobrevivir y por ello, en general, tendrían más prole. Esta prole también tendería a portar algoritmos más eficaces que sus primos, puesto que heredaron de sus padres los ingredientes de estos mejores algoritmos; de este modo los algoritmos mejoraron poco a poco —no necesariamente de manera uniforme puesto que podría haber habido tropiezos importantes en su evolución— hasta que alcanzaron el importante status que encontramos (aparentemente) en el cerebro humano. (Compárese con Dawkins, 1986).

Incluso según mi propio punto de vista tendría que haber *algo* de verdad en esta imagen, puesto que concibo que gran parte de la acción del cerebro es realmente algorítmica y, como el lector habrá deducido de la discusión anterior, soy un creyente convencido en el poder de la selección natural. Pero no veo cómo la selección natural por sí sola pueda hacer evolucionar algoritmos que pudieran tener el tipo de juicios conscientes sobre la *validez* de otros algoritmos que parece que nosotros tenemos.

Imaginemos un programa ordinario de ordenador. ¿Cómo habría llegado a formarse? ¡Es evidente que no (directamente) por selección natural! Algún programador de ordenadores humano lo habría concebido y habría verificado que realiza correctamente las acciones que se supone que debe hacer. (En realidad, muchos programas de ordenador complicados contienen errores —normalmente menores, pero a menudo muy sutiles y que no salen a la luz excepto en circunstancias muy poco comunes. La presencia de tales errores no afecta substancialmente a mi argumento.) A veces un programa de ordenador po-

dría haber sido «escrito» por otro programa, digamos un programa de ordenador «maestro», pero en tal caso el propio programa maestro habría sido el producto del ingenio e intuición humanos; o el programa podría perfectamente ensamblarse a partir de ingredientes, algunos de los cuales son los productos de otros programas de ordenador. Pero en todos los casos la validez y la misma concepción del programa habría sido en última instancia responsabilidad de (al menos) una consciencia humana.

Podemos imaginar, por supuesto, que no es necesario que haya sido así y que, dado el tiempo suficiente, el programa de ordenador podría haber evolucionado espontáneamente por algún proceso de selección natural. Si creemos que las acciones de las consciencias de los programadores de ordenador son ellas mismas simples algoritmos, entonces debemos creer, en efecto, que los algoritmos *han* evolucionado de esta misma forma. Lo que me molesta de esto, sin embargo, es que la decisión sobre la validez de un algoritmo *no* es ella misma un proceso algorítmico! Ya hemos visto algo de esto en el Capítulo 2. (La cuestión de si una máquina de Turing se *parará* o no es algo que no puede decidirse algorítmicamente.) Para decidir si un algoritmo *funcionará* o no, necesitamos *intuiciones*, y no sólo otro algoritmo.

De todas formas, aún sería posible imaginar algún tipo de proceso de selección natural que fuera efectivo para producir algoritmos *aproximadamente* válidos. Sin embargo, yo personalmente encuentro esto muy difícil de creer. Cualquier proceso de selección natural de este tipo actuaría sólo sobre el *output* de los algoritmos * y no directamente sobre las ideas que subyacen a los algoritmos. Esto no sólo es extremadamente ineficaz; creo que sería totalmente impracticable. En primer lugar, no es fácil verificar cuál es realmente un algoritmo mediante un simple examen de su *output*. (Sería bastante sencillo construir dos simples y muy diferentes acciones de máquina de Turing para las que las cintas de salida no difieran hasta, digamos, el lugar 2^{65536} —¡y esta diferencia no se podría reconocer en toda la historia del universo!) Además, la más «ligera» mutación de un algoritmo (por ejemplo, un pequeño cambio en la especificación de una máquina de Turing o en su cinta *input*) tendería a hacerla totalmente inútil, y es difícil ver siquiera cómo de esta forma aleatoria podrían aparecer *mejoras* reales en los algoritmos. (Incluso las mejoras *deli-*

* Está también el asunto algo espinoso de si dos algoritmos deben ser considerados mutuamente equivalentes si simplemente sus *outputs* son iguales en lugar de serlo sus cálculos reales. Ver Capítulo 2, p. 85.

beradas son difíciles sin que estén disponibles los «significados». Esto está particularmente confirmado por los casos no poco frecuentes en los que un programa de ordenador complicado e inadecuadamente documentado necesita ser alterado o corregido y el programador original se ha marchado o quizá ha muerto. Antes de tratar de desentrañar todos los diversos significados e intenciones de los que el programa depende explícitamente, ¡es probablemente más fácil desecharlo sin más y empezar todo de nuevo!)

Quizá pudiera imaginarse una forma mucho más «robusta» de especificar algoritmos que no estuviera sujeta a las críticas anteriores. En cierto modo, esto es lo que yo mismo estoy diciendo. Las especificaciones «robustas» son las *ideas* que subyacen a los algoritmos. Pero las ideas son cosas que, por lo que sabemos, necesitan mentes conscientes para su manifestación. Hemos vuelto al problema de qué es realmente la consciencia y qué cosas puede hacer de las que son incapaces los objetos inconscientes —y cómo demonios la selección natural ha sido lo bastante inteligente para hacer evolucionar *esas* cualidades más importantes.

Los productos de la selección natural son realmente sorprendentes. El pequeño conocimiento que yo mismo he adquirido acerca de cómo funcionan los cerebros humanos —y, de hecho, cualquier otra cosa viviente— me deja casi perplejo de asombro y admiración. El funcionamiento de una neurona individual es extraordinario, pero las propias neuronas están organizadas en conjunto de una forma muy notable con un gran número de conexiones en el instante del nacimiento dispuestas para todas las tareas que sean necesarias más adelante. ¡No es sólo la propia consciencia lo que es notable, sino toda la parafernalia que parece ser necesaria para sostenerla!

Si alguna vez descubrimos en detalle cuál es la cualidad que permite a un objeto físico llegar a ser consciente, entonces sería concebible que pudiéramos ser capaces de construir tales objetos por nosotros mismos —aunque podrían no calificarse como «máquinas» en el sentido de la palabra que ahora entendemos. Podríamos imaginar que estos objetos tendrían una tremenda ventaja sobre nosotros, puesto que podrían diseñarse *específicamente* para la tarea a realizar, a saber, *alcanzar consciencia*. No tendrían que crecer a partir de una sola célula. No tendrían que arrastrar el «equipaje» de su ascendencia (las viejas e «inútiles» partes del cerebro o del cuerpo que sobreviven en nosotros a causa de los «accidentes» de nuestros remotos ancestros). Podríamos imaginar, a la vista de estas ventajas, que tales objetos podrían tener éxito en superar *efectivamente* a los seres humanos

en las tareas en las que (en opinión de gentes como yo mismo) los computadores algorítmicos están condenados a la subordinación.

Pero puede haber más que esto en el asunto de la consciencia. Quizá, de alguna manera, nuestra consciencia depende de nuestra herencia y los miles de millones de años de evolución *real* que hay tras nosotros. A mi modo de ver, hay todavía algo misterioso en la evolución con su aparente «andar a tientas» hacia algún propósito futuro. Al menos, *parece* que las cosas se organizan ellas mismas algo mejor de lo que «deberían» hacerlo sólo sobre la base de la evolución por ciego azar y selección natural. Podría perfectamente suceder que tales apariencias fueran bastante engañosas. Parece que hubiera algo en el modo de trabajar de las leyes de la física que permitiría que la selección natural sea un proceso mucho más eficaz de lo que sería con leyes arbitrarias. El aparente «andar a tientas inteligente» que resulta es un tema interesante y volveré brevemente a él más adelante.

La naturaleza no-algorítmica de la intuición matemática

Como he declarado antes, una buena parte de la razón para creer que la consciencia es capaz de influir en los juicios de verdad de una manera *no-algorítmica* deriva de la consideración del teorema de Gödel. Si podemos ver que el papel de la consciencia es no-algorítmico cuando formamos juicios *matemáticos*, en los que el cálculo y la demostración rigurosa constituyen un factor tan importante, entonces ciertamente podremos persuadirnos de que un ingrediente no-algorítmico semejante podría ser también crucial para el papel de la consciencia en circunstancias más generales (no matemáticas).

Recordemos el argumento, dado en el Capítulo 4 que establecía el teorema de Gödel y su relación con la computabilidad. Se demostró allí que *cualquiera* que sea el algoritmo (suficientemente extenso) que un matemático pudiera utilizar para establecer la verdad matemática —o, lo que es equivalente, cualquiera que sea el *sistema formal* que él * pudiera adoptar para proporcionar su criterio de verdad— habría siempre proposiciones matemáticas, como las proposiciones de Gödel $P_k(k)$ explícitas del sistema (cf. p. 146) para las que su algoritmo no puede proporcionar una respuesta. Si el funcionamiento de la mente de un matemático es completamente algorítmico, entonces el algoritmo (o sistema formal) que utilice realmente para formar sus juicios

* «Él» significa, por supuesto, «él» o «ella». Ver nota en p. 27.

no es capaz de tratar la proposición $P_k(k)$ construida a partir de su algoritmo personal. De todas formas, ¡*nosotros* podemos ver (en principio) que $P_k(k)$ es realmente *verdadera*! Esto parecería presentarle una contradicción puesto que *él* también debería ser capaz de verlo. ¡Quizá esto indique que el matemático *no* estaba utilizando en absoluto un algoritmo!

Sin embargo, para convencernos *realmente* a nosotros mismos de la verdad de $P_k(k)$ necesitaríamos *saber* cuál es realmente el algoritmo del matemático, y también estar convencidos de su validez como medio de llegar a la verdad matemática. Como los defensores de la IA fuerte se apresuran en señalar, si el matemático estuviera utilizando un algoritmo muy complicado dentro de su cabeza no tendríamos *ninguna posibilidad* de conocer realmente cuál es dicho algoritmo y, por consiguiente, no seríamos capaces de construir realmente su proposición de Gödel, y mucho menos convencernos de su validez. Este tipo de objeción se plantea a menudo contra las afirmaciones como la que estoy haciendo aquí de que el teorema de Gödel indica que los juicios matemáticos humanos son no-algorítmicos. Pero yo personalmente no encuentro convincente esta objeción. Supongamos, por el momento, que los matemáticos humanos forman sus juicios conscientes sobre la verdad matemática de una manera realmente algorítmica. Trataremos de reducir esto a un absurdo (*reductio ad absurdum*!) utilizando el teorema de Gödel.

Debemos considerar en primer lugar la posibilidad de que diferentes matemáticos utilicen algoritmos *no equivalentes* para decidir la verdad. ¡Sin embargo, una de las características más sorprendentes de la matemática (quizá casi única entre las disciplinas) es que la verdad de las proposiciones puede efectivamente establecerse mediante argumentos abstractos! Un argumento matemático que convence a un matemático —siempre que no contenga error— también convencerá a otro en cuanto el argumento haya sido completamente captado. Esto también se aplica a las proposiciones tipo-Gödel. Si el primer matemático esta dispuesto a aceptar todos los axiomas y reglas de inferencia de un sistema formal particular que produzca sólo proposiciones *verdaderas*, entonces también debe estar dispuesto a aceptar que su proposición de Gödel describe una proposición verdadera. Sucedería exactamente lo mismo con el segundo matemático. El punto esencial es que los argumentos que establecen la verdad matemática son *comunicables* ².

Por lo tanto, no estamos hablando de varios algoritmos oscuros que podrían estar ejecutándose en las cabezas de distintos matemáti-

cos particulares. Estamos hablando de *un* sistema formal empleado universalmente que *es* equivalente a *todos* los diferentes algoritmos de los matemáticos para juzgar la verdad matemática. Ahora bien, ¡nisi siquiera puede saberse si este presunto sistema «universal», o algoritmo, es el que utilizamos nosotros los matemáticos para decidir la verdad! En efecto, si lo fuera *podríamos* construir su proposición de Gödel y saber que es también una verdad matemática. Por lo tanto, llegamos a la conclusión de que el algoritmo que utilizan realmente los matemáticos para decidir la verdad matemática es tan complicado u oscuro que nunca podremos conocer su propia validez.

¡Pero esto hace caso omiso de lo que es la esencia de la matemática! Toda la gracia de nuestra herencia y aprendizaje matemático es que *no* nos rendimos a la autoridad de algunas reglas oscuras que nunca podemos tener esperanza de comprender. Debemos *ver* —al menos en principio— que cada paso en el argumento puede reducirse a algo simple y obvio. La verdad matemática no es un dogma terriblemente complicado cuya validez está más allá de nuestra comprensión. Es algo construido a partir de semejantes ingredientes simples y obvios —y, cuando los comprendemos, su verdad es evidente y reconocida por todos.

A mi modo de ver, ésta es una *reductio ad absurdum* tan patente como la que se pueda conseguir por otro medio, ¡aunque esté lejos de ser una auténtica demostración matemática! El mensaje debería estar claro. La verdad matemática *no* es algo que averigüemos simplemente utilizando un algoritmo. Creo, también, que nuestra *consciencia* es un ingrediente crucial en nuestra comprensión de la verdad matemática. Debemos «ver» la verdad de un argumento matemático para estar convencidos de su validez. Esta «visión» es la esencia misma de la consciencia. Debe estar presente *dondequiera* que percibimos directamente la verdad matemática. Cuando nosotros mismos nos convencemos de la validez del teorema de Gödel no sólo lo «vemos» sino que al hacerlo revelamos la naturaleza misma no-algorítmica del propio proceso de la «visión».

Inspiración, intuición directa y originalidad

Intentaré hacer algunos comentarios sobre estos ocasionales soplos de nueva intuición que conocemos como inspiración. ¿Son estas ideas e imágenes que proceden misteriosamente de la mente inconsciente, o son, en algún sentido fundamental, producto de la propia consciencia?

cia? Podríamos citar muchos ejemplos en los que grandes pensadores han documentado tales experiencias. Como matemático estoy especialmente interesado en el pensamiento inspirado y original de otros matemáticos, pero imagino que hay mucho en común entre las matemáticas y las otras ciencias y artes. Para un informe muy hermoso remito al lector al librito *The Psychology of Invention in the Mathematical Field*, un clásico del distinguido matemático francés Jacques Hadamard. Él cita numerosas experiencias de inspiración descritas por matemáticos de primera fila, entre otras personas. Una de las más conocidas de éstas la proporciona Henri Poincaré. Poincaré describe, en primer lugar, cómo él tenía períodos intensivos de esfuerzo consciente y deliberado en su área de investigación que llamó funciones fuchsianas, pero había llegado a un punto muerto. Entonces:

... Dejé Caen, en donde vivía, para participar en una excursión geológica organizada por la Escuela de Minas. Las peripecias del viaje me hicieron olvidar mi trabajo matemático. Al llegar a Coutances subimos a un autobús para ir a algún lugar. En ese momento, cuando puse mi pie en el estribo, me vino la idea, sin que nada en mis pensamientos anteriores pareciera haber preparado el camino para ello, de que las transformaciones que yo había utilizado para definir las funciones fuchsianas eran idénticas a las de la geometría no euclídea. No verifiqué la idea; no hubiera tenido tiempo ya que en cuanto tomé asiento en el autobús continué una conversación ya comenzada, pero tenía la certidumbre absoluta. A mi vuelta a Caen, y para quedarme tranquilo, verifiqué el resultado detenidamente.

Lo sorprendente de este ejemplo (y otros muchos citados por Hadamard) es que esta idea complicada y profunda le llegó a Poincaré aparentemente en un soplo, mientras su pensamiento consciente parecía estar en otra parte, y que iba acompañada por esta sensación de certidumbre de que era correcta —como, de hecho, demostraron los cálculos posteriores. Debería quedar claro que la idea misma no sería en absoluto algo fácil de explicar con palabras. Imagino que para dar una idea apropiada a personas ya expertas él habría necesitado aproximadamente un seminario de una hora larga. Evidentemente sólo pudo entrar de golpe en la consciencia de Poincaré, completamente formada, debido a las muchas horas previas de actividad consciente deliberada que le familiarizó con muchos aspectos diverso del problema entre manos. Pero, en cierto sentido, la idea que tuvo Poincaré mientras subía al autobus era una idea «simple», ¡capaz de ser completamente captada en un instante! Aún más nota-

ble era la convicción que tuvo Poincaré de la verdad de la idea, de modo que su posterior verificación detallada parecía casi superflua.

Quizá debería tratar de relacionar esta experiencia con experiencias más propias que podrían ser en cierto modo comparables. En realidad no puedo encontrar ninguna ocasión en que me haya venido una idea tan completamente caída del cielo como parece que lo hizo en el ejemplo de Poincaré (o como en otros muchos ejemplos que se mencionan de auténtica inspiración). En mi caso parece necesario que yo *esté* pensando (quizá vagamente) en el problema entre manos conscientemente, aunque tal vez en un bajo nivel en el fondo de mi mente. También podría ser que esté ocupado en alguna otra actividad más bien relajante; el afeitado sería un buen ejemplo. Probablemente estuviera empezando a pensar en un problema que había dejado de lado durante un rato. Serían ciertamente necesarias las muchas horas difíciles de actividad consciente deliberada, y a veces me llevaría un rato volver a ponerme al corriente con el problema. Pero la experiencia de una idea que viene «en un soplo» bajo tales circunstancias —junto con una fuerte sensación de convicción de su validez— no me es desconocida.

Quizá valga la pena relatar un ejemplo concreto de esto que tiene un curioso punto de interés añadido. En el otoño de 1964 había estado ocupado en el problema de las singularidades de los agujeros negros. Oppenheimer y Snyder habían demostrado, en 1939, que un colapso exactamente *esférico* de una estrella de gran masa conduciría a una singularidad central en el espacio-tiempo, en la que la teoría clásica de la relatividad general se lleva más allá de sus límites (ver Capítulo 7, p. 415). Muchos pensaban que esta desagradable conclusión podría evitarse si se eliminara su (poco razonable) hipótesis de simetría esférica *exacta*. En el caso esférico toda la materia se dirige hacia un mismo punto central en donde, tal vez no muy inesperadamente si se tiene en cuenta esta simetría, se produce una singularidad de densidad infinita. Parecía que no era irrazonable suponer que *sin* dicha simetría la materia llegaría a la región central de una forma más desordenada y no aparecería ninguna singularidad de densidad infinita. Quizá, incluso, la materia se desplegaría de nuevo para dar un comportamiento muy diferente del de el agujero negro idealizado de Oppenheimer y Snyder ³.

Mis propias ideas habían sido estimuladas por el renovado interés en el problema del agujero negro que se derivaba del muy reciente descubrimiento de los cuásares (a comienzos de los años 60). La naturaleza física de estos objetos astronómicos, sorprendentemente bri-

llantes y lejanos, había llevado a mucha gente a especular que en sus centros podría residir algo semejante a los agujeros negros de Oppenheimer y Snyder. Por otro lado, muchos pensaban que la hipótesis de simetría esférica de Oppenheimer-Snyder podría proporcionar una imagen completamente errónea. No obstante, tuve la idea (a partir de la experiencia de un trabajo que había hecho en otro contexto) de que debería poder demostrarse un teorema matemático exacto que mostrara que las singularidades espacio-temporales deberían ser *inevitables* (según la teoría estándar de la relatividad general), y por lo tanto reivindicar la imagen del agujero negro —con tal de que el colapso hubiera alcanzado un «punto de no retorno». No conocía ningún criterio definible matemáticamente para un «punto de no retorno» (sin utilizar simetría esférica) ni mucho menos ningún enunciado o demostración de un teorema apropiado. Un colega (Ivor Robinson) había venido a visitarme desde los EE.UU. y habíamos iniciado una animada conversación sobre un tema muy diferente mientras paseábamos por la calle hacia mi despacho en el Birbeck College en Londres. La conversación se interrumpió momentáneamente al cruzar una calle secundaria y se reanudó al otro lado. ¡Por lo visto, durante esos breves instantes, se me ocurrió una idea pero luego la reanudación de la conservación la borró de mi mente!

Ese mismo día, después de que mi colega hubiera partido, volví a mi despacho. Recuerdo que tenía una extraña sensación de júbilo que no podía explicar. Empecé a repasar en mi mente todas las cosas que me habían sucedido durante el día, intentando encontrar qué era lo que había causado este júbilo. Después de eliminar muchas posibilidades inadecuadas me vino finalmente a la mente la idea que había tenido al cruzar la calle —¡una idea que me había proporcionado una momentánea alegría dándome la solución al problema que había estado dando vueltas en el fondo de mi cabeza! Aparentemente ése era el criterio necesario —que posteriormente denominé una «superficie atrapada»— y entonces no me llevó mucho tiempo construir las líneas generales de una demostración del teorema que había estado buscando (Penrose 1965). Aun así, pasó algún tiempo antes de que la demostración estuviese formulada de un modo completamente riguroso, pero la idea que había tenido mientras cruzaba la calle había sido la clave. (A veces me pregunto que habría pasado si me hubiera sucedido alguna otra experiencia gozosa durante ese día. ¡Quizá no hubiera recordado nunca la idea de la superficie-atrapada!)

Esta anécdota me lleva a otro tema concerniente a la inspiración y la intuición, a saber: que los criterios *estéticos* son enormemente

valiosos al formar nuestros juicios. En las artes se podría decir que los criterios estéticos son soberanos. La estética en las artes es un tema sofisticado, y los filósofos han dedicado vidas enteras a su estudio. Podría argumentarse que en matemáticas y en las ciencias tales criterios son meramente secundarios, siendo soberano el criterio de *verdad*. Sin embargo, parece que es imposible separar uno de otro cuando consideramos los temas de la inspiración e intuición. Mi impresión es que la fuerte convicción de la *validez* de un soplo de inspiración (no 100 por cien fiable, añadiría, pero al menos mucho más fiable que el simple azar) está ligada muy estrechamente con sus cualidades estéticas. Una idea bella tiene mucha mayor probabilidad de ser correcta que una idea fea. Ésa ha sido al menos mi propia experiencia, y sentimientos similares han sido expresados por otros (cf. Chandrasekhar 1987). Por ejemplo, Hadamard (1945, p. 31) escribe:

... es evidente que no puede tener lugar ningún descubrimiento o invención importante sin la *voluntad* de encontrar. Pero con Poincaré vemos algo más: que la intervención del sentido de belleza juega su parte como un *medio* indispensable para encontrar. Hemos llegado a una conclusión doble:

que la invención es elección

que esta elección está gobernada imperiosamente por el sentido de la belleza científica.

También Dirac (1982), por ejemplo, afirma abiertamente que fue su *agudo sentido de la belleza* el que le hizo capaz de descubrir su ecuación para el electrón (la «ecuación de Dirac» mencionada en p. 364), mientras que otros la habían buscado en vano.

Yo puedo ciertamente responder de la importancia de las cualidades estéticas en mi propio trabajo, tanto en lo que se refiere a la «convicción» que se sentirá con ideas que podrían calificarse posiblemente como «inspiradas» como en lo referente a las conjeturas más «rutinarias» que tenemos que hacer continuamente a medida que sentimos que nos acercamos a algún objetivo deseado. He escrito en otro lugar sobre este tema, en particular en relación con el descubrimiento de las teselas aperiódicas descritas en las Figs. 10.3 y 4.11. Indudablemente fueron las cualidades estéticas de la primera de estas pautas de teselación —no solo su apariencia visual, sino también sus intrigantes posibilidades matemáticas— las que habían permitido que me viniera la intuición (¡probablemente en un «soplo» pero con sólo un 60 por ciento de certidumbre!) de que su disposición podía ser forzada mediante adecuadas reglas de ajuste (esto es, ensamblaje en anzuelo).

Dentro de un momento veremos más sobre estas pautas de teselación. (Cf. Penrose, 1974).

Me parece evidente que la importancia de los criterios estéticos no sólo se aplica a los juicios de inspiración instantáneos sino también a los juicios mucho más frecuentes que hacemos todo el tiempo en el trabajo matemático (o científico). ¡Los argumentos rigurosos constituyen normalmente el último paso! Antes de ello tenemos que hacer muchas conjeturas y, para estas, las convicciones estéticas son enormemente importantes —siempre limitadas por el razonamiento lógico y los hechos conocidos.

Son estos juicios lo que considero la impronta del pensamiento consciente. Mi conjetura es que, incluso para el repentino golpe de intuición, aparentemente producido ya listo por la mente inconsciente, es la *consciencia* la que es el árbitro, y la idea será rápidamente rechazada y olvidada si no «sonase cierta». (Curiosamente, yo *olvidé* realmente mi superficie atrapada, pero eso no ocurre en el nivel que entiendo ahora. La idea irrumpió en la consciencia durante el tiempo suficiente como para dejar una impresión duradera.) El rechazo «estético» al que me estoy refiriendo podría ser tal, estoy suponiendo, que prohibiera que las ideas poco atractivas alcancen cualquier nivel apreciablemente permanente de la consciencia.

¿Cuál es entonces mi opinión sobre el papel del *inconsciente* en el pensamiento inspirado? Admito que estas cuestiones no son tan claras como me gustaría que fueran. Esta es un área en la que el inconsciente parece estar jugando realmente un papel esencial, y debo coincidir en la opinión de que los procesos inconscientes son importantes. Debo estar de acuerdo, asimismo, en que no puede tratarse simplemente de que la mente inconsciente esté lanzando ideas aleatoriamente. Debe haber un proceso de selección enormemente poderoso que permite que la mente consciente sea perturbada sólo por ideas que «tienen alguna posibilidad». Sugeriría que estos criterios de selección —principalmente los «estéticos», de alguna especie— han sido ya fuertemente influenciados por los desiderata conscientes (como la sensación de fealdad que acompaña a las ideas matemáticas que son inconsistentes con los principios generales ya establecidos).

En relación con esto debería plantearse la cuestión de lo que constituye la auténtica *originalidad*. Me parece que aquí hay dos factores implicados, a saber: un proceso de «propuesta» y uno de «rechazo». Pienso que la propuesta podría ser fundamentalmente inconsciente y el rechazo fundamentalmente consciente. Sin un proceso de propuesta efectivo no tendríamos ideas nuevas en absoluto. Pero este procedi-

miento tendría poco valor por sí solo. Necesitamos un procedimiento efectivo para formar juicios, de modo que sólo sobrevivirán aquellas ideas con una oportunidad de éxito razonable. En los sueños, por ejemplo, pueden venir fácilmente a la mente ideas inusuales, pero sólo muy raramente sobrevivirán al juicio crítico de la consciencia despierta. (En mi caso nunca me ha venido una idea científica acertada en un sueño, mientras que otros, como el químico Kekulé con su descubrimiento de la estructura del benceno, parecen haber tenido más fortuna.) En mi opinión, es el proceso (juicio) consciente de rechazo, más que el proceso inconsciente de propuesta, el que es capital en el tema de la originalidad; pero sé que muchos otros podrán mantener una opinión contraria.

Antes de dejar las cosas en este estado más bien insatisfactorio mencionaré otra característica sorprendente del pensamiento inspirado, a saber: su carácter *global*. La anécdota anterior de Poincaré era un ejemplo sorprendente puesto que la idea que vino a su mente en un momento fugaz ha dominado un área enorme de pensamiento matemático. Quizá sea más inmediatamente accesible al lector no matemático (aunque no más comprensible, sin duda) el modo en que algunos artistas pueden tener en su mente a la vez la totalidad de sus creaciones. Mozart da vivamente un ejemplo sorprendente (citado en Hadamard, 1945, p. 16).

Cuando me siento bien y de buen humor, o cuando estoy dando un paseo en carruaje o a pie tras una buena comida, o por la noche cuando no puedo dormir, las ideas se agolpan en mi mente tan fácilmente como desee. ¿De dónde y cómo vienen? No lo sé y no tengo nada que ver con ello. Guardo en la mente aquellas que me gustan y las tarareo; al menos eso me han dicho que hago. Una vez que tengo mi tema viene otra melodía ligada con la primera de acuerdo con las necesidades de la composición global: el contrapunto, la parte de cada instrumento y por último todos los fragmentos melódicos dan lugar a la obra completa. Entonces mi alma arde de inspiración. La obra crece; sigo desarrollándola, concibiéndola cada vez con más claridad hasta que tengo acabada en mi cabeza la composición entera por larga que pueda ser. Entonces mi mente la atrapa de la misma forma como una mirada de mi ojo atrapa una imagen bella o un joven apuesto. No viene a mí poco a poco, con las diversas partes trabajadas en detalle a medida que se van haciendo, sino que es en su totalidad como me deja oír las mi imaginación.

Me parece que esto concuerda con un esquema de propuesta/rechazo. La propuesta parece ser inconsciente («no tengo nada que ver con ello») aunque, sin duda, fuertemente selectiva, mientras que el

rechazo es el árbitro consciente del gusto («guardo aquellas que me gustan...»). La globalidad del pensamiento inspirado es particularmente notable en la cita de Mozart («no viene a mí poco a poco... sino en su totalidad»). Además, sostendré que una notable globalidad está ya presente en general en nuestro pensamiento consciente. Volveré a esta cuestión en un momento.

La no verbalidad del pensamiento

Una de las principales puntualizaciones que hace Hadamard en su estudio sobre el pensamiento creativo es una impresionante refutación de la tesis, tan a menudo expresada todavía, de que la verbalización es necesaria para el pensamiento. Difícilmente podríamos hacer algo mejor que repetir una cita de una carta que recibió de Albert Einstein a propósito de la cuestión:

Las palabras o el lenguaje, ya sea escrito o hablado, no parecen jugar ningún papel en mi mecanismo de pensamiento. Las entidades físicas que parecen servir como elementos del pensamiento son ciertos signos e imágenes más o menos claras que pueden reproducirse y combinarse «voluntariamente»... Los elementos antes mencionados son, en mi caso, de tipo visual y muscular. Las palabras u otros signos convencionales tienen que buscarse laboriosamente sólo en una segunda etapa, cuando el citado juego asociativo está suficientemente establecido y puede ser reproducido a voluntad.

También merece ser citado el eminente geneticista Francis Galton:

Resulta una seria desventaja para mí al escribir, y aún más al explicarme directamente, el hecho de que no pienso tan fácilmente en palabras como en otras formas. Sucede a menudo que después de hacer un trabajo duro, y habiendo llegado a resultados que son perfectamente claros y satisfactorios para mí, cuando trato de expresarlos en el lenguaje siento que debo empezar a colocarme en un plano intelectual muy distinto. Tengo que traducir mis ideas en un lenguaje que no corre muy a la par con ellas. Por ello gasto mucho tiempo en buscar palabras y frases apropiadas, y soy consciente, cuando tengo que hablar de improviso, de ser a menudo muy oscuro por simple torpeza verbal y no por falta de claridad de percepción. Éste es uno de los pequeños fastidios de mi vida.

El propio Hadamard escribe también:

Insisto en que las palabras están totalmente ausentes de mi mente cuando realmente pienso, y mi caso está en la misma línea de Galton en el sentido de que, incluso después de leer o escuchar una pregunta, todas las palabras desaparecen en el preciso instante en que empiezo a pensar sobre ello; y coincido plenamente con Schopenhauer cuando escribe «las ideas mueren en el momento en que se encarnan en palabras».

Cito estos ejemplos porque se ajustan mucho a mis propios modos de pensamiento. Casi todo mi pensamiento matemático se hace visualmente y en términos de conceptos no verbales, aunque con frecuencia las ideas vayan acompañadas de algún inane y casi inútil comentario verbal, tal como «esto va con esto y aquello va con eso otro». (Pudiera utilizar a veces palabras para simples inferencias lógicas.) También he experimentado frecuentemente por mí mismo las dificultades que tenían estos pensadores para traducir sus pensamientos en palabras. Con frecuencia la razón es simplemente que no hay palabras disponibles para expresar los conceptos requeridos. De hecho, yo calculo a menudo utilizando diagramas especialmente diseñados que constituyen una especie de taquigrafía para ciertos tipos de expresiones algebraicas (cf. Penrose y Rindler, 1984, pp. 424-434). Sería así un proceso muy molesto tener que traducir estos diagramas en palabras, y esto es algo que sólo hago como último recurso si se hace necesario para dar una explicación detallada a los demás. En relación con esto he notado, en ocasiones, que si me he estado concentrando duramente en matemáticas durante un rato y alguien inicia repentinamente una conversación, entonces me encuentro casi incapaz de hablar durante varios segundos.

Esto no quiere decir que no piense a veces con palabras, sino que simplemente encuentro las palabras casi inútiles para el pensamiento *matemático*. Otros tipos de pensamiento, quizá tales como el *filosofar*, parecen mucho más adecuados para la expresión verbal. ¡Quizá sea ésta la razón del por qué muchos filósofos parecen ser de la opinión de que el lenguaje es esencial para el pensamiento inteligente o consciente! Sin duda, personas diferentes piensan de modos muy diferentes —como ha sido ciertamente mi propia experiencia, incluso entre matemáticos. La polaridad analítico/geométrico parece ser la principal en el pensamiento matemático. Es interesante que Hadamard se considerase a sí mismo en el lado analítico pese a que él utilizaba imágenes visuales más que verbales para su pensamiento matemático. En cuanto a mí mismo, estoy casi en el extremo geométrico de las cosas, pero el espectro entre los matemáticos es generalmente muy amplio.

Una vez que se acepta que mucho del pensamiento consciente puede ser de carácter no verbal —y, en mi opinión, esta conclusión es una conclusión inevitable de las anteriores consideraciones— entonces quizá el lector no encuentre tan difícil creer que tal pensamiento ¡pudiera tener también un ingrediente no-algorítmico!

Recuérdese que en el Capítulo 9 (p. 476) mencioné el punto de vista expresado con frecuencia de que sólo la mitad del cerebro que es capaz del habla (la mitad izquierda, en la inmensa mayoría) sería también capaz de la consciencia. En vista de las discusiones anteriores, estará claro para el lector por qué encuentro este punto de vista totalmente inaceptable. Yo no sé si, en general, los matemáticos tenderán a utilizar más un lado u otro de sus cerebros; pero no puede haber duda del alto nivel de consciencia que se necesita para el genuino pensamiento matemático. Mientras que el pensamiento analítico parece ser principalmente competencia de lado izquierdo del cerebro, se argumenta a menudo que el pensamiento geométrico reside en el lado *derecho*, de modo que es una conjetura muy razonable el que mucha de la actividad matemática *consciente* ¡realmente *tenga lugar* en el lado derecho!

¿Consciencia animal?

Antes de dejar el tema de la importancia de la verbalización para la consciencia abordaré la cuestión, antes brevemente planteada, de si los animales no-humanos pueden ser conscientes. Me parece que a veces la gente se basa en la incapacidad para hablar de los animales como un argumento en contra de que puedan tener ninguna consciencia apreciable —y, por implicación, en contra de que tengan «derechos». El lector puede percibir perfectamente que considero esto una línea de razonamiento insostenible, ya que mucho pensamiento consciente satisficado (v.g. el matemático) puede llevarse a cabo sin verbalización. También se argumenta a veces que el lado derecho del cerebro tiene una consciencia «pequeña» como la de un chimpancé, debido también a su falta de capacidad verbal (cf. LeDoux, 1985, pp. 197-216).

Existe una considerable controversia sobre si los chimpancés y los gorilas son realmente capaces de genuina verbalización cuando se les permite utilizar un *lenguaje de signos* en lugar de hablar de la forma humana normal (lo que ellos no pueden hacer debido a la falta de cuerdas vocales adecuadas). (Ver diversos artículos en Blakemore y

Greenfield, 1987.) Parece claro, pese a la controversia, que estos animales son capaces de comunicar, al menos hasta cierto grado elemental, por estos medios. A mi modo de ver, resulta un poco cicatero por parte de algunas personas el no admitir que esto pueda llamarse «verbalización». ¡Quizá, al negar a los monos la entrada en el club de los verbalizadores, algunos esperan excluirlos del club de los seres conscientes!

Dejando aparte la cuestión del habla, existe una buena evidencia de que los chimpancés son capaces de genuina *inspiración*. Konrad Lorenz (1972) describe un chimpancé en una habitación en la que había un plátano suspendido del techo apenas fuera de su alcance, y una caja en otro lado de la habitación:

La cuestión no le dejaba en paz y volvió otra vez a ello. Entonces, súbitamente —y no hay otra forma de describirlo— su cara antes deprimida «se iluminó». Sus ojos se movían ahora desde el plátano hacia el espacio vacío que había bajo él en el suelo, y desde éste a la caja, luego volvía el espacio y de allí al plátano. A continuación dio un grito de alegría y luego dio una voltereta sobre la caja, de pura animación. Completamente seguro de su éxito empujó la caja hasta ponerla debajo del plátano. Nadie que lo observara podría dudar de la existencia de una genuina experiencia «Ajá» en los simios antropoides.

Una cuestión interesante se plantea en relación con los delfines (y las ballenas). Puede señalarse que los cerebros de los delfines son tan grandes (o mayores) que los nuestros, y los delfines pueden enviar también complejas señales sonoras de uno a otro. Muy bien pudiera ser que sus grandes cerebros fueran necesarios para algún propósito distinto de la «inteligencia» a una escala humana o casi humana. Además, debido a su falta de manos prensiles, no son capaces de construir una «civilización» del tipo de las que podemos apreciar —y aunque, por la misma razón, no puedan escribir libros, ¡podrían a veces ser filósofos y meditar sobre el significado de la vida y por qué están «allí»! ¿Podrían transmitir a veces sus sensaciones de «consciencia» por vía de sus complejas señales sonoras submarinas? No sé de ninguna investigación que indique si ellos utilizan un lado particular de sus cerebros para «verbalizar» y comunicar con otros. Respecto a las operaciones de «escisión cerebral» que se han realizado en humanos, con sus enigmáticas implicaciones sobre la continuidad del «yo», debería subrayarse que no todo el cerebro del delfín duerme simultáneamente sino que cada vez duerme un lado del cerebro. ¡Sería ins-

tructivo para nosotros si pudiéramos preguntarles cómo «sienten» ellos la continuidad de la consciencia!

Contacto con el mundo de Platón

He mencionado que parece haber muchas diferentes maneras de pensar para personas diferentes —e incluso diferentes maneras de pensar sobre matemáticas para matemáticos diferentes. Recuerdo que cuando estaba a punto de entrar en la universidad para hacer mis estudios esperaba encontrar que los otros, quienes iban a ser mis colegas matemáticos, pensarían más o menos como yo lo hacía. Mi experiencia en la escuela había sido que mis compañeros de clase parecían pensar de una forma bastante diferente de la mía, lo que había encontrado bastante desconcertante. ¡«Ahora», me había dicho entusiasmado, «encontraré colegas con quienes pueda comunicar libremente! Algunos pensarán con más eficacia que yo, y algunos con menos; pero todos ellos estarán en mi misma onda de pensamiento». ¡Qué equivocado estaba! ¡Creo que encontré más diferencias en cuanto a modos de pensar de las que había experimentado hasta entonces! Mi propio pensamiento era mucho más geométrico y menos analítico que el de los demás, pero había otras muchas diferencias entre las formas de pensar de mis diversos colegas. Siempre tuve una dificultad particular en comprender una descripción verbal de una fórmula, mientras que muchos de mis colegas parecían no experimentar tal dificultad.

Una experiencia común, cuando algún colega trataba de explicarme algún punto de las matemáticas, era que a menudo lo escuchaba atentamente pero sin comprender prácticamente nada de las conexiones lógicas entre un conjunto de palabras y el siguiente. Sin embargo, en mi mente se formaba cierta imagen conjeturada para las ideas que él estaba tratando de mostrarme —formada completamente en mis propios términos y con poca conexión aparente con las imágenes mentales que habían sido la base de la propia comprensión de mi colega— y yo respondía. Para mi asombro, mis propios comentarios eran aceptados normalmente como apropiados, y la conversación continuaba así de uno a otro. Estaba claro, al final de ello, que había tenido lugar alguna comunicación genuina y positiva. ¡Pero las sentencias reales que pronunciábamos cada uno de nosotros sólo muy infrecuentemente parecían ser comprendidas! En mis años siguientes como matemático (o físico matemático) profesional he encontrado que este fenómeno no es menos cierto de lo que lo era cuando era estudiante. Quizá, a

medida que ha aumentado mi experiencia matemática, he logrado imaginar mejor lo que los otros quieren dar a entender con sus explicaciones, y quizá soy un poco mejor para tener en cuenta los otros modos de pensar cuando yo mismo estoy explicándome las cosas. Pero, en esencia, nada ha cambiado.

Con frecuencia me he devanado la cabeza pensando en cómo es posible cualquier comunicación según este extraño procedimiento, pero ahora me gustaría aventurar una especie de explicación, pues pienso que posiblemente pudiera tener alguna profunda importancia para las otras cuestiones que he estado abordando. La idea es que al transmitir la matemáticas no *simplemente* comunicamos *hechos*. Para que se pueda comunicar de una persona a otra una cadena de hechos (contingentes) es necesario que los hechos sean cuidadosamente enunciados por la primera y que la segunda los pueda aceptar individualmente. Pero en las matemáticas el contenido *factual* es pequeño. Los enunciados matemáticos son verdades necesarias (¡o, por el contrario, falsedades necesarias!) e incluso si el enunciado del primer matemático representa meramente un tanteo de una tal verdad necesaria, será esa misma verdad la que se transmitirá al segundo matemático con tal de que éste la haya comprendido de forma adecuada. Las imágenes mentales del segundo pueden diferir de las del primero, igual que pueden diferir sus descripciones verbales, pero la idea matemática importante habrá pasado de uno a otro.

Ahora bien, este tipo de comunicación no sería posible en absoluto si no fuera por el hecho de que las verdades matemáticas *interesantes* o *profundas* están mínimamente distribuidas entre las verdades matemáticas en general. Si la verdad a ser transmitida fuera, pongamos por caso, el enunciado sin interés $4.897 \times 512 = 2.507.264$, entonces el segundo tendría que haber sido comprendido realmente por el primero para que el enunciado exacto sea transmitido. Pero para un enunciado matemático interesante podemos darnos cuenta a menudo del concepto que se intenta transmitir incluso si se ha proporcionado una descripción del mismo muy imprecisa.

Podría parecer que hay una paradoja en esto, ya que la matemática es un tema en donde la precisión es soberana. En realidad, en las presentaciones escritas se tiene mucho cuidado para estar seguro de que los diversos enunciados son a la vez precisos y completos. Sin embargo, para transmitir una idea matemática (normalmente en descripciones verbales), semejante precisión puede tener a veces un efecto inhibitorio a primera vista, y puede resultar necesaria una forma de comunicación más vaga y descriptiva. Una vez que la idea ha sido

captada en esencia es cuando los detalles pueden ser examinados con posterioridad.

¿Cómo es posible que se puedan comunicar las ideas matemáticas de esta forma? Imagino que siempre que la mente percibe una idea matemática, toma contacto con el mundo platónico de los conceptos matemáticos. (Recordemos que, según el punto de vista platónico, las ideas matemáticas tienen una existencia propia, y habitan en un mundo platónico ideal que sólo es accesible por la vía del intelecto; cf. pp. 135, 206-207.) Cuando «vemos» una verdad matemática, nuestra consciencia irrumpe en este mundo de ideas y toma contacto directo con él («accesible por vía del intelecto»). He descrito esta «visión» en relación con el teorema de Gödel pero es la esencia de la comprensión matemática. Cuando los matemáticos se comunican, esto se hace posible porque cada uno tiene un *camino directo a la verdad*, estando la consciencia de cada uno en disposición de percibir directamente las verdades matemáticas a través de este proceso de «visión». (De hecho, este acto de percepción va acompañado a menudo de palabras como ¡«Oh, lo veo!»!) Puesto que cada uno puede tomar contacto directamente con el mundo platónico, pueden comunicar entre ellos mucho más fácilmente de lo que pudieran esperar. Las imágenes mentales que tiene cada uno de ellos cuando hacen este contacto platónico podrían ser bastante diferentes para cada uno, pero la comunicación es posible porque ¡cada uno está en contacto directo con el *mismo* mundo platónico externamente existente!

Según esta idea, la mente es siempre capaz de este contacto directo. Pero sólo un poco puede trasvasar cada vez. El descubrimiento matemático consiste en un ensanchamiento del área de contacto. Debido al hecho de que las verdades matemáticas son verdades necesarias, ninguna «información» real, en el sentido técnico, pasa al descubridor. Toda la información estaba allí todo el tiempo; ¡era sólo cuestión de poner las cosas juntas y «ver» la respuesta! Esto coincide prácticamente con la propia idea de Platón que de ¡el descubrimiento (digamos matemático) es sólo una forma de *recuerdo*! En realidad, me he sorprendido a menudo de la similitud entre el hecho de no ser capaz de recordar simplemente el nombre de alguien y el de no ser capaz de encontrar simplemente el concepto matemático correcto. En ambos casos, el concepto buscado está en cierto sentido *ya presente* en la mente, aunque ésta es una forma menos usual de hablar en el caso de una idea matemática no descubierta.

Para que esta manera de ver las cosas sea útil, en el caso de la comunicación matemática, debemos imaginar que las ideas matemá-

ticas interesantes y profundas tienen de algún modo una existencia más fuerte que las poco interesantes o triviales. Esto tendrá importancia en relación con las consideraciones especulativas de la próxima sección.

Una visión de la realidad física

Cualquier punto de vista acerca de la aparición de la consciencia dentro del universo de la realidad física debe abordar, al menos implícitamente, la cuestión de la propia realidad física.

El punto de vista de la IA fuerte, por ejemplo, sostiene que una «mente» encuentra su existencia a través de la encarnación de un algoritmo suficientemente complejo, a medida que este algoritmo es activado por algunos objetos del mundo físico. Se supone que no importa cuales sean estos objetos reales. Igual valdrán señales nerviosas que corrientes eléctricas en cables, engranajes, poleas o tuberías. Todo lo que importa es el propio algoritmo. Pero para que un algoritmo «exista» independientemente de cualquier encarnación física particular parece ser esencial un punto de vista platónico. Sería difícil para un defensor de la IA fuerte aceptar la línea alternativa de que «los conceptos matemáticos existen sólo en las mentes», ya que esto sería un círculo vicioso que requiere mentes pre-existentes para la existencia de algoritmos y algoritmos pre-existentes para las mentes! Dicho defensor podría tratar de adoptar la postura de que los algoritmos pueden existir como marcas en un trozo de papel, o direcciones de magnetización en un bloque de hierro, o desplazamientos de carga en la memoria de un ordenador. Pero tales configuraciones de material no constituyen por sí mismas un algoritmo. Para convertirse en algoritmos necesitan tener una *interpretación*, es decir, debe ser posible *descodificar* las configuraciones; y ello dependerá del «lenguaje» en el que estén escritos los algoritmos. Una vez más, parece que se necesita una mente pre-existente para «comprender» el lenguaje, y volvemos donde estábamos. Aceptando, entonces, que los algoritmos habitan en el mundo platónico y, por lo tanto que en *dicho* mundo, según la idea de la IA fuerte, es donde deben encontrarse las mentes, tenemos que afrontar ahora la cuestión de cómo pueden relacionarse el mundo físico y el mundo platónico. Esta es a mi parecer la versión de la IA fuerte del problema mente-cuerpo.

Mi propio punto de vista es diferente de éste, puesto que yo creo

que las mentes (conscientes) *no* son entidades algorítmicas. No obstante, me siento algo desconcertado al descubrir que existen muchos puntos en común entre el punto de vista de la IA fuerte y el mío. He señalado que creo que la consciencia está estrechamente asociada con la sensación de verdades necesarias —y por consiguiente con la consecución de un contacto directo con el mundo platónico de los conceptos matemáticos. Esto no es un procedimiento algorítmico —y no es el algoritmo que pudiera habitar en ese mundo el que tiene interés especial para nosotros— pero de nuevo se ve que el problema mente-cuerpo, según esta idea, está íntimamente ligado a la cuestión de cómo se relaciona el mundo platónico con el mundo «real» de los objetos físicos verdaderos.

Hemos visto en los Capítulos 5 y 6 cómo el mundo físico real parece ajustarse de un modo notable a algunos esquemas matemáticos muy precisos (las teorías SOBERBIAS, cf. p. 199). Se ha subrayado a menudo cuán realmente extraordinaria es esta precisión (cf. especialmente Wigner, 1960). Me resulta difícil creer, como algunos tratan de sostener, que semejantes teorías SOBERBIAS pudiesen haber surgido simplemente por alguna selección natural de ideas que dejase sólo las buenas como supervivientes. Las buenas son sencillamente *demasiado* buenas para ser simplemente las supervivientes de ideas que hayan aparecido de modo aleatorio. En lugar de ello, debe haber alguna profunda razón subyacente para el acuerdo entre matemáticas y física, es decir, entre el mundo platónico y el mundo físico.

Al hablar del «mundo platónico» le estamos asignando algún tipo de realidad que es comparable de alguna forma a la realidad del mundo físico. Por otra parte, la realidad del propio mundo físico parece más nebulosa de lo que había parecido ser antes de la llegada de las teorías SOBERBIAS de la relatividad y la mecánica cuántica (ver los comentarios de las pp. 199, 200 y 361 especialmente). La misma exactitud de estas teorías ha proporcionado una existencia matemática casi abstracta para la verdadera realidad física. ¿Es esto de alguna forma una paradoja? ¿Cómo puede la realidad concreta hacerse abstracta y matemática? Esta es tal vez la otra cara de la moneda a la pregunta de cómo los conceptos matemáticos abstractos pueden conseguir una realidad casi concreta en el mundo platónico. ¿Quizá, en cierto sentido, los dos mundos son realmente el *mismo*? (Cf. Wigner, 1960; Penrose, 1979a; Barrow, 1988; también Atkins, 1987.)

Aunque siento fuertes simpatías por esta idea de identificar realmente estos dos mundos, debe haber algo más que eso. Como he

mencionado en el Capítulo 3, y en una sección precedente de este mismo capítulo, algunas verdades matemáticas parecen tener una realidad platónica más fuerte (¿ «más profunda», «más interesante», «más fructífera» ?) que otras. Tales verdades serían las que se identificarían más estrechamente con el funcionamiento de la realidad física. (El sistema de los números complejos (cf. Capítulo 3) sería un caso concreto por ser estos los ingredientes fundamentales de la mecánica cuántica, las amplitudes de probabilidad.) Con semejante identificación podría ser más comprensible cómo las «mentes» parecerían manifestar alguna curiosa conexión entre el mundo físico y el mundo platónico de las matemáticas. Recordemos también que, como se describió en el Capítulo 4, existen muchas partes del mundo matemático —algunas de sus partes más profundas y más interesantes, por lo demás— que no tienen carácter algorítmico. Parece probable, por lo tanto, sobre la base del punto de vista que trato de exponer, que la acción no-algorítmica debería tener un papel de importancia muy considerable dentro del mundo físico. Estoy sugiriendo que este papel está íntimamente ligado con el propio concepto de «mente».

Determinismo y determinismo fuerte

Hasta aquí poco he dicho sobre la cuestión del «libre albedrío», que normalmente se considera que es el tema fundamental del aspecto *activo* del problema mente-cuerpo. En lugar de ello, me he concentrado en mi sugerencia de que existe un aspecto *no-algorítmico* en el papel de la acción consciente. Normalmente, el tema del libre albedrío se discute en relación con el determinismo en física. Recordemos que en la mayoría de nuestras teorías SOBERBIAS existe un determinismo estricto, en el sentido de que si se conoce el estado del sistema en un instante cualquiera, entonces está completamente fijado para todos los instantes posteriores (o también anteriores) por las ecuaciones de la teoría. Por consiguiente, parece que no hay lugar para el «libre albedrío» puesto que el comportamiento futuro de un sistema parece estar totalmente determinado por las leyes físicas. Incluso la parte **U** de la mecánica cuántica tienen este carácter completamente determinista. Sin embargo, la parte **R** del «salto cuántico» no es determinista e introduce un elemento completamente aleatorio en la evolución temporal. Con anterioridad, diversas personas aventuraron ya la posibilidad de que pudiera haber un papel para el libre albedrío, y que la acción de la consciencia tuviera quizá algún efecto

directo sobre el modo en que podría saltar un sistema cuántico individual. Pero si **R** es *realmente* aleatorio, entonces no sirve de gran ayuda si deseamos hacer algo positivo con nuestras libres voluntades.

Mi propio punto de vista, aunque no esté muy bien formulado a este respecto, sería el de que algún nuevo procedimiento (GQC; cf. Capítulo 8) toma el mando a partir de la línea divisoria cuántico-clásico e interpola entre **U** y **R** (que son ahora ambos considerados como aproximaciones), y que este nuevo procedimiento contendría un elemento *esencialmente no-algorítmico*. Esto implicaría que el futuro *no sería computable* a partir del presente, incluso aunque *pudiera estar determinado* por él. He tratado de ser claro al distinguir el tema de la computabilidad del tema del determinismo en mi discusión en el Capítulo 5. Me parece bastante plausible que la GQC pudiera ser una teoría determinista pero no-computable *. (Recordemos el «modelo de juguete» no computable que describí en el Capítulo 5, p. 221.)

Hay quienes aceptan la idea de que ni siquiera con el determinismo clásico (o U-cuántico) existe determinismo *efectivo*, ya que las condiciones iniciales nunca podrían conocerse suficientemente bien para que se pudiera computar *realmente* el futuro. A veces, cambios muy pequeños en las condiciones iniciales pueden conducir a diferencias muy grandes en el resultado final. Esto es lo que sucede, por ejemplo, en el fenómeno conocido como «caos» en un sistema determinista (clásico) —siendo un ejemplo de ello la incertidumbre en la predicción del tiempo meteorológico. Sin embargo, es muy difícil creer que este tipo de incertidumbre clásica pudiera ser la que nos permite nuestra (¿ilusión de?) libre albedrío. El comportamiento futuro seguirá estando *determinado*, desde el mismo big bang, incluso aunque seamos incapaces de calcularlo (cf. p. 225).

La misma objeción podría levantarse contra mi hipótesis de que una falta de *computabilidad* podría ser intrínseca a las leyes dinámicas —que ahora se suponen de carácter no-algorítmico— antes que a nuestra falta de información acerca de las condiciones iniciales. En esta perspectiva, incluso aunque fuera no computable, el futuro seguiría completamente *fijado* por el pasado —en todo el camino hacia el big bang. De hecho, no voy a ser tan dogmático como para insistir en que la GQC debería ser determinista pero no computable. Mi conjetura sería que la teoría buscada tendría una descripción más sutil que

* Puede señalarse que hay al menos una aproximación a una teoría de gravitación cuántica que parece implicar un elemento de no-computabilidad (Geroch y Hartle, 1987).

esa. Sólo estoy planteando que debería contener elementos de un tipo esencialmente no-algorítmico.

Para cerrar esta sección me gustaría comentar una idea aún más extrema que se podría mantener sobre el tema del determinismo. Ésta es lo que he denominado *determinismo fuerte* (Penrose, 1987b). Según el determinismo fuerte no se trata sólo de que el futuro esté determinado por el pasado: la *historia entera del Universo está fijada*, según algún esquema matemáticamente preciso, *para cualquier instante*. Semejante punto de vista podría tener un atractivo si nos inclinamos a identificar de alguna forma el mundo platónico con el mundo físico, puesto que el mundo platónico está fijado de una vez por todas, ¡sin «posibilidades alternativas! para el Universo! (Me pregunto a veces si Einstein pudiera haber tenido este esquema en la mente cuando escribió: «Lo que realmente me interesa es si Dios podía haber hecho el mundo de una forma diferente; es decir, ¡si la necesidad de simplicidad lógica deja alguna libertad!» (carta a Ernst Strauss; ver Kuznetsov, 1977, p. 285).

Como una variante del determinismo fuerte podríamos considerar la idea de los *muchos-universos* en mecánica cuántica (cf. Capítulo 6, p. 372). Según ésta, no sería una *simple* historia de universo individual la que estaría fijada por un esquema matemáticamente preciso, sino que sería la totalidad de miríadas de historias del universo «posibles» las que estarían determinadas. Pese a la naturaleza desagradable (al menos para mí) de semejante esquema, y la multitud de problemas e insuficiencias que nos presenta, no puede ser descartado como posibilidad.

Me parece que si tenemos determinismo fuerte, pero *sin* muchos universos, entonces el esquema matemático que gobierna la estructura del Universo *tendría* que ser probablemente no-algorítmico ⁶; de otro modo podríamos calcular en principio lo que íbamos a hacer a continuación y entonces podríamos «decidir» hacer algo muy diferente, lo que constituiría una contradicción flagrante entre el «libre albedrío» y el determinismo fuerte de la teoría. Podemos evitar esta contradicción introduciendo no-computabilidad en la teoría —¡aunque tengo que confesar que me siento algo molesto con este tipo de solución, y espero algo mucho más sutil para las *verdaderas* reglas (no-algorítmicas) que gobiernan el modo en que funciona el mundo!

El principio antrópico

¿Qué importancia tiene la consciencia para el Universo como un todo? ¿Podría existir un universo sin habitantes conscientes? ¿Están las leyes de la física especialmente diseñadas para permitir la existencia de vida consciente? ¿Hay algo especial en nuestra localización particular en el Universo, ya sea en el espacio o en el tiempo? Éste es el tipo de preguntas que pretende responder lo que se ha llegado a conocer como el principio antrópico.

Éste principio tiene muchas formas. (Ver Barrow y Tipler 1986.) La más claramente aceptable de estas formas se refiere simplemente a la localización espacio-temporal de la vida consciente (o «inteligente») en el Universo. Éste es el principio antrópico *débil*. El argumento puede ser utilizado para explicar por qué se dan las condiciones justas para la existencia de vida (inteligente) en la Tierra en la época presente: si no fueran las correctas entonces nosotros mismos no estaríamos aquí, sino en alguna otra parte, en alguna otra época apropiada. Este principio fue utilizado de modo muy eficaz por Brandon Carter y Robert Dicke para resolver un problema que había intrigado a los físicos durante muchos años. El problema se refería a varias relaciones numéricas sorprendentes que se ha observado que se dan entre las constantes físicas (la constante gravitatoria, la masa del protón, la edad del Universo, etc.). Un aspecto enigmático de esto era que algunas de estas relaciones son válidas solamente en la época actual de la historia de la Tierra, de modo que casualmente parece que estamos viviendo en un tiempo muy especial (¡con un margen de unos pocos millones de años más o menos!) Esto fue finalmente explicado, por Carter y Dicke, por el hecho de que este período de tiempo coincide con la vida media de las llamadas estrellas de la secuencia principal, como es el Sol. En cualquier otra época, sigue diciendo el argumento, no existiría vida inteligente para medir las constantes físicas en cuestión —de modo que la coincidencia *tenía que* darse simplemente por el hecho, de que ¡sólo existiría vida inteligente en el momento particular en que *se diera* la coincidencia!

El principio antrópico *fuerte* va más lejos. En este caso, no sólo estamos interesados en nuestra localización espacio-temporal en el Universo, sino en una infinitud de universos *posibles*. Ahora podemos sugerir respuestas a preguntas como la de por qué las constantes de la física, o las leyes de la física en general, están especialmente diseñadas para que pueda existir vida inteligente. El razonamiento sería que si las constantes o las leyes fueran diferentes, entonces ¡no esta-

ríamos en este universo particular, sino que estaríamos en algún otro! En mi opinión, el principio antrópico fuerte tiene un carácter algo dudoso y tiende a ser invocado por los teóricos cuando no tienen ninguna teoría bastante buena para explicar los hechos observados (por ejemplo, en las teorías de la física de partículas, en las que las masas de las partículas no están explicadas y se argumenta que, si hubieran tenido valores distintos de los observados, entonces la vida sería imposible presumiblemente, etc.). El principio antrópico débil, por el contrario, me parece incuestionable siempre que seamos muy cuidadosos en la forma de utilizarlo.

Mediante la utilización del principio antrópico —ya sea en la forma fuerte o en la débil— podríamos tratar de probar que la consciencia era *inevitable* en virtud del hecho de que tenía que haber seres sensibles, es decir «nosotros», para observar el mundo, ¡así que *no* necesitamos suponer, como he hecho, que la sensibilidad tenga alguna ventaja selectiva! En mi opinión, este razonamiento es técnicamente correcto y el argumento del principio antrópico débil *podría* proporcionar (al menos) una razón para que la consciencia exista sin que tenga que ser favorecida por la selección natural. Pese a ello, yo no puedo creer que el argumento antrópico sea la razón *auténtica* (o la única razón) para la evolución de la consciencia. Hay evidencia suficiente procedente de otras direcciones para convencerme de que la consciencia *tiene* una poderosa ventaja selectiva, y no creo que se necesite el principio antrópico.

Teselaciones y cuasicristales

Dejaré ahora las especulaciones demasiado generales de las últimas secciones y consideraré en su lugar una cuestión que, aunque todavía algo especulativa, es mucho más científica y «tangible». Esta cuestión parecerá al principio que es una digresión irrelevante; pero su importancia para nosotros se pondrá de manifiesto en la próxima sección.

Recordemos las pautas de teselación mostradas en la Fig. 4.12, p. 184. Estas pautas son bastante notables por cuanto «casi» violan un teorema matemático estándar relativo a las redes cristalinas. El teorema establece que la únicas simetrías rotacionales que están permitidas para una estructura cristalina son la doble, triple, cuádruple y séxtuple. Por estructura cristalina entiendo un sistema discreto de puntos que tiene una *simetría de translación*; es decir, existe una ma-

nera de hacer deslizar la estructura sobre sí misma sin rotarla de modo que finalmente la estructura coincida consigo misma (esto es, es invariante bajo ese movimiento particular) y por lo tanto tiene un *paralelogramo período* (cf. Fig. 4.8). Ejemplos de pautas de teselación con estas simetría de rotación se dan en la Fig. 10.2. Ahora bien, las pautas de la Fig. 4.12, como las mostradas en la Fig. 10.3 (que es esencialmente la teselación producida al ajustar las teselas de la Fig. 4.11) *casi* tienen, por el contrario, simetría de translación y *casi* tienen simetría quíntuple —en donde «casi» significa que podemos encontrar movimientos de la estructura (translacional y rotacional respectivamente) tales que la estructura coincide consigo misma hasta cualquier grado de acuerdo preasignado que se desee, aunque sin llegar al 100 por ciento. No tenemos que preocuparnos ahora del significado exacto de esto. El único punto que será importante para nosotros es que si tuviéramos una substancia en la que los átomos estuvieran dispuestos en los vértices de esta estructura, entonces parecería ser cristalina ¡pese a que exhibiría una simetría quíntuple prohibida!

En diciembre de 1984, el físico israelí Dany Schechtman, que había estado trabajando con otros colegas en el National Bureau of

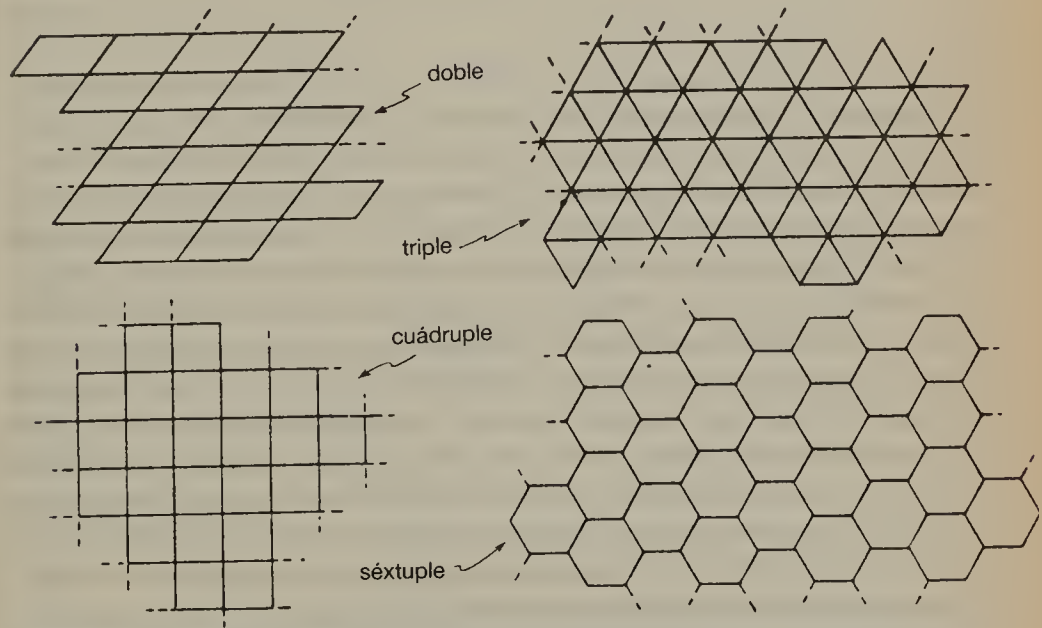


Fig. 10.2. Teselaciones periódicas con diversas simetrías (en las que el centro de simetría, en cada caso, se considera que es el centro de una tesela)

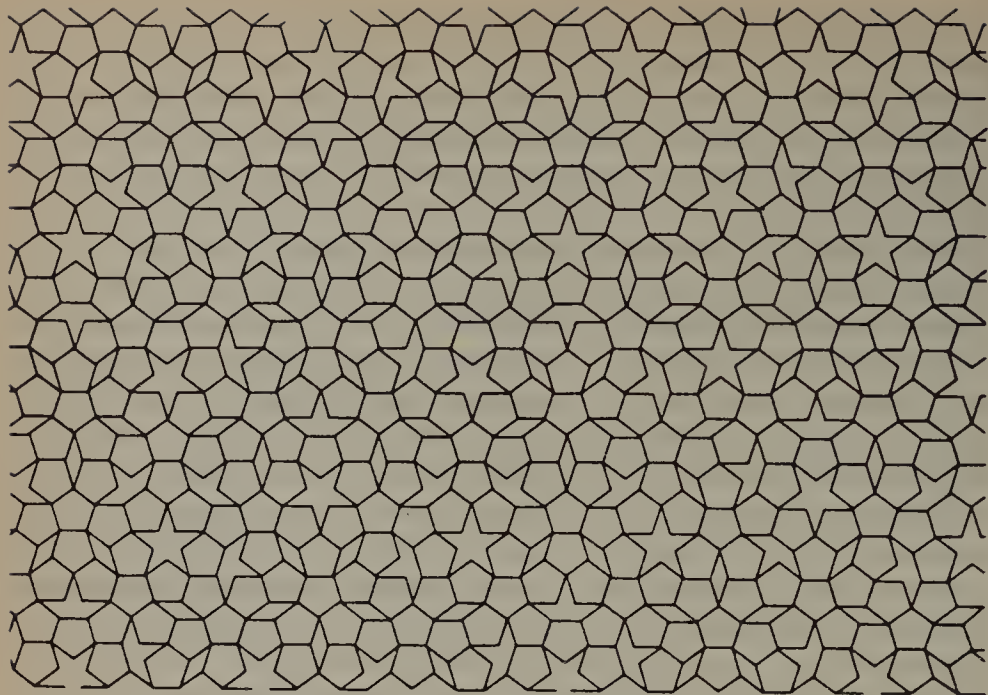


Fig. 10.3. Una teselación cuasi-periódica (básicamente la que se produce ajustando las teselas de la Fig. 4.11) con una cuasi-simetría quíntuple cristalográficamente «imposible»

Standards en Washington DC, EE.UU., anunció el descubrimiento de una fase de una aleación de aluminio-manganeso que parecía ser realmente una sustancia de tipo cristalino —ahora denominada *cua-sicristal*— con simetría quíntuple. En realidad, esta sustancia cuasi-cristalina también mostraba una simetría en *tres* dimensiones, y no sólo en el plano, dando globalmente una simetría *icosaédrica* prohibida (Shechtman *et al.*, 1984). («Icosaedros» tridimensionales análogos a mis teselaciones planas quíntuples han sido encontrados por Robert Ammann en 1975, ver Gardner 1989.) Las aleaciones de Shechtman formaban sólo minúsculos cristales microscópicos, de unos 10^{-3} milímetros de largo, pero posteriormente se descubrieron otras sustancias cuasicristalinas y, en particular, una aleación de aluminio-litio-cobre, para las que las unidades con simetría icosaédrica pueden crecer hasta un tamaño de aproximadamente un milímetro y son perfectamente visibles a simple vista (ver Fig. 10.4).

Ahora bien, una característica notable de las estructuras teselantes cuasicristalinas que he estado describiendo es que su ensamblaje es

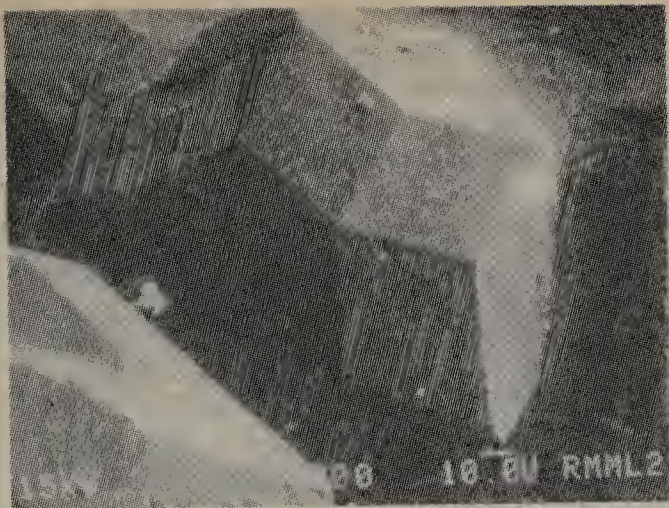


Fig. 10.4. Un cuasicristal (una aleación Al-Li-Cu) con una simetría cristalina aparentemente imposible. (Tomado de Gayle 1987.)

necesariamente *no-local*. Es decir, al ensamblar las estructuras es necesario, de cuando en cuando, examinar el estado de la estructura a muchos y muchos «átomos» de distancia del punto de ensamblaje si queremos estar seguros de no cometer graves errores cuando juntemos las piezas. (Esto es quizá parecido al aparente «andar a tientas inteligente» a que me referí en relación con la selección natural.) Este tipo de característica es un ingrediente de una controversia considerable que rodea actualmente a la cuestión de la estructura y crecimiento de los cuasicristales, y no sería prudente intentar sacar conclusiones definitivas hasta que algunos de los puntos más importantes queden resueltos. De todas formas, podemos especular; y yo aventuraré mi propia opinión. En primer lugar, creo que algunas de estas sustancias cuasicristalinas realmente están fuertemente organizadas, y sus disposiciones atómicas tienen una estructura muy próxima a las estructuras teselantes que he estado considerando. En segundo lugar, soy de la opinión (más tentativa) de que esto implica que su ensamblaje no puede conseguirse razonablemente mediante la adición local de átomos de uno en uno, de acuerdo con la imáaen clásica del crecimiento cristalino, sino que en su lugar debe haber en su ensamblaje un ingrediente mecano-cuántico esencialmente *no-local* ⁷.

El modo en que yo imagino que tiene lugar este crecimiento es que, en lugar de tener átomos que llegan individualmente y se acoplan ellos mismos a una línea de crecimiento en continuo movimiento (crecimiento cristalino clásico), debemos considerar la evolución de una superposición lineal cuántica de muchas diferentes configuraciones al-

ternativas de átomos que se acoplan (mediante el procedimiento cuántico U). ¡De hecho, esto es lo que la mecánica cuántica nos dice que *debe* estar ocurriendo (casi siempre)! No se trata de que sólo suceda una cosa; muchas disposiciones atómicas alternativas deben coexistir en una superposición lineal compleja. Unas pocas de estas alternativas superpuestas crecerán hasta constituir conglomerados más grandes y, en un cierto punto, la diferencia entre los campos gravitatorios de algunas de las alternativas alcanzarán el nivel de un-gravitón (o cualquiera que sea apropiado; ver Capítulo 8, p. 457). En esta etapa, una de las configuraciones alternativas —o, más exactamente, todavía una superposición, aunque una superposición algo reducida— se singularizará como la configuración «real» (procedimiento cuántico R). Este ensamblaje superpuesto, junto con reducciones a configuraciones más definidas, continuará a escala cada vez mayor hasta que se forma un cuasicristal de tamaño razonable.

Normalmente, cuando la Naturaleza busca una configuración cristalina está buscando una configuración de *mínima energía* (suponiendo que la temperatura ambiente sea cero). Estoy concibiendo algo similar en el crecimiento de cuasicristales, con la diferencia de que este estado de mínima energía es mucho más difícil de encontrar, y la «mejor» configuración de átomos no puede descubrirse simplemente añadiendo átomos de uno en uno con la esperanza de poder considerar cada átomo por separado resolviendo simplemente su *propio* problema de minimización. En lugar de ello, tenemos que resolver un problema *global*. Debe haber un esfuerzo cooperativo de un gran número de átomos a la vez. Semejante cooperación, estoy suponiendo, debe ser conseguida mecano-cuánticamente; y el modo de hacerlo es «ensayando» simultáneamente diferentes configuraciones de átomos combinadas en una superposición lineal (un poco quizá como el computador cuántico considerado al final del Capítulo 9). La selección de una solución apropiada (aunque probablemente no la mejor) al problema de la minimización debe conseguirse cuando se alcanza el criterio de un gravitón (o una alternativa apropiada) —lo que presumiblemente sólo ocurrirá cuando se den las condiciones físicas precisas.

Posible relevancia para la plasticidad cerebral

Permítaseme llevar estas especulaciones un poco más lejos y preguntar si pudieran tener alguna relevancia para la cuestión del fun-

cionamiento del cerebro. La posibilidad más probable, por lo que puedo ver, está en el fenómeno de la plasticidad cerebral. Recordemos que el cerebro no es realmente muy parecido a un ordenador sino que se parece más a un ordenador que está cambiando continuamente. Estos cambios pueden suceder aparentemente por la activación o desactivación de sinapsis a través del crecimiento o contracción de espina dendríticas (ver Capítulo 9, p. 491; Fig. 9.15). Me atrevo a especular que este crecimiento o contracción podría estar gobernado por algo semejante al proceso implicado en el crecimiento de los cuasicristales. En tal caso, no sólo se ensaya una sino un inmenso número de posibles configuraciones alternativas, todas superpuestas en una superposición lineal compleja. Mientras los efectos de estas alternativas se mantengan por debajo del nivel de un-gravitón (o cualquiera que sea) entonces ellas coexistirán (y casi invariablemente *deben* coexistir según las leyes de la mecánica cuántica U). Si se mantienen por debajo de este nivel, pueden empezar a ejecutarse cálculos superpuestos simultáneos, de un modo que está en concordancia con los principios de un ordenador cuántico. Sin embargo, parece improbable que estas superposiciones puedan mantenerse durante mucho tiempo, ya que las señales nerviosas producen campos eléctricos que perturbarían significativamente el material circundante (aunque su vaina de mielina podría ayudar a aislarlas). Especulemos que tales superposiciones de cálculos pueden mantenerse realmente durante al menos el intervalo de tiempo suficiente para que algo de importancia *sea* efectivamente calculado antes de que se alcance el nivel de un-gravitón (o cualquiera que sea). El resultado acertado de tal cálculo sería el «objetivo» que sustituye al «objetivo» de la simple minimización de la energía en el crecimiento del cuasicristal. ¡Por lo tanto la consecución de dicho objetivo es similar al crecimiento con éxito del cuasicristal!

Hay obviamente mucha vaguedad e indecisión en estas especulaciones, pero creo que representan una analogía realmente plausible. El crecimiento de un cristal o un cuasicristal está fuertemente influenciado por las concentraciones de los átomos e iones apropiados en su vecindad. Análogamente, se podría concebir que el crecimiento o contracción de familias de espinas dendríticas podría perfectamente estar influenciado por las concentraciones de las diversas sustancias neurotransmisoras que pudieran estar alrededor (tanto como pudiera estar afectado por las emociones). Cualesquiera que sean las configuraciones atómicas que finalmente se resuelven (o «reducen») en la *realidad* del cuasicristal, ellas suponen la solución de un problema de minimización de la energía. De modo análogo, así estoy especulando,

el pensamiento real que sale a la superficie en el cerebro es de nuevo la solución de cierto problema, aunque ahora no es sólo un problema de minimización de energía. Implicaría en general un objetivo de naturaleza mucho más complicada, que incluye deseos e intenciones que están en sí mismos relacionados con los aspectos y capacidades computacionales del cerebro. Estoy especulando que la acción del pensamiento consciente está muy ligada a la resolución de alternativas que previamente estaban en la superposición lineal. Todo esto está relacionado con la física desconocida que gobierna la línea divisoria entre **U** y **R** y que, estoy afirmando, depende de una aún no descubierta teoría de la gravitación cuántica-GQC!

¿Podría semejante acción física ser de naturaleza no-algorítmica? Recordemos que el problema general de la teselación, como se describió en el Capítulo 4, no tiene solución algorítmica. Se podría concebir que los problemas del ensamblaje de átomos podrían compartir esta propiedad no-algorítmica. Si estos problemas pueden «resolverse» en principio por el tipo de medios que he estado insinuando, entonces existe alguna posibilidad para un ingreciente no algorítmico en el tipo de acción cerebral que tengo en mente. Para que esto sea así, sin embargo, necesitamos algo no-algorítmico en la GQC. Evidentemente hay mucha especulación aquí. Pero *algo* de carácter no-algorítmico me parece definitivamente necesario a la vista de los argumentos antes expuestos.

¿Con qué rapidez tienen lugar estos cambios en la conexión cerebral? La cuestión parecer ser algo controvertida entre los neurofisiólogos pero, puesto que la memoria permanente puede asentarse en unas pocas fracciones de segundo, es plausible que tales cambios en las conexiones puedan efectuarse en un tiempo de ese orden. Para que mis propias ideas tengan una oportunidad sería necesaria una velocidad de este tipo.

Los retardos temporales de la consciencia

A continuación deseo describir dos experimentos (descritos en Harth 1982) que se han realizado en sujetos humanos y que parecen tener implicaciones notables para nuestras consideraciones sobre este punto. Estos tienen que ver con el tiempo que necesita la consciencia para actuar y ser activada. El primero de estos concierne al papel activo de la consciencia, y el segundo a su papel pasivo. Consideradas juntas, estas implicaciones son aún más sorprendentes.

El primero fue realizado por H. H. Kornhuber y colaboradores en Alemania en 1976. (Deecke, Grötzinger y Kornhuber 1976.) Cierta número de sujetos humanos se prestaron voluntarios para que se registrasen las señales eléctricas en un punto de sus cabezas (electroencefalogramas, esto es, EEG), y se les pedía que flexionaran varias veces, y repentinamente, el dedo índice de sus manos derechas *totalmente a su propia elección*. La idea es que los registros de los EEG indicarían algo de la actividad mental que está teniendo lugar dentro del cráneo y que está involucrada en la decisión real consciente de flexionar el índice. Para obtener una señal significativa a partir de las trazas del EEG es necesario promediar las trazas de varias series diferentes, y la señal resultante no es muy definida. Sin embargo, se encuentra algo curioso: hay un aumento gradual del potencial eléctrico registrado durante un *segundo entero*, o quizá incluso hasta un segundo y medio, *antes* de que el dedo sea flexionado. Esto parece indicar que el proceso de decisión consciente ¡necesita un segundo para actuar! Esto puede ser contrastado con el tiempo mucho más corto que lleva responder a una señal externa si el modo de respuesta ha sido establecido por adelantado. Por ejemplo, la flexión del dedo podría ser la respuesta al destello de una señal luminosa en lugar de ser «libremente voluntaria». En tal caso es normal un tiempo de reacción de aproximadamente un quinto de segundo, que es unas cinco veces más rápido que la acción «voluntaria» que se ponía a prueba en los datos de Kornhuber (ver Fig. 10.5).

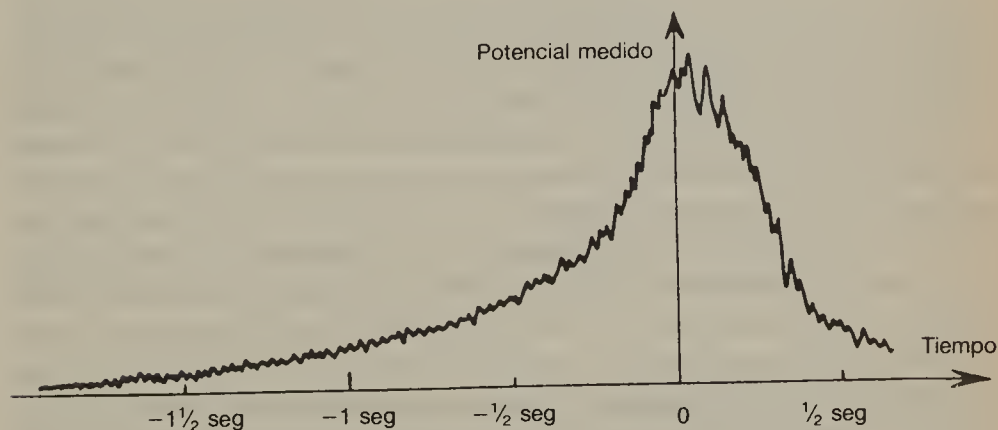


Fig. 10. Experimento de Kornhuber. La decisión de flexionar el dedo parece hacerse en el tiempo 0, pese a que la señal precursora (promediada sobre muchos ensayos) sugiere un «preconocimiento» de la intención de flexionar.

En el segundo experimento, Benjamit Libet, de la Universidad de California, en colaboración con Bertram Feinstein del Instituto Neurológico del Monte Sión en San Francisco (Libet *et al.*, 1979), ponían a prueba sujetos que tenían que sufrir intervención quirúrgica en el cerebro por alguna razón que no tuviera relación con el experimento y que consintieron en tener electrodos colocados en puntos del córtex somatosensorial del cerebro. El resultado del experimento de Libet fue que, cuando se aplicaba un estímulo en la piel de estos pacientes, transcurría aproximadamente medio segundo antes de que fueran conscientes de dicho estímulo, pese al hecho de que el propio cerebro había recibido la señal del estímulo en sólo una centésima de segundo y podía lograrse una respuesta pre-programada «refleja» del cerebro a dicho estímulo (cf. arriba) en aproximadamente una décima de segundo (Fig. 10.6). Además, a pesar del retardo de medio segundo antes de que el estímulo alcanzase la consciencia, existía la impresión subjetiva de los propios pacientes de que ¡no había habido ningún retraso en su toma de conciencia del estímulo! (Algunos de los experimentos de Libet implicaban estimulación del tálamo, cf. p. 471, con resultados similares a los del córtex somatosensorial.)

Recordemos que el córtex somatosensorial es la región del cerebro por donde entran las señales nerviosas. Por lo tanto, la estimulación eléctrica de un punto del córtex somatosensorial, correspondiente a un punto particular de la piel, sería para el sujeto igual que si algo hubiera tocado realmente la piel en el punto correspondiente. Sin embargo, resulta que si la estimulación eléctrica es demasiado breve —durante menos de aproximadamente medio segundo— entonces el sujeto no llega a ser consciente de ninguna sensación. Esto debe contrastarse con una estimulación directa del punto de la misma piel, puesto que un contacto momentáneo en la piel sí puede ser sentido.

Supongamos ahora que se toca primero la piel y luego se estimula eléctricamente el punto del cortex somatosensorial. ¿Qué siente el paciente? Si la estimulación eléctrica se inicia aproximadamente un cuarto de segundo después de tocar la piel, entonces ¡el contacto en la piel no se siente en absoluto! Éste es un efecto conocido como *enmascaramiento retroactivo*. De algún modo la estimulación del córtex sirve para impedir que la sensación normal de contacto en la piel sea sentida conscientemente. La percepción consciente puede ser inhibida («enmascarada») por un suceso posterior, siempre que dicho suceso ocurra antes de medio segundo. Esto nos dice que la percepción consciente de tal sensación ocurre algo así como ¡medio segundo después del suceso real que produce esa sensación!

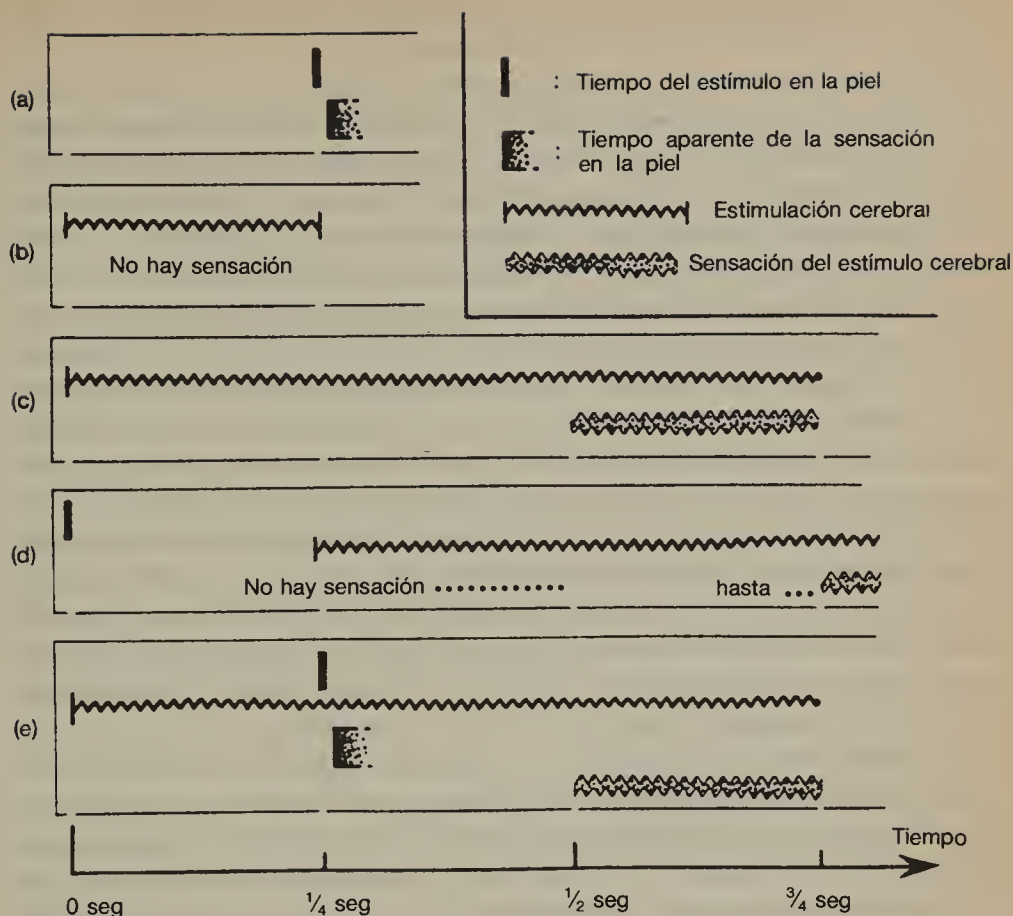


Fig. 10.6. Experimento de Libet. (a) El estímulo en la piel «parece» percibirse aproximadamente en el momento real del estímulo. (b) Un estímulo cortical de menos de medio segundo no es percibido. (c) Un estímulo cortical de más de medio segundo se percibe a partir de medio segundo en adelante. (d) Tal estímulo cortical puede «enmascarar retroactivamente» un estímulo anterior en la piel, lo que indica que la conciencia del estímulo en la piel *no ha tenido lugar aún* en el momento del estímulo cortical. (e) Si un estímulo en la piel se aplica inmediatamente *después* del estímulo cortical, la percepción de la piel es «remitida hacia atrás» pero la percepción cortical no lo es.

Sin embargo, no parece que seamos «conscientes» de un retardo temporal tan largo en nuestras percepciones. Una manera de hacer comprensible este curioso descubrimiento podría consistir en imaginar que el «tiempo» de todas nuestras «percepciones» está retrasado efectivamente alrededor de medio segundo respecto al «tiempo real»

—como si el reloj interno de cada uno simplemente estuviese «equivocado» en medio segundo aproximadamente.

Quizá algo de esta naturaleza queda corroborado en una segunda parte del experimento de Libet, en donde él iniciaba *primero* una estimulación eléctrica del córtex, continuando esta estimulación durante un intervalo mucho mayor de medio segundo y tocando también la piel mientras seguía esta estimulación, aunque menos de medio segundo después de su inicio. Tanto la estimulación cortical como el contacto en la piel se percibían por separado, y resultaba claro para el sujeto cuál era cada uno. Sin embargo, cuando se le preguntaba qué estímulo ocurrió *primero*, el sujeto respondía que fue el contacto en la piel, ¡pese al hecho de que la estimulación cortical se inició mucho antes! Por lo tanto, el sujeto parece remitir la percepción del contacto en la piel *hacia atrás en el tiempo* en aproximadamente medio segundo (ver Fig. 10.6). Sin embargo, parece que esto no es simplemente un error «global» en el tiempo internamente percibido sino una readaptación de la percepción temporal de los sucesos, pues la estimulación cortical, suponiendo que se perciba realmente no más de medio segundo después de su inicio, *no* parece estar remitida hacia atrás de esta forma.

A partir del primero de los experimentos anteriores parece deducirse que la acción consciente necesita algo así como un segundo o un segundo y medio antes de poder llevarse a cabo, mientras que, según el segundo experimento, la consciencia de un suceso externo no parece ocurrir hasta medio segundo después de que haya tenido lugar el suceso. Imaginemos lo que sucede cuando respondemos a algún acontecer externo imprevisto. Supongamos que la respuesta es algo que requiere un momento de contemplación consciente. Parecería, sobre la base de los experimentos de Libet, que debiera transcurrir medio segundo antes de que la consciencia sea llamada a juego; y luego, como parecen implicar los datos de Kornhuber, se necesita bastante más de un segundo antes de que pueda tener efecto una respuesta «voluntaria». El proceso total, desde el input sensorial hasta el output motor, parecería necesitar un tiempo ¡del orden de dos segundos! La implicación aparente de estos dos experimentos considerados juntos es que la consciencia no puede siquiera entrar en juego *en absoluto* en respuesta a un suceso externo, ¡si dicha respuesta tiene que tener lugar en menos de un par de segundos aproximadamente!

El extraño papel del tiempo en la percepción consciente

¿Podemos tomar estos experimentos al pie de la letra? Si es así, parece que estamos conducidos a la conclusión de que actuamos completamente como «autómatas» cuando llevamos a cabo cualquier acción que necesite menos de un segundo o dos en modificar una respuesta. Sin duda la consciencia es de acción lenta comparada con otros mecanismos del sistema nervioso. Yo mismo me he encontrado en ocasiones observando inútilmente cómo mi mano cierra la puerta del coche un instante después de haber reparado en alguna cosa que hubiera querido recoger del interior del coche, y mi orden voluntaria para detener el movimiento de mi brazo actúa de forma preocupantemente lenta —demasiado lenta para detener el cierre de la puerta. Pero ¿se requiere un segundo o dos para esto? Me parece poco probable que esté involucrada una escala de tiempo tan grande. Por supuesto, mi percepción *consciente* del objeto dentro del coche, junto con mi supuesta «libre voluntariedad» de la orden para detener mi mano, *podrían* haber ocurrido perfectamente después de ambos sucesos. Quizá la consciencia es, después de todo, un espectador que no experimenta más que una «acción repetida» del drama entero. De modo similar, y frente a esto, sobre la base de los descubrimientos anteriores no habría tiempo para que la consciencia juegue el más mínimo papel cuando, por ejemplo, ejecutamos un golpe de tenis —¡y mucho menos de ping-pong! Sin duda los expertos en estas tareas tendrán todos los elementos esenciales de sus respuestas soberbiamente preprogramados en el control cerebelar. Pese a ello, encuentro un poco difícil dar crédito a que la consciencia *no* esté jugando el más mínimo papel en la decisión de qué golpe jugar en cada ocasión. Sin duda hay mucho de anticipación de lo que el adversario podría hacer, y muchas respuestas preprogramadas deberían estar disponibles para cada posible acción del adversario, pero me parece que esto es ineficaz y encontraría difícil aceptar una *total* ausencia de implicación consciente en ese instante. Comentarios semejantes serán aún más pertinentes en relación con la conversación ordinaria. También aquí, aunque pudiéramos anticipar en parte lo que el otro vaya a decir, debe haber normalmente algo inesperado en los comentarios del interlocutor ¡o la conversación sería totalmente innecesaria! ¡Ciertamente no se necesita tanto como un par de segundos para responder a alguien en el modo normal de conversación!

Hay quizá alguna razón para dudar que los experimentos de Kornhuber demuestren que la consciencia «necesite» realmente un segundo

y medio para actuar. Aunque es cierto que el *promedio* de todas las trazas del EEG para la intención de flexionar el dedo daba una señal tan prematura, pudiera ser que sólo en *algunos* casos haya una intención tan anticipada de flexionar el dedo —donde a menudo esta intención consciente *no* se materializaría realmente— mientras que en muchos otros casos la acción consciente ocurra mucho más próxima a la flexión del dedo.

Aceptemos, por el momento, que ambas conclusiones experimentales *son* realmente válidas. Me gustaría hacer una alarmante sugerencia en relación con esto. ¡Sugiero que es probable que estemos equivocados al aplicar las reglas físicas usuales para el *tiempo* cuando consideramos la consciencia! Existe, efectivamente, algo muy singular en el modo en que entra el tiempo en nuestras percepciones conscientes en cualquier caso, y pienso que es posible que sea necesaria una concepción muy diferente cuando tratamos de colocar las percepciones conscientes en un marco convencional de ordenación temporal. La consciencia es, después de todo, el único fenómeno que conocemos según el cual el tiempo ¡necesita «fluir»! El modo de considerar el tiempo en la física moderna no es esencialmente diferente de la manera en que se considera el *espacio* *, y el «tiempo» de las descripciones físicas no «fluye» en absoluto; ¡sólo tenemos un «espacio-tiempo» fijo de apariencia estática en el que están dispuestos los sucesos de nuestro universo! No obstante, según nuestras percepciones, el tiempo *fluye* (ver Capítulo 7). Mi conjetura es que aquí también existe algo ilusorio, y el tiempo de nuestras percepciones no fluye «realmente» en la forma de avance lineal en que lo percibimos fluir (¡cualquier cosa que sea lo que esto pueda significar!) El ordenamiento temporal que nos «parece» percibir es, estoy afirmando, algo que imponemos a nuestras percepciones para poder darles sentido en relación con la progresión temporal uniforme hacia adelante de una realidad física externa.

Algunas personas podrían detectar muchas «incorrecciones filosóficas» en los comentarios anteriores —y sin duda tendrán razón en

* La simetría entre tiempo y espacio sería aún más sorprendente para un espacio-tiempo *bi*-dimensional. Las ecuaciones de la física en un espacio-tiempo bidimensional serían esencialmente simétricas con respecto al intercambio del espacio con el tiempo —pese a que nadie consideraría que el espacio «fluye» en la física en dos dimensiones. Es difícil de creer que lo que hace que el tiempo «fluya realmente», en nuestras experiencias del mundo físico que conocemos, sea simplemente la asimetría entre el número de dimensiones espaciales (3) y dimensiones temporales (1) que resulta tener nuestro espacio-tiempo.

estas acusaciones. ¿Cómo podemos estar «errados» sobre lo que realmente percibimos? Ciertamente, nuestras percepciones reales son, *por definición*, precisamente las cosas de las que somos directamente conscientes; por lo tanto, no podemos estar «errados» sobre ello. De todas formas, pienso que es probable que *estemos* «errados» sobre nuestras percepciones del progreso temporal (a pesar de mis inexactitudes en el uso del lenguaje ordinario para describir esto) y que hay evidencia para apoyar tal creencia (ver Chrchland, 1984).

Un ejemplo extremo (p. 524) es la capacidad de Mozart para «captar de golpe» una composición musical entera «por larga que pueda ser». Debemos suponer, por la descripción de Mozart, que este «golpe» contenía la esencia de la composición entera, pese a que el intervalo de tiempo real, en términos físicos ordinarios, de este acto consciente de percepción no fuera en modo alguno comparable con el tiempo que se necesitaría para ejecutar la composición. Podríamos imaginar que la percepción de Mozart habría tomado una forma completamente diferente, quizá distribuida espacialmente como una escena visual o toda una partitura musical desplegada. Pero incluso una partitura musical necesitaría un tiempo considerable para ser leída —y dudo mucho que la percepción de Mozart de sus composiciones pudiera haber tomado inicialmente esta forma (¡o seguro que lo habría dicho!). La escena visual parecería más próxima a sus descripciones, pero (como con las más comunes de las imágenes matemáticas que me son personalmente más familiares) dudaría mucho que existiera algo semejante a una traducción directa de la música en términos visuales. Me parece mucho más probable que la mejor interpretación del «golpe» de Mozart debe tomarse *musicalmente*, con las connotaciones característicamente temporales que tendría la audición (o ejecución) de una pieza musical. La música consiste en sonidos cuya ejecución lleva un tiempo definido, el *tiempo* que en la descripción real de Mozart permite «... que mi imaginación me deje oírla».

Oigamos la cuádruple fuga en la parte final del *Arte de la fuga* de J. S. Bach. Nadie a quien le guste la música de Bach puede evitar el quedar conmovido cuando la música se detiene después de diez minutos de ejecución, precisamente tras la entrada del tercer tema. La composición como un todo parece que de algún modo sigue «ahí», pero ahora se nos ha desvanecido en un instante. Bach murió antes de poder completar la obra y su partitura se detiene simplemente en este punto, sin ninguna indicación escrita sobre cómo pensaba continuarla. Pero empieza con tal seguridad y completo dominio que resulta imposible imaginar que no tuviese la esencia de la composición

entera en su cabeza en ese mismo instante. ¿Tendría necesidad de tocarla en su totalidad para sí mismo en su mente, al ritmo de ejecución normal, ensayando una y otra vez, y otra vez más, a medida que se le ocurrían las diferentes mejoras? No puedo imaginar que lo hiciera de esta forma. Al igual que Mozart, él debió haber sido capaz de concebir de algún modo la obra en su totalidad, con la intrincada complejidad y sentido artístico que exige la escritura fugada, evocada toda a la vez. Pero la cualidad temporal de la música es uno de sus ingredientes esenciales. ¿Cómo es que la música puede seguir siendo música si no se está ejecutando en «tiempo real»?

La concepción de una novela o una historia podría presentar un problema comparable (aunque aparentemente menos enigmático). Para la comprensión de la vida entera de un individuo necesitaríamos contemplar diversos sucesos cuya adecuada apreciación parecería requerir su actualización mental en «tiempo real». Pero no parece que esto sea necesario. Incluso las impresiones de los recuerdos de nuestras propias experiencias extendidas en el tiempo parecen estar de algún modo tan «comprimidas» que virtualmente podemos «revivirlas» ¡en un instante de rememoración!

Hay quizá una estrecha similitud entre la composición musical y el pensamiento matemático. La gente podría suponer que una demostración matemática se concibe como una progresión lógica, en donde cada paso se sigue de los que le precedían. Pero es apenas probable que la concepción de un nuevo razonamiento proceda de esta forma. Existe una globalidad y un contenido conceptual aparentemente vago que son necesarios en la construcción de un argumento matemático; y esto puede guardar poca relación con el tiempo que parecería necesitarse para apreciar completamente una demostración presentada serialmente.

Supongamos, entonces, que aceptamos que el ritmo y la progresión temporal de la consciencia no concuerda con el de la realidad física externa. ¿No estamos en peligro de llegar a una paradoja? Supongamos que haya incluso algo vagamente teleológico en los efectos de la consciencia, de modo que una impresión futura pueda afectar a una acción pasada. ¿Nos llevaría *esto* con seguridad a una contradicción, como las implicaciones paradójicas de las señales más-rápidas-que-la-luz que consideramos —y precisamente descartamos— en nuestra discusión hacia el final del Capítulo 5? (cf. p. 273) Quiero sugerir que *no* hay necesariamente una paradoja —por la propia naturaleza de lo que estoy sosteniendo que realmente consigue la consciencia. Recuérdesse mi propuesta de que la consciencia es, en esencia, la

«visión» de una verdad necesaria; y que puede representar un tipo de contacto real con el mundo platónico de los conceptos matemáticos ideales. Recuértese que el mismo mundo platónico es intemporal. La percepción de la verdad platónica no conlleva información —en el sentido técnico de «información» que puede transmitirse mediante un mensaje— ¡y no habría contradicción real envuelta si tal percepción consciente se propagara incluso hacia atrás en el tiempo!

Pero aun si aceptamos que la propia consciencia tiene esa curiosa relación con el tiempo —y que representa, en cierto sentido, un contacto entre el mundo físico externo y algo intemporal— ¿cómo puede esto ajustarse a una acción físicamente determinada y ordenada temporalmente del cerebro material? Una vez más, parece que nos quedamos con un simple papel de «espectador» para la consciencia si no podemos reflejar la progresión normal de las leyes físicas. Pero yo *estoy* argumentando en favor de algún tipo de papel activo para la consciencia, y en realidad para un papel poderoso, con una fuerte ventaja selectiva. La respuesta a este dilema, creo yo, reside en la extraña forma en que debe actuar la GQC en su resolución del conflicto entre los dos procesos mecano-cuánticos **U** y **R** (cf. pp. 438, 456).

Recuérdense los problemas con el tiempo que encuentra el proceso **R** cuando tratamos de hacerlo compatible con la relatividad (especial) (Capítulos 6, 8, pp. 361, 461). El proceso no parece tener sentido en absoluto cuando se describe en términos de espacio-tiempo. Consideremos un estado cuántico de un par de partículas. Tal estado será normalmente un estado *correlacionado* (es decir, *no* de la simple forma $|\psi\rangle|\chi\rangle$, donde cada uno de los $|\psi\rangle$ y $|\chi\rangle$ describe sólo una de las partículas, sino una suma, como $|\psi\rangle|\chi\rangle + |\alpha\rangle|\beta\rangle + \dots + |\varrho\rangle|\sigma\rangle$). En tal caso, una observación en una de las partículas afectará a la otra de una forma no-local que no puede describirse en términos de espacio-tiempo ordinario de forma compatible con la relatividad especial (EPR; el efecto Einstein-Podolsky-Rosen). Tales efectos no-locales estarían implícitamente involucrados en mi sugerida analogía entre el «cuasicristal» y el crecimiento y contracción de las espinas dendríticas.

Aquí, estoy interpretando «observación» en el sentido de un amplificación de la acción de cada partícula observada hasta que se satisfaga algo parecido al criterio de un-gravitón. En términos más «convencionales», una «observación» es una cosa mucho más oscura y es difícil ver cómo podría empezarse a desarrollar una descripción teórico-cuántica de la acción cerebral ¡cuando perfectamente podríamos

tener que considerar el cerebro como «observándose a sí mismo» continuamente!

Mi propia idea es que la GQC, por el contrario, proporciona una teoría física *objetiva* de la reducción del vector de estado (**R**) que *no* tendría que depender de ninguna idea de consciencia. Todavía no tenemos una teoría semejante, ¡pero al menos su descubrimiento no será obstaculizado por los profundos problemas para decidir que «es» realmente la consciencia!

Imagino que una vez que la GQC haya sido finalmente descubierta, puede hacerse *entonces* posible elucidar en sus términos el fenómeno de la consciencia. En realidad, creo que las propiedades exigidas a la GQC, cuando llegue la teoría, estarán incluso más alejadas de una descripción espacio-temporal convencional de lo que lo están los enigmáticos fenómenos EPR para las dos partículas citadas antes. Si, como estoy sugiriendo, el fenómeno de la consciencia depende de esta deseada GQC, entonces ¡la propia consciencia se adaptará sólo de forma bastante incómoda a nuestras actuales descripciones espacio-temporales!

Conclusión: una visión de niño

En este libro he presentado muchos argumentos intentando mostrar lo insostenible del punto de vista —aparentemente prevaleciente en la corriente filosófica actual— de que nuestro pensamiento es básicamente lo mismo que la acción de un computador muy complicado. Se ha adoptado aquí la terminología de Searle «IA fuerte» para la suposición explícita de que la simple operación de un algoritmo puede provocar *conocimiento consciente*. Otros términos tales como «funcionalismo» son utilizados a veces de una manera algo menos definida.

¡Algunos lectores pueden haber considerado desde el principio que el «defensor de la IA fuerte» era quizá más que nada un enemigo ficticio! ¿No es obvio que la simple computación no puede provocar placer o dolor; que no puede percibir la poesía, o la belleza del cielo al atardecer, o la magia de los sonidos; que no puede tener esperanza o amar o desesperar; que no puede tener un objetivo genuino autónomo? Pero la ciencia parece habernos llevado a aceptar que todos somos simplemente pequeñas partes de un mundo gobernado en todo detalle (incluso si finalmente resultara ser quizá de manera probabilista) por leyes matemáticas muy precisas. Nuestros propios cerebros, que parecen controlar todas nuestras acciones, están gobernados tam-

bién por estas mismas leyes precisas. Ha emergido la imagen de que toda esta precisa actividad física es, en efecto, nada más que la activación de alguna enorme computación (quizá probabilista) —y por ello nuestros cerebros y nuestras mentes tienen que ser comprendidos solamente en términos de semejantes computaciones. Quizá cuando las computaciones se vuelvan extraordinariamente complicadas pueden empezar a adoptar las cualidades más poéticas o subjetivas que asociamos al término «mente». No obstante, es difícil evitar un sentimiento incómodo de que siempre se echará algo de menos en esta imagen.

En mis propios argumentos he tratado de apoyar esta idea de que debe haber algo esencial que está ausente de cualquier imagen puramente computacional. Pero mantengo también la esperanza de que es a través de la ciencia y las matemáticas como deben salir a la luz algunos avances profundos en la comprensión de la mente. Existe aquí un aparente dilema, pero he tratado de mostrar que existe una salida real. Computabilidad no es en modo alguno la misma cosa que ser matemáticamente exacto. Existe en el mundo matemático platónico exacto tanto misterio y belleza como pudiéramos desear, y la mayor parte de este misterio reside en conceptos que yacen fuera de la parte relativamente limitada de él donde residen los algoritmos y la computación.

La consciencia me parece un fenómeno de tal importancia que sencillamente no puedo creer que sea algo que sólo es «accidentalmente» producido por una computación complicada: es el fenómeno en el que se hace conocida la misma existencia del universo. Podemos argumentar que un universo gobernado por leyes que no permiten la consciencia no es universo en absoluto. Diría incluso que todas las descripciones matemáticas del universo que se han dado hasta ahora deben incumplir este criterio. ¡Es sólo el fenómeno de la consciencia el que puede conjurar un presunto universo «teórico» a la existencia real!

Algunos de los argumentos que he dado en estos capítulos pueden parecer tortuosos y complicados. Algunos son decididamente especulativos, aunque yo creo que no existe escapatoria real de unos u otros. Pero por encima de estos tecnicismos está el sentimiento de que es realmente «obvio» que la mente *consciente* no puede trabajar como un computador, incluso aunque mucho de lo que está realmente implicado en la actividad mental podría hacerlo.

Éste es el tipo de obviada que puede ver un niño —aunque ese niño pueda, en su vida posterior, verse llevado a creer que los pro-

blemas obvios son «no-problemas», de los que se argumentaba su no-existencia mediante cuidadosos razonamientos y definiciones inteligentemente escogidas. A veces los niños ven las cosas más claramente de como se han oscurecido en la vida avanzada. Cuando las preocupaciones de las actividades del «mundo real» han comenzado a cargarse sobre nuestros hombros solemos olvidar las maravillosas sensaciones de cuando éramos niños. Los niños no tienen miedo de hacer preguntas básicas que nosotros, como adultos, sentiríamos vergüenza de plantear. ¿Qué sucede a cada uno de nuestros flujos de consciencia después de morir?, ¿dónde estaba antes de que naciera cada uno?, ¿podríamos convertirnos en, o haber sido, algún otro?, ¿por qué percibimos en absoluto?, ¿por qué estamos aquí?, ¿por qué hay un universo en el que podamos estar? Estos son enigmas que tienden a llegar con el despertar de la consciencia en cualquiera de nosotros—y sin duda, con el despertar de la verdadera auto-consciencia, dentro de cualquier criatura u otra entidad en que primero llegue.

Recuerdo haber sido turbado, yo mismo, por muchos de estos enigmas cuando era niño. Quizá mi propia consciencia podría ser intercambiada repentinamente con la de algún otro. ¿Cómo sabría yo alguna vez si no pudiera haberme sucedido tal cosa antes —suponiendo que cada persona porta sólo los recuerdos importantes para esa persona concreta? ¿Cómo podría explicar semejante experiencia de «intercambio» a algún otro? ¿Realmente significa algo? Quizá estoy viviendo simplemente las mismas experiencias de diez minutos una y otra vez, cada vez con exactamente las mismas percepciones. Quizá sólo «existe» para mí el instante presente. Quizá el «mí» de mañana, o el de ayer, es realmente una persona muy diferente con una consciencia independiente. Quizá estoy viviendo realmente hacia atrás en el tiempo, con mi flujo de consciencia marchando hacia el pasado de modo que mi memoria me dice realmente lo que *va* a sucederme más que lo que me *ha* sucedido —de modo que la experiencia desagradable en la escuela es realmente algo que me aguarda y, desgraciadamente, encontraré en muy breve plazo. ¿«Significa» algo realmente la distinción entre *eso* y la progresión temporal experimentada realmente, de modo que uno es «falso» y la otra «correcta»? Para que las respuestas a semejantes preguntas sean resolubles en principio, sería necesaria una teoría de la consciencia. Pero ¿cómo podríamos siquiera *empezar* a explicar la substancia de tales problemas a una entidad que no sea ella misma consciente...?

Epílogo

«... gustaría? Oh, ... una pregunta muy interesante, muchacho... eh... yo también quisiera saber la respuesta», dijo el Diseñador Jefe. «Veamos lo que nuestro amigo tiene que decir a eso... esto es extraño... eh... Ultronic dice que él no ve... ni siquiera puede comprender lo que planteas!». Los murmullos de risas por la sala estallaron en grandes carcajadas.

Adam se sintió extremadamente molesto. Podían haber hecho cualquier otra cosa, pero no reírse.

Notas del Capítulo 1

1. Ver, por ejemplo, Gardner (1958), Gregory (1981), y las referencias que allí figuran.
2. Ver, por ejemplo, Resnikoff y Wells (1984), pp. 181-4. Para un informe clásico sobre los calculadores prodigio en general, ver Rouse Ball (1892); también Smith (1983).
3. Ver Gregory (1981), pp. 285-7, Grey Walter (1953).
4. Este ejemplo está sacado de Delbrück (1986)
5. Ver los artículos de O'Connell (1988) y Keene (1988). Para más información sobre los computadores que juegan al ajedrez, ver Levy (1984).
6. A lo largo de este libro he adoptado la terminología de Searle «IA fuerte» para este punto de vista extremo, sólo por ser concreto. El término «funcionalismo» se utiliza frecuentemente para lo que en esencia es el mismo punto de vista, pero quizá no siempre de una forma tan concreta. Algunos proponentes de este punto de vista son Minsky (1968), Fodor (1983) y Hofstadter (1979).
7. Ver Searle (1987), p. 211, para un ejemplo de tal afirmación.
8. En su crítica del artículo original de Searle, tal como está reimpressa en «The Mind's I», Douglas Hofstadter se queja de que ningún ser humano podría concebiblemente «internalizar» la descripción completa de la mente de otro ser humano debido a la complicación que supone. ¡Ciertamente no! Pero, tal como yo lo veo, no es esa la idea exacta. Estamos interesados simplemente en llevar a cabo parte de un algoritmo que pretende encarnar la ocurrencia de un simple estado mental. Este podría ser alguna «comprensión consciente» momentánea de la contestación a una pregunta de un test de Turing, o podría ser algo más simple. ¿Realmente requeriría tal «suceso» un algoritmo de enorme complicación?

9. Ver pp. 368, 372 del artículo de Searle (1980) en Hofstadter y Dennett (1981).
10. Algunos lectores, expertos en tales materias, podrían poner reparos a alguna diferencia de signo. ¡Pero incluso esa (defendible) diferencia desaparece si rotamos uno de los electrones 360° al hacer el intercambio! (Ver Capítulo 6, p. 279 para una explicación).
11. Ver la Introducción a Hofstadter y Dennett (1981).

Notas del Capítulo 2

1. Estoy siguiendo la terminología moderna usual que ahora incluye el cero entre los «números naturales».
2. Existen muchas otras maneras, bien conocidas para los matemáticos, de codificar pares, tríos, etc., de números como números individuales, aunque son menos convenientes para nuestro propósito. Por ejemplo, la fórmula $\frac{1}{2}((a + b)^2 + 3a + b)$ representa unívocamente los pares de números naturales como un simple número natural. ¡Inténtelo!
3. No me he molestado, en lo que precede, en introducir ninguna marca para *iniciar* la secuencia de números (o instrucciones, etc.). Esto no es necesario para el input, ya que las cosas empiezan precisamente cuando se encuentra el primer 1. Sin embargo, para el output se necesita algo más, puesto que podríamos no saber *a priori* hasta dónde mirar en la cinta output para alcanzar el primer 1 (esto es, el que está más a la izquierda). Incluso aunque hubiéramos encontrado una larga cadena de ceros que se extiende hacia la izquierda, esto no garantizaría que no hubiera un 1 aún *más lejos* hacia la izquierda. Podemos adoptar diversos puntos de vista sobre esto. Uno de estos consistiría en utilizar siempre una marca especial (digamos, codificada por 6 en el procedimiento de contracción) para iniciar el output completo. Pero, por simplicidad, adoptaré un punto de vista diferente en mis descripciones: el de que siempre se «conoce» cuanta cinta ha sido realmente recorrida por el dispositivo (v.g. podemos imaginar que deja una «estela» de algún tipo), de modo que, en principio, no tenemos que examinar una cantidad infinita de cinta para estar seguros de que ha sido revisado todo el output.
4. Un modo de codificar la información de dos cintas en una sola cinta consiste en intercalar las cintas. Así, las marcas impares en la cinta simple representarían las marcas de la primera cinta, mientras que las marcas pares representarían las marcas de la segunda cinta. Un esquema análogo funciona para tres o más cintas. La «ineficacia» de este procedimiento se deriva del hecho de que el dispositivo de lectura tendría que estar fintando hacia adelante y atrás y dejar marcadores en la cinta para conservar la huella de dónde está, tanto en las partes pares como en las partes impares de la cinta.
5. Este procedimiento se refiere sólo al modo en que una cinta marcada

puede interpretarse como un número natural. No altera los números de nuestras máquinas de Turing específicas, tales como EUC o $XN + 1$.

6. Si T_n no está correctamente especificada, entonces U procederá como si el número para n hubiera terminado en cuanto se alcanza la primera cadena de cuatro 1s o más en la expresión binaria de n . Leerá el resto de esta expresión como parte de la cinta para m , ¡de modo que procederá a ejecutar un cálculo sin sentido! Podríamos eliminar esta característica, si quisiéramos, disponiendo que n se exprese en notación binaria *expandida*. ¡No he querido hacer esto para no complicar más la descripción de la pobre máquina universal de Turing U !
7. Debo agradecer a David Deutsch el que haya derivado la forma decimal de la descripción binaria para u que he calculado más abajo. También le agradezco el haber comprobado que este valor binario de u da efectivamente una máquina universal de Turing. El valor binario para u es de hecho:

```

1 0 0 0 0 0 0 0 0 1 0 1 1 1 0 1 0 0 1 1 0 1 0 0 0 1 0 0 1 0 1 0 1 0 1 1 1 0
1 0 0 0 1 1 0 1 0 0 0 1 0 1 0 0 0 0 0 1 1 0 1 0 1 0 0 1 1 1 0 1 0 0 0 1 0 1 0
1 0 0 1 0 1 1 0 1 0 0 0 0 1 1 0 1 0 0 0 0 1 0 1 0 1 1 0 1 0 0 1 0 0 1 1 1 1 0
1 0 0 1 0 1 0 0 1 0 0 1 0 1 1 1 0 1 0 1 0 0 0 1 1 1 0 1 0 1 0 1 0 0 1 0 0 1
0 1 0 1 1 1 0 1 0 1 0 1 0 0 1 1 0 1 0 0 0 1 0 1 0 0 0 1 0 1 0 1 1 0 1 0 0 0
0 0 1 1 0 1 0 0 1 0 0 0 0 0 1 0 1 0 1 1 0 1 0 0 0 1 0 0 1 1 1 0 1 0 0 1 0 1
0 0 0 0 1 0 1 0 1 1 1 1 0 1 0 0 1 0 0 0 1 1 1 0 1 0 0 1 0 1 0 1 0 0 0 0 1 0 1
1 1 0 1 0 0 1 0 1 0 0 1 1 0 1 0 0 0 0 1 0 0 0 0 1 1 1 0 1 0 1 0 0 0 0 1 1 1
0 1 0 1 0 0 0 0 1 0 0 1 0 0 1 1 1 0 1 0 0 0 1 0 1 0 1 0 1 1 0 1 0 1 0 0 1 0
1 0 1 1 0 1 0 0 0 0 0 1 1 0 1 0 1 0 1 0 0 1 0 1 1 0 1 0 0 1 0 0 1 0 0 0 1 1
0 1 0 0 0 0 0 0 0 0 1 1 0 1 0 0 0 0 0 0 1 1 1 0 1 0 1 0 0 1 0 1 0 1 0 1 0 1
1 1 0 1 0 0 0 0 1 0 0 1 1 1 0 1 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 1 1 0 1 0
0 0 0 1 0 1 0 1 0 1 1 1 0 1 0 0 0 0 1 0 1 0 0 0 1 0 1 1 1 0 1 0 0 0 1 0 1 0
0 1 1 0 1 0 0 1 0 0 0 0 1 0 1 0 0 1 1 0 1 0 0 1 0 1 0 0 1 0 0 1 1 0 1 0 0 1
0 0 0 1 0 1 1 1 0 1 0 1 0 0 0 1 0 1 1 1 0 1 0 0 1 0 0 1 0 1 0 1 1 1 0 1 0 0 1
0 1 0 0 0 1 1 1 0 1 0 1 0 0 1 0 1 0 0 1 0 0 1 1 1 0 1 0 1 0 1 0 1 0 0 0 0 1
1 0 1 0 0 1 0 1 0 1 0 1 0 1 1 1 0 1 0 1 0 0 1 0 0 0 1 0 1 1 0 1 0 1 0 0 0 0
1 0 1 1 0 1 0 1 0 0 0 1 0 0 1 1 0 1 0 1 0 1 0 1 0 1 0 0 0 1 0 1 1 0 1 0 0 1
0 1 0 1 0 0 1 0 0 1 0 1 1 1 0 1 0 1 0 0 1 0 0 1 0 1 1 1 0 1 0 1 0 1 0 0 1 0 1
0 1 1 1 0 1 0 1 0 0 1 0 1 0 0 1 1 0 1 0 1 0 1 0 0 0 0 1 1 1 0 1 0 0 0 1 0 0
1 0 0 1 0 1 0 1 1 1 0 1 0 1 0 1 0 0 1 0 1 0 1 1 1 0 1 0 1 0 1 0 0 0 0 0 1 1
1 0 1 0 1 0 0 1 0 0 0 0 0 1 1 0 1 0 1 0 1 0 1 0 0 1 0 1 1 1 0 1 0 1 0 0 1 0
1 0 1 1 0 1 0 0 0 1 0 0 1 0 0 0 1 1 1 0 1 0 0 0 0 0 0 0 1 1 1 0 1 0 0 1 0 1
0 0 1 0 1 0 1 0 1 0 1 1 1 0 1 0 0 1 0 1 0 0 1 0 0 1 0 1 0 1 1 1 0 1 0 0 0 0
0 1 0 1 0 1 1 1 0 1 0 0 0 0 1 0 0 0 1 1 1 0 1 0 0 0 0 0 1 0 1 0 1 0 0 1 1 1
0 1 0 0 0 0 1 0 1 0 0 1 1 1 0 1 0 0 0 0 0 1 0 0 0 1 0 1 1 1 0 1 0 0 0 1 0 0
0 0 1 1 1 0 1 0 0 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 0 0 0 1 0 0 0 0 1 0 1 1 0 1
0 0 0 1 0 1 0 0 1 0 1 1 1 0 1 0 0 0 1 0 1 0 0 1 0 1 1 0 1 0 0 1 0 0 0 0 0 1
0 1 1 0 1 0 0 0 1 0 1 0 1 0 0 1 0 0 1 1 0 1 0 0 0 1 0 1 0 1 0 1 0 1 1 1 0 1

```


001000000111010010010101010111010101010
01101001000101011010010010010110100000
00101101000001000110100000100101101000
00000011010010100010111010010101000110
10010100101011010000010011101001010100
10110100100111010100000010101110101000
00011010101000101010110100101010110101
00010111010101011101010011010100011101
00001101010010001011010100101001101010
10001011101010010100101110101010000010
11101010100000101110100000011101010100
0010101110100101010110101010100001011101
010001010101110101010010010010110101010
00011101010000000111010010010000110100
10010001011010101010100111010000000010
11010010000110101010101001011101001000
01101001000101010111010000100011101000
10000111010000110100000001011010000010
01011101010100101010110100010001001011
10100000100111010101001101000001010101
10100001000011101001000010001110101010
10101001110100001001001110100010010000
11101000010100101101000010100001110101
01010101011101000100100110100010010011
01010010100101110100010001010111010000
00011101000100100101110100110100100100
00101101010101001101000101000101110100
00110101000010001011010100110101010010
10010110101010011010010010101110100110
10010000010110100010101010001110100100
00101011010000001001101001000100101110
1001000011010101000001001011101001001010
01101001001010101101001101001001010010
11010011010010100000101101001000001110
1010010011010101010100001010111010100001
01110100101010101110101000100101101001
0011101001010101000101110100010011101010
00010110100100111010010101010101110100
10001110100101010100101110100100011101
01000001010101110011010100000101101001
00111010100000010111010100101101010000
01010110100101001011101010000100101110
10000110101000100001011010100110101000
1000101101010101010010111010100010100101
10100010101010111010010000101011010100

01011101010010010101011101010100100101
11010100011101010001110101001001001011
10101000111010100101000101110101000101
11010100001001011101010001110100010100
01011101001010010111010100101010010111
01001010101010101101010000101010101101
00001001110100001010101010111010101000
10101110101010001010111010000001110101
01000100101110100000011101010100100010
1110101000000110101010000101101000000111
01001000000101110101000111010100100010
10111010100110101010100010101101000001
10101010100101010110100000010011010101
01001001110101001101010101001001011010
10011010010010011101000001101010101010
01010110101000100110100010100101010111
0100000110101010101010100101101000100011
10100010101010101011010001000111010000
10101110100010010000111010011010000000
10011101000000100101110100010001010011
10100000010010111010010101010100101101
00001010101011101000100101001011101000
00100010111010101001011010001000100111
01000001001010111010000001010101101000
01000111001111010000100000111010000100
10011101000001010010111010000010100101
10100001000101011101000010001001101000
10000111010111101000010010010111010000
10010010111010000000101011101000010101
00011010001001011101000010000011101000
01001110100010000010111010101001011010
00100000101110100001010101011101000000
10101011101000100001010111010001000010
10111010010000011101010010010011010000
00101011101000100010010111010101000011
10101001010110100101010100001101000001
01001101000000011101000001001001110100
10110100100010100101101010100110100010
10010010110101010011010001010100010110
01101010010010010111010101001101000101010
10101100110101000101010110011010010001
01010101110100010001110100100101010101
01101001010010100011010010000001011101
00000110101010010101010110100101010110
100100010001011110100010101011010100000

```

1 0 1 0 1 1 0 1 0 0 0 1 0 0 0 0 0 1 1 0 1 0 0 1 0 0 0 1 0 1 0 1 1 0 1 0 0 0
0 1 0 0 1 1 1 0 1 0 1 0 0 1 0 1 0 1 0 1 0 1 1 1 0 1 0 0 1 0 1 1 0 1 0 0
1 0 0 1 0 0 0 1 0 1 0 1 1 0 0 1 1 0 1 0 0 1 0 0 1 0 0 1 0 1 0 1 0 1 1 1 0 1
0 0 1 1 0 1 0 0 1 0 0 1 0 0 1 0 1 0 1 1 0 1 0 0 1 0 1 1 0 1 0 0 1 0 0 1 0 0
1 0 0 1 0 1 1 0 1 0 0 1 0 1 1 0 1 0 0 1 0 0 1 0 1 0 0 0 1 0 1 1 0 0 1 1 0 1
0 0 1 0 0 1 0 1 0 0 1 0 1 0 1 1 1 0 1 0 0 0 1 0 1 0 1 1 1 0 1 0 0 1 0 0 1 0
1 1 1 0 0 1 1 0 1 0 0 1 0 0 1 0 1 0 1 0 0 1 0 1 1 1 0 0 1 1 0 1 0 0 1 0 1 0
0 0 1 0 1 0 1 0 1 1 1 0 1 0 0 0 1 0 0 0 1 1 1 0 1 0 0 0 0 1 0 1 0 0 1 0 1 1
0 1 0 0 1 0 1 0 0 0 1 0 1 1 1 0 1 0 0 1 0 1 0 0 0 1 0 1 0 1 1 1 0 1 0 0 0 1 0
0 1 1 1 0 1 0 0 0 1 0 1 0 0 0 1 0 0 1 0 1 1 1 0 1 0 0 0 1 0 0 1 1 1 0 1 0 0 1
0 1 0 0 1 0 0 0 1 0 1 1 1 0 0 1 1 0 1 0 0 1 0 0 0 1 0 0 0 1 1 1 0 1 0 0 0 1
0 0 1 1 1 0 1 0 0 1 0 1 0 0 1 0 1 0 1 0 1 1 1 0 0 1 1 0 1 0 0 1 0 1 0 0 0 0
0 1 1 1 0 0 1 1 0 1 0 1 0 1 0 1 0 1 0 1 1 0 1 0 0 0 0 0 0 0 1 1 1 0 1 0 0 1
0 1 0 1 0 0 1 0 1 0 1 0 1 0 1 1 1 1 0 1 0 0 1 0 0 0 1 1 1 0 1 0 0 1 0 1 0 1 0 0
1 0 1 0 1 1 1 0 0 1 1 0 1 0 0 0 0 1 0 1 0 0 1 0 0 1 1 0 0 1 1 0 1 0 1 0 0 0 0
0 0 1 1 0 1 0 0 0 0 0 0 0 0 1 1 1 0 1 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 1 1 0 0
1 1 0 1 0 1 0 0 0 1 0 0 0 0 1 1 0 1 0 0 0 0 0 0 0 1 1 1 0 1 0 0 0 1 0 0 1 0
1 0 1 0 1 0 1 1 1 0 1 0 0 0 1 0 0 0 1 1 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0
1 1 0 1 0 0 0 0 1 0 0 1 1 1 0 1 0 0 1 0 0 0 1 0 0 1 0 1 0 1 1 1 0 1 0 0 1 0
1 0 1 0 0 0 1 0 0 1 1 0 1 0 1 0 0 0 0 0 0 0 1 0 1 1 0 1 0 0 1 0 0 1 1 1 0 1
0 1 0 0 0 0 1 0 1 0 1 1 1 0 1 0 0 1 0 0 0 0 1 1 0 1 0 1 0 0 0 0 0 0 0 1 0 1
1 0 1 0 0 1 0 0 0 1 1 1 0 1 0 0 1 0 0 1 0 1 1 1 0 1 0 0 0 0 1 1 0 1 0 1 0 0
0 0 1 0 1 0 1 0 1 1 0 1 0 1 0 0 0 1 0 1 1 1 0 1 0 1 0 0 0 0 1 0 1 0 0 1 0 1
1 1 0 1 0 1 0 0 0 1 0 1 1 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 0 1 1 1 0 0 1 0 1 1
0 1 0 0 0 1 0 1 0 1 1 0 1 0 0 0 0 1 1 0 1 0 1 0 0 0 1 0 0 1 0 1 0

```

El lector animoso, que tenga un computador en casa, puede ocuparse en comprobar, utilizando las recetas dadas en el texto, que el código anterior da realmente una acción de una máquina universal de Turing, aplicándola a diversas máquinas de Turing de números sencillos.

Hubiera sido posible una rebaja importante del valor de u con una especificación diferente para una máquina de Turing. Por ejemplo, podríamos prescindir de STOP y adoptar en su lugar la regla de que la máquina se para cuando vuelve a entrar en el estado interno o después de que haya estado en algún otro estado interno. Esto no sería una gran ganancia (suponiendo que lo fuese). Resultaría una ganancia mayor si permitiésemos cintas con otras marcas además de 0 ó 1. Efectivamente se han descrito en la literatura máquinas de Turing de apariencia muy concisa, pero esta concisión es algo decepcionante ya que dependen de codificaciones extraordinariamente complejas para las descripciones de las máquinas de Turing en general.

8. Para una discusión no técnica de las cuestiones relacionadas con este famoso enunciado, ver Devlin (1988).
9. Por supuesto, también podríamos derrotar a este algoritmo mejorado aplicando simplemente el procedimiento precedente una vez más. Podemos

entonces utilizar este nuevo conocimiento para mejorar aún más nuestro algoritmo; pero también podríamos derrotar a este otro, y así sucesivamente. El tipo de consideración a que nos lleva este procedimiento iterativo será discutido en relación con el teorema de Gödel, en el Capítulo 4, cf. p. 109.

Notas del Capítulo 3

1. Ver Mandelbrot (1986). La particular secuencia de ampliaciones que he escogido está adaptada de las de Peitgen y Richter (1986), en donde se encontrarán muchas imágenes notablemente coloreadas del conjunto de Mandelbrot. Para más ilustraciones sorprendentes, ver Peitgen y Saupe (1988).
2. Por lo que yo conozco, es un punto de vista consistente, aunque poco convencional, el exigir que hubiera siempre *algún* tipo de regla que determine cuál es realmente el n -ésimo dígito para un número real arbitrario, aunque esa regla pueda no ser efectiva ni siquiera definible en un sistema formal preasignado (ver Capítulo 4). ¡Yo espero que *sea* consistente, puesto que es el punto de vista al que más me gustaría adherirme!

Notas del Capítulo 4

1. Al considerar conjuntos cuyos miembros pueden ser a su vez conjuntos debemos tener cuidado en distinguir entre los miembros de ese conjunto y los miembros de los *miembros* de dicho conjunto. Por ejemplo, supongamos que S es el conjunto de los *subconjuntos no-vacíos* de otro cierto conjunto T , en donde los miembros de T son una manzana y una naranja. T tiene la propiedad de «paridad» y no «trinidad», pero S tiene realmente la propiedad «trinidad»; en efecto, los tres miembros de S son: un conjunto que sólo contiene una manzana, un conjunto que sólo contiene una naranja, y un conjunto que contiene una manzana y una naranja —tres conjuntos en total, que son los *tres* miembros de S . Análogamente, el conjunto cuyo único miembro es el *conjunto vacío* posee la propiedad de «unidad» no de «nulidad» —¡tiene un miembro, a saber, el conjunto vacío! Por supuesto, *el propio* conjunto vacío tiene cero miembros.
2. De hecho, el razonamiento en el teorema de Gödel puede presentarse de forma que no dependa de un concepto totalmente externo de «verdad» para proposiciones tales como $P_k(k)$. Sin embargo, depende todavía de una interpretación del «significado» real de *algunos* de los símbolos: en particular de que « $\sim\exists$ » *significa* realmente que «no existe ningún (número natural)... tal que...».
3. En lo que sigue, las letras minúsculas representan números naturales, y las mayúsculas, conjuntos finitos de números naturales. Sea $m \rightarrow [n, k, r]$

la representación del enunciado «si x es cualquier conjunto de m -elementos de números naturales cuyos subconjuntos de k -elementos son asignados a r cajas, entonces existe un «gran» subconjunto de n -elementos Y de X tal que todos los subconjuntos de k -elementos de Y entran en la misma caja». «Grande» significa aquí que Y tiene más elementos que el número natural que es el elemento más pequeño de Y . Consideremos la proposición: «Para cualquier elección de k , r y n existe un m_0 tal que, para todo m mayor que m_0 , el enunciado $m \rightarrow [n, k, r,]$ es siempre verdadero». J. Paris y L. Harrington (1977) han demostrado que esta proposición es equivalente a una proposición tipo-Gödel para los axiomas estándar (Peano) de la aritmética, indemostrable a partir de estos axiomas aunque afirma algo acerca de estos axiomas que es «obviamente verdadero» (a saber, en este caso, que las proposiciones deducibles de los axiomas son ellas mismas verdaderas).

4. El título era «Sistemas de lógica basados en ordinales», y algunos lectores estarán familiarizados con la notación para los *números ordinales* de Cantor que he estado utilizando en los subíndices. La jerarquía de sistemas lógicos que se obtiene mediante el procedimiento que he descrito arriba está caracterizada por *números ordinales computables*.

Existen algunos teoremas matemáticos que son bastante naturales y fáciles de enunciar para los que, si intentamos demostrarlos utilizando las reglas estándar (Peano) de la aritmética, tendríamos que utilizar el procedimiento de «gödelización» anterior hasta un grado escandalosamente grande (extendiendo el procedimiento enormemente más allá de lo que he esbozado arriba). Las demostraciones matemáticas de estos teoremas no son en absoluto del tipo de las que dependen de cualquier vago o cuestionable razonamiento que pareciera estar fuera de los procedimientos de la argumentación matemática normal. Ver Smorynski (1983).

5. La hipótesis del continuo que fue mencionada en el Capítulo 3, p. 85 (y que establece que $C = \aleph_1$) es el enunciado matemático más «extremo» que hemos encontrado aquí (aunque con frecuencia se consideran enunciados mucho más extremos que éste). La hipótesis del continuo tiene un interés adicional debido a que el propio Gödel, junto con Paul J. Cohen, estableció que la hipótesis del continuo es en realidad *independiente* de los axiomas y reglas de inferencia estándar de la teoría de conjuntos. Por consiguiente, la actitud de cada uno hacia el status de la teoría del continuo distingue entre los puntos de vista formalista y platónico. Para un formalista la hipótesis del continuo es «indecidable» puesto que no puede ser establecida ni refutada utilizando el sistema formal estándar (Zermelo-Frankel), y constituye un «sinsentido» el llamarla «verdadera» o «falsa». Sin embargo, para un buen platónico la hipótesis del continuo es realmente o verdadera o falsa, aunque establecer cuál de los dos es el caso requeriría alguna forma nueva de razonamiento —que va incluso más allá del empleo de proposiciones tipo-Gödel para el sistema formal de Zermelo-Frankel. (¡El propio Cohen (1966) sugirió un principio de

reflexión que hace la hipótesis del continuo «obviamente falsa»!)

6. Para un informe vivo y no-técnico sobre estas cuestiones, ver Rucker (1984).
7. El propio Brower parece haber partido de esta línea de razonamiento debido en parte a las quejas y críticas acerca de una «no-constructividad» en su demostración de uno de sus propios teoremas, «el teorema del punto fijo de Brower» de la topología. El teorema afirma que si se toma un disco —es decir, un círculo junto con su interior— y se deforma de una forma continua en el interior de la región en la que estaba situado inicialmente, entonces existe al menos un punto del disco —llamado un punto fijo— que termina exactamente donde empezó *. Podemos no tener idea de cuál es exactamente este punto, o de si pudiera haber varios de estos puntos; lo que afirma el teorema es simplemente la *existencia* de alguno de estos puntos. (Como son los teoremas de existencia en matemáticas, este es en verdad completamente «constructivo». De un orden diferente de no-constructividad son los teoremas de existencia que dependen de lo que se conoce como el «Axioma de Elección» o «lema de Zorn» (cf. Cohen 1966, Rucker 1984). En el caso de Brower la dificultad es similar a la siguiente: si f es una función continua real de variable real que toma valores positivos y negativos, encontrar un lugar donde f se anula. El procedimiento usual supone la bisección repetida del intervalo en el que f cambia de signo, pero puede no ser «constructivo», en el sentido requerido por Brower, el decidir si los valores intermedios de f son positivos, negativos o cero.
8. Enumeramos los conjuntos $\{v, w, x, \dots, z\}$, donde v representa la función f según algún esquema lexicográfico. Hacemos una comprobación (recursivamente) en cada paso para ver si $f(w, x, \dots, z) = 0$ y retenemos la proposición $\exists w, x, \dots, z [f(w, x, \dots, z) = 0]$ sólo en caso de que sea así.
9. Existe una nueva teoría de la computabilidad para funciones reales de números reales (en contraste con las convencionales funciones que toman valores naturales para números naturales), debida a Blum, Shub y Smale (1989), cuyos detalles sólo muy recientemente han venido a mi atención. Esta teoría se aplicaría también a funciones complejas y podría tener una relación importante con algunos de los temas planteados en el texto. Algunos de los resultados de este nuevo trabajo prestan un fuerte apoyo a la conjetura de que el conjunto de Mandelbrot es realmente no-recursivo según una definición apropiada.
10. Este problema particular se denomina con más propiedad «el problema de las palabras para semigrupos». Existen también otras formas del problema de las palabras, en las que las reglas son ligeramente diferentes. Estas no nos interesan aquí.

* Una versión intuitiva de este teorema es que si se quita un mantel que recubre una mesa circular, se arruga sin desgarrarlo y se tira así sobre la mesa, habrá al menos un punto del mantel que no habrá cambiado de lugar (citado del «Diccionario de Matemáticas», dirigido por François le Lionnais, Editorial Akal). [Nota del traductor.]

11. Hanf (1974) y Myers (1974) han demostrado, además, que hay un solo conjunto (de un gran número de teselas) que teselará el plano sólo de un modo *no-computable*.
12. De hecho, mediante el uso del ingenio, este número de pasos puede reducirse a algo del orden de $n \log n \log \log n$ para n grande —que, por supuesto, sigue estando en **P**. Ver Knuth (1981) para más información sobre estas materias.

Notas del Capítulo 5

1. Es sorprendente el hecho de que *todas* las desviaciones establecidas de la imagen newtoniana han estado asociadas, de alguna forma esencial, con el comportamiento de la *luz*. En primer lugar, están los campos incorpóreos que transportan energía en la teoría electromagnética de Maxwell. En segundo lugar está, como veremos, el papel crucial que juega la velocidad de la luz en la teoría de la relatividad especial de Einstein. En tercer lugar, las pequeñísimas desviaciones de la teoría gravitatoria de Newton que exhibe la teoría de la relatividad *general* de Einstein, se hacen significativas sólo cuando las velocidades son comparables con la de la luz. (La desviación de la luz por el Sol, el movimiento de Mercurio, las velocidades de escape comparables a la de la luz para agujeros negros, etc.) En cuarto lugar, está la dualidad onda-corpúsculo de la teoría cuántica, observada en primer lugar en el comportamiento de la luz. Finalmente está la electrodinámica cuántica, que es la teoría cuántica de la luz y las partículas cargadas. Es razonable especular que el propio Newton hubiera estado dispuesto a aceptar que había profundos problemas para su imagen del mundo ocultos tras el misterioso comportamiento de la luz cf. Newton (1730); también Penrose (1987a).
2. Hay un magnífico cuerpo de conocimientos físicos bien establecido —la *termodinámica* de Carnot, Maxwell, Kelvin, Boltzmann y otros— que he omitido clasificar. Esto puede intrigar a algunos lectores pero la omisión ha sido deliberada. Por razones que se harán evidentes en el Capítulo 7, yo mismo sería bastante reacio a colocar la termodinámica, tal como está, en la categoría de las teorías SOBERBIAS. ¡Sin embargo, muchos físicos considerarán probablemente un *sacrilegio* colocar un cuerpo de ideas tan bello y fundamental en una categoría tan baja como la de simplemente ÚTILES. En mi opinión, la termodinámica, como se entiende normalmente, siendo algo que se aplica sólo a *promedios* y no a los constituyentes individuales de un sistema —y siendo parcialmente una deducción de otras teorías— no es una teoría física en el sentido en que lo entiendo aquí (lo mismo se aplica a la estructura matemática subyacente de la *mecánica estadística*). Estoy utilizando este hecho como una excusa para evitar el problema y dejarlo fuera de la clasificación. Como veremos en el Capítulo 7, estoy afirmando una íntima conexión entre la termodi-

námica y un tema que he citado antes como perteneciente a la categoría de UTILES, a saber el modelo estándar para el big bang. Una unión apropiada entre estos dos conjuntos de ideas (que falta en parte actualmente) debería ser considerada, eso creo, como una teoría física en el sentido exigido —incluso perteneciente a la categoría de SOBERBIAS. Esto es algo a lo que tendré que volver más tarde.

3. Mis colegas me han preguntado dónde colocaría la «teoría de twistors» —una elaborada colección de ideas y procedimientos con la que yo mismo he estado asociado durante un gran número de años. En la medida en que una teoría de twistor *es* una teoría diferente sobre el mundo físico, no puede ser en otra que en la categoría de TENTATIVAS; pero en buena medida no es una teoría en absoluto, siendo una descripción matemática de teorías físicas previamente bien establecidas.
4. El nombre de Newton se asocia a este modelo —y en realidad a la mecánica «newtoniana» en general— simplemente como una *etiqueta* conveniente. Las propias opiniones de Newton respecto a la *verdadera* naturaleza del mundo físico parecen haber sido mucho menos dogmáticas y más sutiles que esto. (La persona que promovió con más fuerza este modelo «newtoniano» parece haber sido R. G. Boscovich, 1711-1787.)
5. Como se sugirió en el Capítulo 4 (nota 9, p. 566), la nueva teoría de Blum-Shub-Smale (1989) quizá proporcione un modo de resolver algunos de estos puntos de una forma más aceptable matemáticamente.
6. Las auténticas ecuaciones de Hamilton, aunque quizá no su punto de vista particular, eran ya conocidas para el gran matemático italo-francés Joseph C. Lagrange (1736-1813) unos 24 años antes que Hamilton. Un desarrollo anterior igualmente importante fue la formulación de la mecánica en términos de las *ecuaciones de Euler-Lagrange*, según las cuales las ecuaciones de Newton pueden verse como deducidas de un principio superior: el *principio de acción estacionaria* (P.L.M. de Maupertuis). Además de su gran importancia teórica, las ecuaciones de Euler-Lagrange proporcionan métodos de cálculo de considerable potencia y valor práctico.
7. La situación es, en realidad, «peor» en el sentido de que el volumen del espacio de fases de Liouville es sólo uno entre una familia entera de «volúmenes» de diferentes dimensiones (conocidos como invariantes de Poincaré) que permanecen constantes bajo evoluciones hamiltonianas. Sin embargo, he sido un poco injusto en la radicalidad de mis afirmaciones. Podemos imaginar un sistema en el que grados de libertad físicos (que contribuyen a parte del volumen del espacio de fases) pueden estar «tirados» en alguna parte en la que no estamos interesados (como radiación que escapa al infinito) de modo que el volumen en el espacio de fases de la parte que *nos interesa* podría reducirse.
8. Este segundo hecho, en particular, resulta de tremenda buena fortuna para la ciencia pues, sin ello, el comportamiento dinámico de los cuerpos grandes podría haber sido incomprensible y hubiera dado pocas pistas

sobre las leyes exactas que se aplicarían a las propias partículas. Mi conjetura es que una razón para la fuerte insistencia de Newton en su tercera ley era que, sin ella, esta transferencia de comportamiento dinámico desde los cuerpos microscópicos a los macroscópicos simplemente no se cumpliría.

Otro hecho «milagroso», que fue vital para el desarrollo de la ciencia, es que la ley del inverso del cuadrado es la única ley de potencias (decreciente con la distancia) para la que las órbitas generales en torno a un cuerpo central son formas geométricas sencillas. ¿Qué hubiera hecho Kepler si la ley de fuerzas hubiera sido una ley del inverso de la distancia o del inverso del cubo?

9. En efecto, tenemos un número *infinito* de x_i y p_i ; pero existe la complicación adicional de que no podemos utilizar simplemente los valores de los campos en esas coordenadas, siendo necesario un cierto «potencial» para el campo de Maxwell para que pueda aplicarse el esquema hamiltoniano.
10. Es decir, no dos veces diferenciable.
11. Las ecuaciones de Lorentz nos dicen cuál es la *fuerza* sobre una partícula cargada, debida al campo electromagnético en el que está inmersa; entonces, si conocemos su masa, la segunda ley de Newton nos dice cuál es la aceleración de la partícula. Sin embargo, las partículas cargadas se mueven con frecuencia a velocidades próximas a la de la luz y los efectos de la relatividad especial comienzan a ser importantes, lo que afecta a cuál debe ser la masa de la partícula que debemos tomar (ver la próxima sección). Fueron razones de este tipo las que retrasaron el descubrimiento de la correcta ley de fuerzas sobre una partícula cargada hasta el nacimiento de la relatividad especial.
12. De hecho, en cierto sentido, cualquier partícula mecano-cuántica en la Naturaleza actúa por sí misma como tal reloj. Como veremos en el Capítulo 6, existe una oscilación asociada a cada partícula cuántica, cuya frecuencia es proporcional a la masa de la partícula; ver p. 283. Los relojes modernos de gran precisión (relojes atómicos, relojes nucleares) dependen en última instancia de este hecho.
13. El lector podría estar preocupado por el hecho de que, puesto que en la línea-de-universo del viajero hay una «esquina» en B , el viajero sufre, como se muestra, una aceleración infinita en el suceso B . Esto es accesorio. Con una aceleración finita, la línea-de-universo del viajero tiene la esquina en B suavizada, y esto apenas supone diferencia en el tiempo total que experimenta, que aún se mide por la «longitud» minkowskiana de la línea-de-universo total.
14. Estos son los espacios de sucesos que M juzgaría simultáneos según la *definición de simultaneidad de Einstein*, que utiliza las señales luminosas enviadas por M y devueltas a M desde los sucesos en cuestión. Ver, por ejemplo, Rindler (1982).
15. Esta es la derivada *segunda* (o «aceleración») inicial de la forma con

- respecto al tiempo. El ritmo de variación (o «velocidad») de la forma se toma inicialmente nula, puesto que la esfera está en reposo.
16. La descripción matemática de esta reformulación de la teoría de Newton fue llevada a cabo por primera vez por el célebre matemático francés Élie Cartan (1923) —lo que, por supuesto, fue *después* de la teoría de la relatividad general de Einstein.
 17. Los espacios curvos que son localmente euclideos en este sentido (también en mayores dimensiones) se llaman *variedades de Riemann* —en honor del gran Bernhard Riemann (1826-1866), quien investigó por primera vez estos espacios, siguiendo algunos importantes trabajos anteriores de Gauss para el caso bi-dimensional. Aquí necesitamos una modificación significativa de la idea de Riemann, a saber, permitir que la geometría sea localmente *minkowskiana* en lugar de euclidea. Tales espacios se denominan a menudo variedades *lorentzianas* (que pertenecen a una clase llamada *pseudo-riemanniana* o, menos lógicamente, variedades *semi-riemannianas*).
 18. ¡El lector puede estar preocupado por cómo este valor *cero* puede representar el valor *máximo* de la «longitud»! De hecho lo es, pero en un sentido vacuo: una geodésica de longitud-cero se caracteriza por el hecho de que no existen líneas-de-universo de *ninguna otra* partícula que conecten cualquier par de sus puntos (localmente).
 19. En realidad esta división en efectos distorsionantes y cambio de volumen no es ni mucho menos tan clara como la he presentado. El mismo tensor de Ricci puede dar una cierta cantidad de distorsión de marea. (Con rayos de luz la división *es* completamente tajante; cf. Penrose y Rindler (1986), Capítulo 7.) Para una definición precisa de los tensores de Weyl y Ricci ver, por ejemplo, Penrose y Rindler (1984), pp. 240, 210. (Hermann Weyl, de origen alemán, fue una figura matemática sobresaliente de este siglo; el italiano Ricci fue un geómetra muy influyente que fundó la teoría de tensores en el siglo pasado.)
 20. La forma correcta de las auténticas ecuaciones fue encontrada también por David Hilbert en noviembre de 1915, pero las ideas físicas de la teoría son debidas por entero a Einstein.
 21. Para quienes sepan de tales materias, estas ecuaciones diferenciales son las *identidades de Bianchi* en las que se han substituido las ecuaciones de Einstein.
 22. Hay algunas vías (no muy satisfactorias) de evitar este argumento, cf. Wheeler y Feynman (1945).
 23. Técnicamente, el término «hipersuperficie» es más apropiado que «superficie» puesto que es tri-dimensional en lugar de bi-dimensional.
 24. Serían de gran utilidad e interés *teoremas* rigurosos sobre estos puntos. Por el momento faltan.
 25. Incalculable, en la teoría actual «que da la respuesta (provisional) inútil: ¡infinito!

1. He dado por supuesto que cualquier punto de vista filosófico «serio» debería contener al menos una buena dosis de realismo. ¡Siempre me sorprende al enterarme de que pensadores aparentemente serios, con frecuencia físicos interesados en las implicaciones de la mecánica cuántica, adoptan la visión fuertemente subjetiva de que, en realidad, no existe en absoluto *ningún* mundo real «ahí fuera»! El hecho de que yo adopte una línea realista siempre que sea posible no quiere decir que no sea consciente de que tales visiones subjetivas se defienden seriamente muy a menudo —simplemente soy incapaz de entenderlas. Para un poderoso y divertido ataque a este subjetivismo, ver Gardner (1983), Capítulo 1.
2. En particular, J. J. Balmer señaló, en 1885, que las frecuencias de las líneas espectrales del hidrógeno tenían la forma $R(n^{-2} - m^{-2})$ en donde n y m son enteros positivos (siendo R una constante).
3. Quizá no debiéramos descartar muy a la ligera esta imagen «enteramente de campos». Einstein, quien (como veremos) era profundamente consciente del carácter discreto que manifestaban las partículas cuánticas, dedicó los treinta últimos años de su vida a tratar de encontrar una teoría completamente global de este tipo clásico general. Pero los intentos de Einstein, como todos los demás, fracasaron. Parece necesitarse alguna otra cosa además de un campo clásico para explicar la naturaleza discreta de las partículas.
4. Estos dos procedimientos de evolución se describieron en una obra clásica del famoso matemático húngaro/americano John von Neumann (1955). Su «proceso 1» es el que yo he llamado $\mathbf{R}$ —«reducción del vector de estado»— y su «proceso 2» es $\mathbf{U}$ —«evolución unitaria» (lo que significa que las amplitudes de probabilidad se conservan efectivamente en la evolución). De hecho, existen otras —aunque equivalentes— descripciones de la evolución $\mathbf{U}$ del estado cuántico, en las que se podría no utilizar el término «ecuación de Schrödinger». En la «*imagen de Heisenberg*», por ejemplo, el estado se *describe* de modo que parece no evolucionar en absoluto, siendo considerada la evolución dinámica como un continuo desplazamiento de los significados de las coordenadas de posición/momento. Las diversas distinciones no nos interesan aquí pues las diferentes descripciones del proceso $\mathbf{U}$ son completamente equivalentes.
5. Para ser completos deberíamos especificar también todas las leyes algebraicas requeridas que, en la notación (de Dirac) utilizada en el texto, son:

$ \psi\rangle + \chi\rangle = \chi\rangle + \psi\rangle,$	$ \psi\rangle + (\chi\rangle + \varphi\rangle) = (\psi\rangle + \varphi\rangle) + \varphi\rangle,$
$(z + w) \psi\rangle = z \psi\rangle + w \psi\rangle,$	$z(\psi\rangle + \chi\rangle) = z \psi\rangle + z \chi\rangle,$
$z(w \psi\rangle) = (zw) \psi\rangle,$	$1 \psi\rangle = \psi\rangle,$
$ \psi\rangle + \mathbf{0} = \psi\rangle,$	$0 \psi\rangle = \mathbf{0}, \text{ y } z\mathbf{0} = \mathbf{0}.$
6. Hay una operación importante, conocida como *producto escalar* (o producto interno) de dos vectores, que puede ser utilizada para expresar los

conceptos de «vector unitario», «ortogonalidad» y «amplitud de probabilidad» de un modo muy simple. (En el álgebra vectorial ordinaria el producto escalar es $ab\cos\theta$, donde a y b son las longitudes de los vectores y θ es el ángulo entre sus direcciones.) El producto escalar entre vectores del espacio de Hilbert da un número *complejo*. Para dos vectores de estado $|\psi\rangle$ y $|\chi\rangle$ escribimos el producto escalar $\langle\psi|\chi\rangle$. Existen reglas algebraicas $\langle\psi(|\chi\rangle + |\varphi\rangle)\rangle = \langle\psi|\chi\rangle + \langle\psi|\varphi\rangle$, $\langle\psi(q|\chi\rangle)\rangle = q\langle\psi|\chi\rangle$, y $\langle\psi|\chi\rangle = \langle\bar{\chi}|\bar{\psi}\rangle$, en donde la barra denota conjugación compleja. (El complejo conjugado de $z=x+iy$, es $\bar{z}=x-iy$, siendo x e y reales; nótese que $|z|^2 = z\bar{z}$). La ortogonalidad entre $|\psi\rangle$ y $|\chi\rangle$ se expresa como $\langle\psi|\chi\rangle = 0$. El cuadrado de la longitud de $|\psi\rangle$ es $|\psi|^2 = \langle\psi|\psi\rangle$, de modo que la condición para que $|\psi\rangle$ esté normalizada como un vector unitario es $\langle\psi|\psi\rangle = 1$. Si un «acto de medida» provoca que un estado $|\psi\rangle$ salte o bien a $|\chi\rangle$ o a algún otro ortogonal a $|\chi\rangle$, entonces la amplitud de que salte a $|\chi\rangle$ es $\langle\chi|\psi\rangle$, suponiendo que $|\psi\rangle$ y $|\chi\rangle$ estén ambos normalizados. Si no están normalizados, la probabilidad de saltar de $|\psi\rangle$ a $|\chi\rangle$ puede escribirse $\langle\chi|\psi\rangle\langle\psi|\chi\rangle/\langle\chi|\chi\rangle\langle\psi|\psi\rangle$. (Ver Dirac, 1947.)

7. Para aquellos familiarizados con el formalismo de los operadores de la mecánica cuántica, esta medida se define (en la notación de Dirac) por el operador hermítico acotado $|\chi\rangle\langle\chi|$. El valor propio 1 (para $|\chi\rangle$ normalizado) significa **SÍ** y el valor propio 0 significa **NO**. (Los vectores $\langle\chi|$, $\langle\psi|$, etc., pertenecen al espacio *dual* del espacio de Hilbert original.) Ver von Neumann (1955), Dirac (1947).
8. En mis primeras descripciones de un sistema cuántico consistente en una sola partícula he hecho demasiadas simplificaciones al ignorar el spin y suponer que el estado puede describirse en términos de su posición únicamente. *Existen* realmente ciertas partículas —llamadas partículas *escalares*, de las que son ejemplo las partículas nucleares llamadas piones (mesones p , cf. p. 282), o ciertos átomos— para las que el valor del spin resulta ser cero. Para estas partículas (pero sólo para éstas) será suficiente la descripción anterior en términos de posición únicamente.
9. Tómese $|\bar{\chi}\rangle = \bar{z}|\uparrow\rangle - \bar{w}|\downarrow\rangle$, donde $\bar{z}$ y $\bar{w}$ son los complejos conjugados de z y w . (Ver nota 6.)
10. Hay un dispositivo experimental estándar, conocido como aparato de Stern-Gerlach, que puede ser utilizado para medir los spines de átomos apropiados. Los átomos se proyectan en un haz que atraviesa un campo magnético fuertemente inhomogéneo, y la dirección de la inhomogeneidad del campo proporciona la dirección de la medida del spin. El haz se desdobra en dos (para un átomo de spin 1/2, o en más de dos haces para un spin mayor): un haz que da los átomos para los que la respuesta a la medida del spin es **SÍ** y el otro para el que la respuesta es **NO**. Por desgracia, hay razones técnicas, irrelevantes para nuestros propósitos, por las que este aparato no puede ser utilizado para la medida del spin electrónico y debe utilizarse un procedimiento más indirecto. (Ver Mott y

Massey, 1965.) Por ésta y otras razones prefiero no ser muy concreto sobre el modo en que realmente se está midiendo el spin.

11. El lector decidido puede ocuparse en verificar la geometría dada en el texto. Resulta más fácil si orientamos nuestra esfera de Riemann de modo que la dirección α sea «arriba» y la dirección β está en el plano que determinan «arriba» y «derecha», esto es, dado por $q = \tan(\theta/2)$ en la esfera de Riemann, y luego usamos la receta $\langle \chi | \psi \rangle \langle \psi | \chi \rangle / \langle \chi | \chi \rangle \langle \psi | \psi \rangle$ para la probabilidad de saltar de $|\psi\rangle$ a $|\chi\rangle$. Ver nota 6.
12. En lenguaje matemático decimos que el espacio de los estados de dos partículas es el *producto tensorial* del espacio de estados de la primera partícula por el de la segunda partícula. El estado $|\chi\rangle|\varphi\rangle$ es entonces el producto tensorial del estado $|\chi\rangle$ y el $|\varphi\rangle$.
13. Wolfgang Pauli, un brillante físico austriaco y figura prominente en el desarrollo de la mecánica cuántica, propuso su principio de exclusión como una hipótesis en 1925. El tratamiento mecano-cuántico de lo que ahora llamamos «fermiones» fue desarrollado en 1926 por el muy influyente y original físico italo(-americano) Enrico Fermi y por el gran Paul Dirac, a quien ya hemos encontrado en varias ocasiones antes de ahora. El comportamiento estadístico de los fermiones sigue la «estadística de Fermi-Dirac», cuyo nombre la distingue de la «estadística de Boltzmann» —la estadística clásica de las partículas distinguibles. La «estadística de Bose-Einstein» para los bosones fue desarrollada para el tratamiento de los fotones por el famoso físico indio S. N. Bose y Albert Einstein en 1924.
14. Este es un resultado tan famoso e importante que vale la pena dar otra versión de él. Supongamos que hay sólo *dos* posiciones para el medidor E, arriba $[\uparrow]$ y derecha $[\rightarrow]$, y dos para el medidor-P 45° hacia arriba y a la derecha $[\nearrow]$ y 45° hacia abajo y a la derecha $[\searrow]$. Considérese que las posiciones *reales* son $[\rightarrow]$ y $[\nearrow]$, para los medidores E y P respectivamente. Entonces la probabilidad de que los medidores E y P den coincidencia es $\frac{1}{2}(1 + \cos 135^\circ) = 0,146\dots$, que es un poco menos del 15 por ciento. Una larga serie de experimentos con estas posiciones dadas, por ejemplo,

E: S N N S N S S S N S S N N S N N N S S N...

P: N S S N N N S N S N N S S N S S S N S N N S...

dará precisamente algo menos de un 15 por ciento de coincidencias. Supongamos ahora que las medidas-P no están influenciadas por la posición-E —de modo que *si* la posición-E hubiera sido $[\uparrow]$ en lugar de $[\rightarrow]$, entonces la serie de resultados-P habría sido exactamente la misma —y puesto que el ángulo entre $[\uparrow]$ y $[\nearrow]$ es el mismo que el ángulo entre $[\rightarrow]$ y $[\nearrow]$, habría otra vez algo menos de un 15 por ciento de coincidencias entre las medidas-P y las nuevas medidas-E, digamos E'. Por otro lado, si la posición-E hubiera sido $[\rightarrow]$, como antes, pero la posición-P fuera $[\searrow]$ en lugar de $[\nearrow]$, entonces la serie de resultados-E habría sido

la misma que antes pero los nuevos resultados-P, digamos P' , estarían algo por debajo del 15 por ciento de coincidencias con los resultados-E originales. Se sigue que no podría haber más de un 45 por ciento (15 por ciento + 15 por ciento + 15 por ciento) de coincidencias entre la medida- P' [$\searrow$] y la medida-E' [$\uparrow$] si hubieran sido *estas* las posiciones reales. Pero el ángulo entre [$\searrow$] y [$\uparrow$] es 135° y no 45° , de modo que la probabilidad de coincidencia *debería* ser justo algo mayor del 85 por ciento, no el 45 por ciento. ¡Esto es una contradicción que muestra que la hipótesis de que la elección de la medida a hacer en E no puede influir en los resultados de P (y *viceversa*) debe ser falsa! Estoy en deuda con David Mermin por este ejemplo. La versión dada en el texto principal está tomada de su artículo Mermin (1985).

15. Los primeros resultados fueron debidos a Freedman y Clauser (1972) basados en ideas sugeridas por Clauser, Horne, Shimony y Holt (1969). Existe todavía un punto de discusión, en estos experimentos, debido al hecho de que los detectores de fotones que se utilizan están muy lejos de alcanzar una eficiencia del 100 por ciento, de modo que sólo una fracción relativamente pequeña de los fotones emitidos son realmente detectados. Sin embargo, el acuerdo con la teoría cuántica es tan perfecto, con estos detectores relativamente ineficientes, que ¡es difícil ver como la mejora de los detectores vaya a producir un acuerdo *peor* con la teoría!
16. La teoría cuántica de campos parece ofrecer alguna perspectiva para la no-computabilidad (cf. Komar, 1964).

Notas del Capítulo 7

1. Algunos «puristas» de la relatividad prefieren utilizar los conos de luz de los observadores más que sus espacios simultáneos. Sin embargo, esto no supone la más mínima diferencia en las conclusiones.
2. Se gana entropía al combinar núcleos ligeros (v.g. de hidrógeno) en las estrellas para dar núcleos más pesados (v.g. de helio o, en última instancia, de hierro). Análogamente, existe mucha «pequeñez de entropía» en el hidrógeno presente en la Tierra, parte de la cual podemos utilizar convirtiendo el hidrógeno en helio en las centrales nucleares de «fusión». La posibilidad de ganar entropía por estos medios aparece únicamente debido a que la gravitación ha hecho posible que los núcleos se concentren al margen de los fotones mucho más numerosos que han escapado a la inmensidad del espacio y constituyen ahora la radiación de fondo de cuerpo negro de 2.7 K (cf. p. 404). Esta radiación contiene una entropía inmensamente mayor que la que está presente en la materia de las estrellas ordinarias, y si se concentrara de nuevo en el material de las estrellas serviría para ¡desintegrar nuevamente la mayor parte de estos núcleos pesados en sus partículas constituyentes! Por ello, la ganancia de entropía en la fusión es sólo «temporal», y se hace posible únicamente a través de la presencia de los efectos concentrantes de la gravedad. Veremos más adelante que,

incluso aunque la entropía disponible por la vía de la fusión de los núcleos es muy grande en relación con mucha de la que se ha obtenido hasta ahora directamente a través de la gravedad —y la entropía en la radiación de fondo es enormemente mayor—, éste es un estado de cosas puramente local y temporal. Las reservas de entropía de la gravitación son *enormemente* mayores que las de la fusión o las de la radiación de 2.7 K (cf. p. 428).

3. Alguna evidencia reciente a partir de los pozos ultra-profundos perforados en Suecia puede interpretarse como un apoyo a la teoría de Gold pero el tema es muy controvertido, habiendo explicaciones alternativas convencionales.
4. Estoy suponiendo que se trata de lo que se conoce como una supernova de «tipo II». Si hubiera sido una supernova de «tipo I» estaríamos pensando de nuevo en términos de la ganancia «temporal» de entropía proporcionada por la «fusión» (cf. nota 2). Sin embargo, es poco probable que las supernovas Tipo I produzcan mucho uranio.
5. He llamado a los modelos con curvatura nula o negativa modelos *infinitos*. Existen, no obstante, maneras de «plegar» estos modelos para hacerlos espacialmente finitos. Esta consideración —de no poco probable relevancia para el universo real— no afectaría en gran medida a la discusión, y no me voy a ocupar aquí de ello.
6. Las bases experimentales para esta confianza proceden principalmente de dos tipos de datos. En primer lugar, el comportamiento de las partículas cuando chocan entre sí a velocidades que son importantes y rebotan, se fragmentan y crean nuevas partículas, se conoce por medio de los aceleradores de partículas de alta energía construidos en varios lugares de la Tierra, y mediante el comportamiento de las partículas de los rayos cósmicos que inciden en la Tierra procedentes del espacio exterior. En segundo lugar, se sabe que los parámetros que gobiernan el modo de interacción de las partículas no han variado ni siquiera una parte en 10^6 , en 10^{10} años (cf. Barrow 1988), de modo que es altamente probable que no hallan cambiado de manera significativa (y probablemente nada en absoluto) desde el tiempo de la bola de fuego primordial.
7. En realidad el principio de Pauli no prohíbe que los electrones estén en el mismo «lugar» sino que prohíbe que dos de ellos estén en el mismo «estado» —que incluye también el modo en que los electrones se desplazan y giran. El argumento real es un poco delicado y fue objeto de mucha controversia, en particular por parte de Eddington, cuando fue desarrollado por primera vez.
8. Este razonamiento ya fue desarrollado en 1784 por el astrónomo inglés John Michell e, independientemente, por Laplace un poco más tarde. Ellos llegaron a la conclusión de que los cuerpos con más masa y más concentrados del universo podría ser totalmente invisibles —como agujeros negros— pero sus argumentos (ciertamente proféticos) utilizaban la teoría newtoniana por lo que su conclusiones son, en el mejor de los

casos, discutibles. Un tratamiento adecuado a partir de la relatividad general fue dado por primera vez por John Robert Oppenheimer y Hartland Snyder (1939).

9. De hecho, la localización *exacta* del horizonte, en el caso de un agujero negro general no estacionario, no es algo que pueda determinarse por medidas directas. ¡Depende en parte de un conocimiento de todo el material que caerá en el agujero negro en su *futuro*!
10. Ver las discusiones de Belinskii, Khalatnikov y Lifshitz (1970) y Penrose (1979*b*).
11. Resulta tentador identificar la contribución gravitatoria a la entropía de un sistema con alguna medida de la curvatura total de Weyl, pero todavía no se ha encontrado ninguna medida apropiada. (En general tendría que tener algunas complejas propiedades no-locales.) Afortunadamente no necesitamos esta medida de la entropía gravitatoria para nuestra discusión.
12. Existe una concepción muy popular en estos momentos, conocido como el «escenario inflacionario», que se propone explicar, entre otras cosas, el porqué de que el universo sea tan uniforme a gran escala. Según este punto de vista, el universo sufrió una enorme expansión en sus primeras etapas —de un orden enormemente mayor que la expansión «ordinaria» del modelo estándar. La idea es que cualquier irregularidad pudo ser reforzada por esta expansión. Sin embargo, sin alguna ligadura inicial aún mayor, como la que ya proporciona la hipótesis de curvatura de Weyl, la inflación no puede funcionar. No introduce ningún ingrediente con asimétrica temporal que pudiera explicar la diferencia entre las singularidades inicial y final. (Además está basado en teorías físicas no confirmadas —las teorías GUT— cuyo status no pasa de TENTATIVAS, en la terminología del Capítulo 5. Para una discusión crítica de la «inflación» en el contexto de las ideas de este capítulo, ver Penrose (1989*b*.)

Notas del Capítulo 8

1. Estas modificaciones populares son: (i) cambiar las ecuaciones reales de Einstein $\text{RICCI} = \text{ENERGY}$ (vía «lagrangianos de alto orden»); (ii) cambiar el número de dimensiones del espacio-tiempo de cuatro a algún número mayor (como en las llamadas «teorías de tipo Kaluza-Klein»); (iii) introducir «supersimetría» (una idea tomada en préstamo del comportamiento cuántico de bosones y fermiones, combinada en un amplio esquema y aplicada, no de forma completamente lógica, a las coordenadas del espacio-tiempo); (iv) teoría de cuerdas (un esquema radical muy popular en este momento en el que las «líneas-de-universo» se reemplazan por «historias-de-cuerdas» —combinada normalmente con las ideas de (ii) y (iii). Todas estas propuestas, por su popularidad y vigorosa presentación, entran firmemente en la categoría de TENTATIVAS en la terminología del Capítulo 5.

2. Por lo que puedo deducir, un punto de vista de este tipo está implícito en las propuestas actuales de Hawking para una explicación de estas cuestiones por medio de la gravitación cuántica (Hawking 1987, 1988). Una propuesta de Hartle y Hawking (1984) de un origen cuántico-gravitatorio para el estado inicial es posiblemente el tipo de cosa que *podría* dar substancia teórica a una condición inicial del tipo $WEYL = 0$, pero en estas ideas no está ni mucho menos ausente (en mi opinión) un input *esencial* con asimetría temporal.
3. Estos hechos son algo más transparentes en términos de la operación del *producto escalar* $\langle \psi | \chi \rangle$ dada en la nota 6 del Capítulo 6. En la descripción hacia delante en el tiempo calculamos la probabilidad p mediante

$$p = |\langle \psi | \chi \rangle|^2 = |\langle \chi | \psi \rangle|^2$$

y en la descripción hacia atrás, mediante

$$p = |\langle \chi' | \psi' \rangle|^2 = |\langle \psi' | \chi' \rangle|^2$$

El hecho de que estas sean iguales se sigue de $\langle \psi' | \chi' \rangle = \langle \psi | \chi \rangle$, que es lo que esencialmente entendemos por «evolución unitaria».

4. Algunos lectores pueden tener dificultad en comprender lo que puedo querer decir al preguntar cuál es la probabilidad de un suceso pasado dado un suceso futuro! Sin embargo no hay un problema esencial en esto. Imaginemos toda la historia del universo representada en el espacio-tiempo. Para encontrar la probabilidad de que ocurra p , dado que ha ocurrido q , imaginémonos examinando todas las ocurrencias de q y contando qué fracción de estas están acompañadas de p . Esta es la probabilidad buscada. No importa si q es el tipo de suceso que normalmente ocurriría más tarde o antes de p .
5. Debe estar permitido que existan los llamados *gravitones longitudinales* —los gravitones «virtuales» que componen un campo gravitatorio estático. Desgraciadamente hay problemas teóricos implicados al definir tales cosas de una forma matemática precisa e «invariante».
6. Mis propios burdos intentos originales para calcular este valor fueron muy mejorados por Abhay Ashtekar, y estoy utilizando aquí su valor (ver Penrose, 1987a). Sin embargo, él me ha enfatizado que hay bastante arbitrariedad en algunas de las hipótesis que parecemos obligados a utilizar, de modo que hay que tener mucha precaución al adoptar el valor exacto obtenido para la masa.
7. Otros diversos intentos para proporcionar una teoría objetiva de la reducción del vector de estado han aparecido de tarde en tarde en la literatura. Los más importantes son Károlyházy (1974), Károlyházy, Frenkel y Lukács (1986), Komar (1969), Pearle (1985, 1988), Ghirardi, Rimini y Weber (1986).
8. Yo mismo he estado interesado, durante años, en tratar de desarrollar una

teoría no-local del espacio-tiempo, motivada generalmente desde otras direcciones, denominada «teoría de twistor» (ver Penrose y Rindler 1986, Huggett y Tod 1985, Ward y Wells 1989). Sin embargo, a esta teoría aún le faltan, en el mejor de los casos, algunos ingredientes esenciales, y no sería apropiado entrar ahora en una discusión sobre ella.

Notas del Capítulo 9

1. En una emisión radiofónica de la BBC; ver Hodges (1983), p. 419.
2. Los primeros experimentos de este tipo se realizaron con gatos (cf. Myers y Sperry 1953). Para más información sobre experimentos de escisión cerebral, ver Sperry (1966), Gazzaniga (1970), MacKay (1987).
3. Para un informe fácil de leer sobre el funcionamiento del cortex visual, ver Huberl (1988).
4. Ver Hubel (1988) p. 221. Experimentos anteriores registraron células sensibles sólo a la imagen de una mano.
5. La teoría, ahora bien establecida, de que el sistema nervioso consiste en células individuales separadas, las neuronas, fue vigorosamente expuesta por el gran neuronatista español Ramón y Cajal alrededor de 1900.
6. De hecho, *todas* las puertas lógicas pueden construirse a partir de « $\sim$ » y « $\&$ » (o incluso sólo a partir de *una* operación $\sim(A\&B)$).
7. En realidad, el uso de puertas lógicas está más próximo a la construcción de un computador electrónico de lo que lo están las consideraciones detalladas de la máquina de Turing del Capítulo 2. En este capítulo se hacía énfasis en la aproximación de Turing por razones teóricas. El desarrollo de los computadores reales surgió tanto del trabajo del matemático húngaro-americano John von Neumann como del de Turing.
8. Estas comparaciones son equívocas en ciertos aspectos. La inmensa mayoría de los transistores en los computadores actuales conciernen a la «memoria» más que a la acción lógica, y la memoria del computador siempre puede ser aumentada por medios exteriores, indefinidamente en la práctica. Con la operación paralela aumentada, más transistores estarían involucrados directamente en la acción lógica de lo que sucede normalmente en la actualidad.
9. Deutsch prefiere utilizar, en sus descripciones, la interpretación de los muchos universos con respecto a la teoría cuántica. Sin embargo, es importante darse cuenta de que esto es completamente accesorio, y que el concepto de computador cuántico es igualmente apropiado cualquiera que sea el punto de vista que adoptemos con respecto a la mecánica cuántica estándar.

1. Ver Lucas (1961) para un argumento de que el teorema de Gödel implica no-computabilidad, y Good (1969), Benacerraf (1967), Bowie (1982) para varios contra-argumentos.
2. Algunos lectores pueden quedar turbados por el hecho de que haya diferentes puntos de vista entre los matemáticos. Recuérdesse la discusión en el Capítulo 4. Sin embargo, las diferencias, donde existen, no nos tienen que interesar aquí necesariamente: se refieren sólo a las cuestiones esotéricas relativas a conjuntos muy grandes, así que basta con que restrinjamos nuestra atención a proposiciones de la aritmética (con un número finito de cuantificadores existenciales y universales) y se aplicará la discusión precedente. (Quizá esto exagera algo la situación, puesto que un principio de reflexión relativo a conjuntos infinitos puede utilizarse a veces para derivar proposiciones de la aritmética.) En cuanto al formalista muy dogmático, inmune a Gödel, que afirma no reconocer siquiera que *exista* tal cosa como una verdad matemática, simplemente lo ignoro, ¡puesto que aparentemente no posee la cualidad de adivinación de la verdad que acabe la discusión!
3. El término «agujero negro» se hizo de uso común sólo mucho más tarde, hacia 1968 (básicamente por vía de las ideas proféticas del físico americano John A. Wheeler).
4. Me parece que el hecho de que los animales necesiten un período de sueño en el que parecen a veces *soñar* (como se nota a menudo en los perros) es una evidencia de que pueden poseer consciencia. En efecto, algo de consciencia parece ser un ingrediente importante en la distinción entre dormir soñando o no.
5. En el caso de la relatividad especial o general, léase «espacios simultáneos» o «superficies de tipo-espacio» pp. 260, 276) en lugar de «tiempos».
6. Existe, sin embargo, un obstáculo en el caso de un universo espacialmente infinito, ya que entonces (de modo análogo al caso de los muchos universos) resulta que ¡habría infinitas copias de uno mismo y del entorno inmediato de uno! El comportamiento futuro de cada copia podría ser ligeramente diferente, ¡y nunca estaríamos bastante seguros sobre cuál de las copias aproximadas de uno mismo modeladas en las matemáticas podría-mos «ser» realmente!
7. Incluso el crecimiento de ciertos cristales reales podría implicar problemas similares, por ejemplo, en los casos en que la celda básica unidad involucra varios cientos de átomos, como en las llamadas fases de Frank-Casper. Debería mencionarse, por el contrario, que un procedimiento de crecimiento teórico «casi local» (aunque todavía no local) para la simetría quíntuple ha sido sugerido por Onoda, Steinhardt y Socolar (1988).

Referencias

- Aharonov, Y. y Albert, D. Z. (1981): «Can we make sense out of the measurement process in relativistic quantum mechanics? *Phys. Rev.*, **D24**, 359-370.
- Aharonov, Y.; Bergmann, P. y Lebowitz, J. L. (1964): «Time symmetry in the quantum process of measurement». En *Quantum Theory and measurement* (ed. J. A. Wheeler y W. H. Zurek), Princeton University Press, 1983; originalmente aparecido en *Phys. Rev.*, **134B**, 1410-1416.
- Ashtekar, A.; Balachandran, A. P. y Sang Jo (1989): «The CP problem in quantum gravity». *Int. J. Mod. Phys.*, **A6**, 1493-1514.
- Aspect, A. y Grangier, P. (1986): «Experiments on Einstein-Podolsky-Rosen-type correlations with pairs of visible photons». En *Quantum concepts in space and time* (ed. R. Penrose y C. J. Isham), Oxford University Press.
- Atkins, P. W. (1987): *Why mathematics works*. Oxford University Extension Lecture in series: Philosophy and the New Physics (13 marzo).
- Barrow, J. D. (1988): *The world within the world*. Oxford University Press.
- Barrow, J. D. y Tipler, F. J. (1986): *The anthropic cosmological principle*. Oxford University Press.
- Baylor, D. A.; Lamb, T. D. y Yau, K. W. (1979): «Responses of retinal rods to single photons». *J. Physiol.*, **288**, 613-634.
- Bekenstein, J. (1972): «Black holes and entropy». *Phys. Rev.*, **D7**, 2333-2346.
- Belifante, F. J. (1975): *Measurement and time reversal in objective quantum theory*. Pergamon Press, New York.
- Belinskii, V. A.; Khalatnikov, I. M. y Lifshitz, E. M. (1979): «Oscillatory approach to a singular point in the relativistic cosmology». *Adv. Phys.*, **19**, 525-573.
- Bell, J. S. (1987): *Speakable and unspeakable in quantum mechanics*. Cambridge University Press. (Hay traducción española: *Lo decible y lo indecible en mecánica cuántica*. Alianza Universidad, Madrid 1991.)
- Benacerraf, P. (1967): «God, the Devil and Gödel». *The Monist*, **51**, 9-32.
- Blakemore, C. y Greenfield, S. (ed) (1987): *Mindwaves: thoughts on intelligence, identity and consciousness*. Basil Blackwell, Oxford.

- Blum, L.; Shub, M. y Smale, S. (1989): «On a theory of computation and complexity over the real numbers: NP completeness, recursive functions and universal machines. *Bull. Amer. Math. Soc. (New series)*, **21**, n.º 1, pp. 1-46.
- Bohm, D. (1951): «The Paradox of Einstein, Rosen and Posolsky». En *Quantum theory and measurement* (ed. J. A. Wheeler y W. H. Zurek), Princeton University Press, 1983; originalmente aparecido en el *Quantum Theory*, D. Bohm, Ch. 22, sect. 15-19. Prentice-Hall, Englewoods-Clifts.
- Bohm, D. (1952): «A suggested interpretation of the quantum theory in terms of "hidden" variables, I and II», en *Quantum theory and measurement*, (ed., J. A. Wheeler y W. H. Zurek), Princeton University Press, 1983; originalmente aparecido en *Phys. Rev.*, **85**, 166-193.
- Bondi, H. (1960): «Gravitational waves in general relativity». *Nature (London)*, **186**, 535.
- Bowie, G. L. (1982): «Lucas' number is finally up». *J. of Philosophical Logic*, **11**, 279-285.
- Cartan, E. (1923): «Sur les variétés a connexion affine et la théorie de la relativité generalisée. *Ann. Sci. Ec. Norm. Sup.*, **40**, 325-412.
- Chandrasekhar, S. (1987): *Truth and beauty: aesthetics and motivations in science*, University of Chicago Press.
- Church, A. (1941): *The calculi of lambda-conversion*. Annals of Mathematical Studies, n.º 6. Princeton University Press.
- Churchland, P. M. (1984): *Matter and consciousness*. Bradford Books, MIT Press, Cambridge, Mass.
- Clauser, J. F.; Horne, A. H.; Shymony, A. y Holt, R. A. (1969): «Proposed experiment to test local hidden-variable theories». En *Quantum theory and measurement* (ed. J. A. Wheeler y W. H. Zurek), Princeton University Press, 1983; originalmente aparecido en *Phys. Rev. Lett.*, **23**, 880-884.
- Close, F. (1983): *The cosmic onion: quarks and the nature of the universe*. Heinemann, London. (Hay traducción española: *La cebolla cósmica*, Editorial Crítica, Barcelona, 1988).
- Cohen, P. C. (1966): *Set theory and the continuum hypothesis*. Benjamin, Menlo Park, CA.
- Cutland, N. J. (1980): *Computability: an introduction to recursive function theory*. Cambridge University Press.
- Davies, P. C. W. (1974): *The physics of time-asymmetry*. Surrey University Press.
- Davies, P. C. W. y Brown, J. (1988): «Superstrings: a theory of everything?». Cambridge University Press. (Hay traducción española: *Supercuerdas ¿una teoría de todo?* Alianza Editorial, Madrid, 1990).
- Davies, R. D.; Lasenby, A. N.; Watson, R. A.; Daintree, E. J.; Hopkins, J.; Beckman, J.; Sánchez-Almeida, J. y Rebol, R. (1987): «Sensitive measurement of fluctuations in the cosmic microwave background». *Nature*, **326**, 462-465.
- Davis, M. (1988): «Mathematical logic and the origin of modern computers».

- En *The universal Turing machine: a half-century survey* (ed. R. Herken), Kammerer & Unverzagt, Hamburg.
- Dawkins, R. (1986): *The blind watchmaker*. Longman, London. (Hay traducción española: *El relojero ciego*, Editorial Labor, Barcelona, 1989).
- de Broglie, L. (1956): *Tentative d'interpretation causale et nonlineaire de la mecanique ondulatoire*. Gauthier-Villars, París.
- Deeke, L.; Grötzinger, B. y Kornhuber, H. H. (1976): «Voluntary finger movements in man: cerebral potentials and theory. *Biol. Cybernetics*, **23**, 99.
- Delbrück, M. (1986): *Mind from Matter?* Blackwell Scientific Publishing, Oxford.
- Dennet, D. C. (1978): *Brainstorms*. Philosophical Essays on Mind and Psychology, Harvester Press, Hassocks, Sussex.
- Deutsch, D. (1985): «Quantum theory, the Church-Turing principle and the universal quantum computer». *Proc. Roy. Soc. (Lond.)*, **A400**, 97-117.
- Devlin, K. (1988): *Mathematics: the new golden age*. Penguin Books, London.
- De Witt, B. S. y Graham, R. D. (ed.) (1973): *The many-worlds interpretation of quantum mechanics*. Princeton University Press.
- Dirac, P. A. M. (1928): «The quantum theory of the electron». *Proc. Roy. Soc. (Lond.)*, **A117**, 610-624; *ditto*, part II, *ibid.*, **A118**, 361.
- (1938): «Classical theory of radiating electrons». *Proc. Roy. Soc. (Lond.)*, **A167**, 148.
- (1939): «The relations between mathematics and physics» *Proc. Roy. Soc., Edinburgh*, **59**, 122.
- (1947): *The principles of quantum mechanics* (3.^a ed.), Oxford University Press. (Hay traducción española: *Principios de Mecánica Cuántica*, Editorial Ariel, Barcelona, 1958).
- (1982): «Pretty mathematics». *Intr. J. Theor. Phys.*, **21**, 603-605.
- Drake, S. (trans.) (1953): *Galileo Galilei: dialogue concerning the two chief world systems-Ptolemaic and Copernican*. University of California, Berkeley, 1973. (Hay traducción española de los diálogos de Galileo: *Diálogo sobre los sistemas máximos*, traducción, prólogo y notas de José Manuel Revuelta, Aguilar, Buenos Aires, 1975).
- (1957): *Discoveries and opinions of Galileo*. Doubleday, New York.
- Eccles, J. C. (1973): *The understanding of the brain*. McGraw-Hill, New York.
- Einstein, A., Podolsky, B. y Rosen, N. (1935): «Can quantum-mechanical description of physical reality be considered complete En *Quantum mechanics and measurement* (ed. J. A. Wheeler y W. H. Zurek), Princeton University Press, 1983; originalmente aparecido en *Phys. Rev.*, **47**, 777-780.
- Everett, H. (1957): «“Relative State” formulation of quantum mechanics». En *Quantum theory and measurement* (ed. J. A. Wheeler y W. H. Zurek), Princeton University Press, 1983; originalmente aparecido en *Rev. of Mod. Phys.*, **29**, 454-462.
- Feferman, S. (1988): «Turing in the Land of O(z)». En *The universal Turing*

- machine: a half-century survey* (ed. R. Herken). Kammerer & Unverzagt, Hamburg.
- Feynman, R. P.; Leighton y Sands (1965): *The Feynman Lectures*. Addison Wesley. (Hay edición bilingüe inglés-español en Fondo Educativo Interamericano, Panamá, 1972).
- (1985): *QED: the strange theory of light and matter*. Princeton University Press. (Hay traducción española: *Electrodinámica cuántica. La extraña teoría de la luz y la materia*. Alianza Editorial, 1988).
- Fodor, J. A. (1983): *The modularity of mind*. MIT Press, Cambridge, Mass.
- Fredkin, E. y Toffoli, T. (1982): «Conservatory Logic». *Int. J. Theor. Phys.*, **21**, 219-253.
- Freedman, S. J. y Clauser, J. F. (1972): «Experimental test of local hidden-variable theories». En *Quantum theory and measurement*. (ed. J. A. Wheeler y W. H. Zurek) Princeton University Press, 1983; originalmente aparecido en *Phys. Rev. Lett.*, **28**, 938-941.
- Galilei, G. (1638): *Dialogues concerning two new sciences*. Macmillan ed. 1914; Dover Inc. (Hay traducción española: *Consideraciones y demostraciones matemáticas sobre dos nuevas ciencias*. Edición preparada por C. Solís y J. Sádaba. Editora Nacional, Madrid, 1976).
- Gandy, R. (1988): «The Confluence of Ideas in 1936». En *The universal Turing machine: a half-century survey* (ed. R. Herken), Kammerer & Unverzagt, Hamburg.
- Gardner, M. (1958): *Logic machines and diagrams*. University of Chicago Press. (Hay traducción española: *Máquinas y diagramas lógicos*, Alianza Editorial, Madrid, 1985).
- (1983): *The whys of a philosophical scrivener*. William Morrow and Co., Inc., New York. (Hay traducción española: *Los porqués de un escriba filósofo*, Editorial Tusquets, Barcelona, 1989).
- (1989): *Penrose tiles to trapdoor ciphers*. W. H. Freeman and Company, New York.
- Gayle, F. W. (1987): «Free-surface solidification habit and point group symmetry of a faceted icosahedral Al-Li-Cu phase». *J. Mater. Res.*, **2**.
- Gazzaniga, M. S. (1970); *The bisected brain*. Appleton-Century-Crofts, New York.
- Gazzaniga, M. S.; LeDoux, J. E. y Wilson, D. H. (1977): «Language, praxis, and the right hemisphere: clues to some mechanism of consciousness». *Neurology*, **27**, 1144-1147.
- Geroch, R. y Hartle, J. B. (1986): «Computability and physical theories. *Found. Phys.*, **16**, 533.
- Ghirardi, G. C., Rimini, A. y Weber, T. (1980): «A general argument against superluminal transmission through the quantum mechanical measurement process. *Lett. Nuovo Cim.*, **27**, 293-298.
- (1986): «Unified dynamics for microscopic and macroscopic systems». *Phys. Rev.*, **D34**, 470.
- Gödel, K. (1931): «Über formal unentscheidbare Sätze der Principia Mathe-

- matica und verwandter Systema I». *Monatshefte für Mathematik und Physik*, **38**, 173-198. (Hay traducción española en Cuadernos Teorema, Universidad de Valencia, 1981; y también en: *Obras completas*. K. Gödel. Alianza Editorial, Madrid, 1987).
- Good, I. J. (1969): «Gödel's theorem is a red herring». *Brit. J. Philos. Sci.*, **18**, 359-373.
- Gregory, R. L. (1981): *Mind in science: A history of explanations in psychology and physics*, Weidenfeld and Nicholson Ltd.
- Grey Walter, W. (1953): *The living brain*. Gerald Duckworth and Co. Ltd.
- Grünbaum, B. y Shephard, G. C. (1981): «Some problems on plane tilings». En *The mathematical Gardner*. (ed. D. A. Klarner), Prindle, Weber And Schmidt, Boston.
- (1987): *Tilings and patterns*. W. H. Freeman.
- Hadamard, J. (1945): *The psychology of invention in the mathematical field*. Princeton University Press. [Reeditado en 1954 en Dover Publishing Ltd. El informe completo de Poincaré «La invención matemática» puede encontrarse en *Ciencia y método*, Espasa Calpe, Madrid, 1968 (Nota del traductor)].
- Hanf, W. (1974): «Nonrecursive tilings of the plane, I. *J. Symbolic Logic*, **39**, 283-285.
- Harth, E. (1982): *Windows of the Mind*. Harvester Press, Hassoks, Sussex.
- Hartle, J. B. y Hawking, S. W. (1983): «Wave function of the universe». *Phys. Rev.*, **D31**, 1777.
- Hawking, S. W. (1975): «Particle creation by black holes». *Commun. Math. Phys.*, **43**, 199-220.
- (1987): «Quantum cosmology». En *300 years of gravitation* (ed. S. W. Hawking y W. Israel), Cambridge University Press.
- (1988): *A brief history of time*. Bantam Press. London. (Hay traducción española: *Historia del tiempo*, Editorial Crítica, Barcelona, 1988; y Alianza Editorial, Madrid, 1990.)
- Hawking, S. W. y Penrose, R. (1970): «The singularities of gravitational collapse and cosmology». *Proc. Roy. Soc. (London)*, **A134**, 529-48.
- Hebb, D. O. (1954): «The problem of consciousness and introspection». En *Brain mechanism and consciousness* (ed. J. F. Delafresnaye), Blackwell, Oxford.
- Hecht, S., Shlaer, S. y Pirenne, M. H. (1941): «Energy, quanta and vision». *J. of Gen. Physiol.*, **25**, 891-40.
- Herken, R., ed. (1988): *The universal Turing machine: a half-century survey*, Kammerer & Unverzagt, Hamburg.
- Hiley, B. J. y Peat, F. D. eds. (1987): *Quantum implications. Essays in honour of David Bohm*. Routledge and Kegan Paul, London & New York.
- Hodges, A. P. (1983): *Alan Turing: the enigma*. Burnett Books and Hutchinson, London; Simon y Schuster, New York.
- Hofstadter, D. R. (1979): *Gödel, Escher, Bach: an eternal golden braid*. Har-

- vester Press, Hassoks, Sussex. (Hay traducción española: *Gödel, Escher, Bach*, Editorial Tusquets, Barcelona 1989.)
- (1981): «A conversation with Einstein's brain. En *The mind's I* (ed. D. R. Hofstadter y D. C. Dennett), Basic Books, inc.; Penguin Books, Ltd; Harmondsworth, Middx.
- Hofstadter, D. R. y Dennett, D. C., ed. (1981): *The mind's I*, Basic Books, Inc.; Penguin Books, Ltd; Harmondsworth, Middx.
- Hubel, D. H. (1988): *Eye, brain and vision*. Scientific American Library Series 22.
- Huggett, S. A. y Tod, K. P. (1985): *An introduction to twistor theory*. London Math. Soc. student texts, Cambridge University Press.
- Jaynes, J. (1980): *The origin of consciousness in the breakdown of the bicameral mind*. Penguin Books Ltd. Harmondsworth, Middx.
- Kandel, E. R. (1976): *The cellular basis of behaviour*. Freeman, San Francisco.
- Károlyházy, F. (1974): «Gravitation and quantum mechanics of macroscopic bodies. *Magyar Fizikai Polyoirat*, 12, 24.
- Károlyházy, F., Frenkel, A. y Lukács, B. (1986): «On the possible role of gravity on the reduction of the wave function». En *Quantum concepts in space and time* (ed. R. Penrose y C. J. Isham), Oxford University Press.
- Keene, R. (1988): «Chess: Henceforward». *The Spectator*, 261 (n.º 8.371), 52
- Knuth, D. M. (1981): *The art of computer programming*, Vol. 2 (2nd edn), Addison-Wesley, Reading, MA.
- Komar, A. B. (1964): «Undecidability of macroscopically distinguishable states in quantum field theory». *Phys. Rev.*, 133B, 542-4.
- Kuznetsov, B. G. (1977): *Einstein: Leben, Tod, Unslerblichkeit* (traducido al alemán por H. Fuchs). Birkhauser, Basel.
- Le Doux, J. E. (1985): «Brain, mind and language». En *Brain and mind* (ed. D. A. Oakley), Methuen, London y New York.
- Levy, D. W. L. (1984): *Chess computer handbook*, Batsford.
- Libet, B., Wright, E. W. Jr., Feinstein, B. y Pearl, D. K. (1979): «Subjective referral of the timing for a conscious sensory experience». *Brain*, 102, 193-224.
- Lorenz, K. (1972): Citado en: *From ape to Adam* por H. Wendt, Bobbs Merrill, Indianapolis.
- Lucas, J. R. (1961): «Minds, machines and Gödel». *Philosophy*, 36, 120-4; reimpresso en Alan Ross Anderson (1964): *Minds and machines*, Englewood Cliffs.
- MacKay, D. (1987): «Divided brains-divided minds?» En *Mindwaves* (ed. C. Blakemore y S. Greenfield), Basil Blackwell, Oxford.
- Majorana, E. (1932): «Atomi orientati in campo magnetico variabile». *Nuovo Cimento*, 9, 43-50.
- Mandelbrot, B. B. (1986): «Fractals and the rebirth of iteration theory». En *The beauty of fractals: images of complex dynamical systems*, H. O.). Peitgen y P.-H. Richter, Springer-Verlag, Berlín, pp. 151-60.

- Maxwell, J. C. (1865): «A dynamical theory of the electromagnetic field». *Philos. Trans. Roy. Soc. (Lond.)*, **155**, 459-512.
- Mermin, D. (1985): «Is the moon there when nobody looks? Reality and the quantum theory. *Physics Today*, **38** (n.º 4), 38-47.
- Michie, D. (1988): «The fifth generation's unbridged gap». En *The universal Turing machine: a half-century survey* (ed. R. Herken), Kammerer & Unverzagt, Hamburg.
- Minsky, M. L. (1968): «Matter, mind, and models». En *Semantic information processing* (ed. M. L. Minsky), M.I.T. Press, Cambridge, Mass.
- Misner, C. W. (1969): «Mixmaster universe. *Phys. Rev. Lett.*, **22**, 1071-4.
- Morouzzii, G. y Magoun, H. W. (1949): «Brainstem reticular formation and activation of the EEG. *Electroencephalography and Clinical Neurophysiology*, **1**, 455-73.
- Mott, N. F. (1929): «The wage mechanics of the α -tracks». En *Quantum theory and measurement* (ed. J. A. Wheeler y W. H. Zurek), Princeton University Press, 1983; originalmente aparecido en *Proc. Roy. Soc. (Lond.)*, **A126**, 79-84.
- Mott, N. F. y Massey, H. S. W. (1965): «Magnetic moment of the electron». En *Quantum theory and measurement* (ed. J. A. Wheeler y W. H. Zurek), Princeton University Press, 1983; originalmente aparecido en *The theory of atomic collisions* de N. F. Mott y H. S. W. Massey (Clarendon Press, Oxford, 1965).
- Myers, D. (1974): «Nonrecursive tilings of the plane, II». *J. Symbolic Logic*, **49**, 286-94.
- Myers, R. E. y Sperry, R. W. (1953): «Interocular transfer of a visual form discrimination habit in cats after the section of the optic chiasm and corpus callosum». *Anatomical Record*, **175**, 351-2.
- Nagel, E. y Newmann, J. R. (1958): *Gödel's proof*. Routledge and Keegan Paul Ltd. (Hay traducción española: *El teorema de Gödel*, Editorial Tecnos, Madrid, 1979).
- Nelson, D. R. y Halperin, B. I. (1985): «Pentagonal and icosahedral order in rapidly cooled metals». *Science*, **229**, 233.
- Newton, I. (1687): *Principia*. Cambridge University Press. (Hay dos ediciones españolas, la primera preparada por A. Escotado, Editora Nacional, 1980, y Tecnos, Madrid, 1988; y la segunda por E. Rada, Alianza Editorial, Madrid, 1988).
- (1730): *Optiks*, 1952 Dover Inc. (Hay edición española preparada por C. Solís, Alfaguara, Madrid, 1970).
- Oakley, D. A. (ed.) (1985): *Brain and mind*, Methuen, London and New York.
- Oakley, D. A. y Eames, L. C. (1985): «The plurality of consciousness». En *Brain and mind* (ed. D. A. Oakley), Methuen, London y New York.
- O'Connell, K. (1988): «Computer chess». *Chess*, **15**.
- O'Keefe, J. (1985): «Is consciousness the gateway to the hippocampal cogni-

- tive map? A speculative essay on the neural basis of mind». En *Brain and mind* (ed. D. A. Oakley), Methuen, London y New York.
- Onoda, G. Y., Steinhardt, P. J., DiVincenzo, D. P. y Socolar, J. E. S. (1988): «Growing perfect quasicrystals». *Phys. Rev. Lett.*, **60**, 2688.
- Oppenheimer, J. R. y Snyder, H. (1939): «On continued gravitational contraction». *Phys. Rev.* **56**, 455-9.
- Pais, A. (1982): «*Subtle is the Lord...*»: *the science and life of Albert Einstein*. Clarendon Press, Oxford. (Hay traducción española: *El Señor es sutil. La ciencia y la vida de Albert Einstein*, Editorial Ariel, Barcelona, 1984). [La carta de Einstein citada en p. 353 figura en *Correspondencia Einstein-Born*, Siglo XXI, México, 1973 (Nota del traductor)]
- Paris, J. y Harrington, L. (1977): «A mathematical incompleteness in Peano arithmetic». En *Handbook of mathematical logic* (ed. J. Barwise), North-Holland, Amsterdam.
- Pearle, P. (1985): «Models for reduction». En *Quantum concepts in space and time* (ed. C. J. Isham y R. Penrose), Oxford University Press.
- (1989): «Combining stochastic dynamical state-vector reduction with spontaneous localization». *Phys. Rev. A*, **39**, 2277-89.
- Peitgen, H.-O. y Richter, P. H. (1986): *The beauty of fractals*, Springer-Verlag, Berlin and Heidelberg.
- Peitgen, H.-O. y Saupe, D. (1988): *The science of fractals images*, Springer-Verlag, Berlín.
- Penfield, W. y Hasper, H. (1947): *Highest Level Seizures*, Research Publications of the Association for Research in Nervous and Mental Diseases (New York) **26**, 252-71.
- Penrose, R. (1965): «Gravitational collapse and space-time singularities». *Phys. Rev. Lett.*, **14**, 57-9.
- (1974): «The role of aesthetics in pure and applied mathematical research». *Bull. Inst. Math. Applications* **10**, n.º 7/8, 266-71.
- (1979a): «Einstein's vision and the mathematics of the natural world». *The Sciences* (Marzo) 6-9.
- (1979b): «Singularities and time-asymmetry». En *General relativity: an Einstein centenary* (ed. S. W. Hawking y W. Israel), Cambridge University Press.
- (1987a): «Newton, quantum theory and reality». En *300 years of gravity* (ed. S. W. Hawking y W. Israel), Cambridge University Press.
- (1987b): *Quantum implications. Essays in honour of David Bohm*. Routledge y Keegan Paul, London & New York.
- (1989a): «Tilings and quasicrystals: a non-local growth problem?» En *Aperiodicity and order 2* (ed. M. Jaric), Academic Press, New York.
- (1989b): «Difficulties with inflationary cosmology», en *Proceeding of the 14th Texas Symposium on Relativistic Astrophysics* (ed. E. Fenes), NY Acad. Sci., New York.
- Penrose, R. y Rindler, W. (1984): *Spinors and space-time*. Vol. 1: *Two-spinor calculus and relativistic fields*, Cambridge University Press.

- (1986): *Spinors and space-time*. Vol. 2: *Spinor and twistor methods in space-time geometry*, Cambridge University Press.
- Pour-El M. B. y Richards, I. (1979): «A computable ordinary differential equation which possesses no computable solution». *Ann. Math. Logic*, **17**, 61-90.
- (1981): «The wave equation with computable initial data such that its unique solution is not computable. *Adv. in Math.*, **39**, 215-39.
- Pour-El, M. B. y Richards, I. (1982): «Noncomputability in models of physical phenomena». *Int. J. Theor. Phys.*, **21**, 553-5.
- Rae, A. (1986): *Quantum physics: illusion or reality?* Cambridge University Press. (Hay traducción española: *Física Cuántica: ¿ilusión o realidad?*, Alianza Editorial, Madrid, 1988).
- Resnikoff, H. L. y Wells, R. O. Jr. (1973): *Mathematics and civilization*, Holt, Rinehart y Winston, Inc., New York, reimpresso con adiciones 1984 Dover Publications, Inc. Mineola, NY.
- Rindler, W. (1977): *Essential relativity*, Springer-Verlag, New York.
- (1982): *Introduction to special relativity*, Clarendon Press, Oxford.
- Robinson, R. M. (1971): «Undecidability and nonperiodicity for tilings of the plane». *Invent. Math.*, **12**, 177-209.
- Rouse Ball, W. W. (1892): «Calculating prodigies». En *Mathematical recreations and essays*.
- Rucker, R. (1984): *Infinity and the mind: the science and philosophy of the infinite*. Paladin Books, Granada Publishing Ltd., London (publicado inicialmente por Harvester Press Ltd, 1982).
- Sachs, R. K. (1962): «Gravitational waves in general relativity. VIII. Waves in Asymptotically flat space-time». *Proc. Roy. Soc. London*, **A270**, 103-126.
- Schank, R. C. y Abelson, R. P. (1977): *Scripts, plants, goals and understanding*. Erlbaum, Hillsdale, NJ.
- Shechtman, D., Blech, I., Gratias, D. y Cahn, J. W. (1984): «Metallic phase with long-range orientational order and no translational symmetry. *Phys. Rev. Lett.*, **53**, 1951.
- Schrödinger, E. (1935): «Die gegenwärtige Situation in der Quantenmechanik», *Naturwissenschaften*, **23**, 807-812, 823-828, 844-849. (Traducido al inglés por J. T. Trimmer (1980). En *Proc. Amer. Phil. Soc.*, **124**, 323-338.) En *Quantum theory and measurement* (ed. J. A. Wheeler y W. H. Zurek), Princeton University Press, 1983.
- Schrödinger, E. (1967): «*What is life*» and *Mind and Matter*, Cambridge University Press. (Hay traducción española: *¿Qué es la vida?*, Editorial Tusquets, Barcelona, 1984, y *Orbis*, Barcelona, 1986; *Mente y materia*, Editorial Tusquets, Barcelona, 1985. También hay traducción catalana en Edicions 62, Barcelona, 1984).
- Searle, J. (1980): «Mind, brains and programs», en *The behavioral and brain sciences*, vol. 3. Cambridge University Press, reimpresso en *The mind's I* (ed. D. R. Hofstadter y D. C. Dennet), Basic Books, Inc., Penguin Books Ltd., Harmondsworth, Middx., 1981.

- Searle, J. R. (1987): «Minds and brains without programs». En *Mindwaves* (ed. C. Blakemore y S. Greenfield), Basil Blackwell, Oxford.
- Smith, S. B. (1983): *The great mental calculators*. Columbia University Press.
- Smorynski, C. (1983): «“Big” news from Archimedes to Friedman». *Notices Amer. Math. Soc.*, **30**, 251-256.
- Sperry, R. W. (1986): «Brain bisection and consciousness». En *Brain and conscious experience* (ed. J. Eccles), Springer, New York.
- Squires, E. (1985): *To acknowledge the wonder*, Adam Hilger Ltd., Bristol.
- (1986): *The mystery of the quantum world*, Adam Hilger Ltd., Bristol.
- Tipler, F. J.; Clarke, C. J. S. y Ellis, G. F. R. (1980): «Singularities and horizons —a review article». En *General relativity and gravitation* (ed. A. Held), Vol. 2, pp. 97-206. Plenum Press, New York.
- Turing, A. M. (1937): «On computable numbers, with an application to the Entscheidungsprobleme». *Proc. Lond. Math. Soc. (ser. 2)*, **42**, 230-265; corrección en **43**, 544-546.
- (1939): «Systems of logic based on ordinals». *Proc. Lond. Math. Soc.*, **45**, 161-228.
- (1950): «Computing machinery and intelligence». *Mind* **59**, n.º 236; reimpresso en *The mind's I* (ed. R. D. Hofstadter y D. C. Dennett), Basic Books, Inc.; Penguin Books, Ltd; Harmondsworth, Middx., 1981. (Hay traducción española: «Maquinaria de cómputo e inteligencia», en *Perspectivas de la revolución de los computadores*, selección de Z. W. Pylyshyn, Alianza Editorial, Madrid, 1975).
- von Neumann, J. (1955): *Mathematical foundations of quantum mechanics*. Princeton University Press. (Hay traducción española: *Fundamentos matemáticos de la mecánica cuántica*, Instituto Jorge Juan, C.S.I.C., Madrid, 1955).
- Waltz, D. L. (1982): «Artificial intelligence». *Scientific American*, **247** (4), 101-122. (Traducción española en *Investigación y Ciencia*, diciembre 1982, 48-61).
- Ward, R. S. y Wells, R. O. Jr. (1989): *Twistor geometry*, Cambridge University Press.
- Weinberg, S. (1977): *The first three minutes: A modern view of the origin of the universe*. Andre Deutsch, London. (Hay traducción española: *Los tres primeros minutos del Universo*, Alianza Editorial, Madrid, 1978).
- Weiskrantz, L. (1987): «Neuropsychology and the nature of consciousness». En *Mindwaves* (ed. C. Blakemore and S. Greenfield), Blackwell, Oxford.
- Westfall, R. S. (1980): *Never at rest*, Cambridge University Press.
- Wheeler, J. A. (1983): «Law without law». En *Quantum theory and measurement* (ed. J. A. Wheeler y W. H. Zurek), Princeton University Press, pp. 182-213.
- Wheeler, J. A. y Feynman, R. P. (1945): «Interaction with the absorber as the mechanism of radiation». *Revs. Mod. Phys.*, **17**, 157-181.
- Wheeler, J. A. y Zurek, W. H. (ed.) (1983): *Quantum theory and measurement*. Princeton University Press.

- Whittaker, E. T. (1910): *The history of the theories of aether and electricity*, Longman, London. [Reeditado (1989), junto con la segunda parte, en Dover Publishing, Inc. (Nota del traductor)].
- Wigner, E. P. (1960): «The unreasonable effectiveness of mathematics. *Commun. Pure Appl. Math.*, **13**, 1-14.
- (1961): «Remarks on the mind-body question». En *The scientist speculates* (ed. I. J. Good), Heinemann, London. (Reimpreso en E.-Wigner (1967). *Symmetries and reflections*, Indiana University Press, Bloomington). En *Quantum theory and measurement* (ed. J. A. Wheeler y W. H. Zurek), Princeton University Press, 1983.
- Will, C. M. (1987): «Experimental gravitation from Newton's *Principia* to Einstein's general relativity». En *300 years of gravitation* (ed. S. W. Hawking y W. Israel), Cambridge University Press.
- Wilson, D. H.; Reeves, A. G.; Gazzaniga, M. S. y Culver, C. (1977): «Cerebral commissurotomy for the control of intractable seizures», *Neurology*, **27**, 708-715.
- Winograd, T. (1972): «Understanding natural language». *Cognitive Psychology*, **3**, 1-191.
- Wooters, W. K. y Zurek, W. H. (1982): «A single quantum cannot be cloned». *Nature*, **299**, 802-803.

Índice Analítico

Nota: Las entradas que se refieren a las notas se indican mediante el sufijo «n»; las Figuras y las Tablas se indican mediante *números en itálica*.

- abstracción (operación), 101, 102
- actividad cerebral, aspectos mecano-cuánticos de, 495, 497
- agrupamiento gravitacional, 402-403, 423
- agujeros blancos, 417, 420, 439, 452-453
- agujeros negros, 413-420
 - de Oppenheimer-Snyder, 574n, 520
 - entropía, 426-427
 - horizonte de sucesos de, 418-419
- Aitken, Alexander, 28
- ajedrez, computadores que juegan, 35
- álgebra, origen de la palabra, 57
- algoritmo
 - ¿cómo superar a uno?, 98-100
 - de Schank, 41-42
 - de Euclides, 57-60
 - máquina de Turing para el, 69, 71, 74
 - selección natural de, 513-516
 - significado del término, 40, 56-61
- al-Khowarizm, Abu Ja'far Mohammed ibn Musá, 57
- Ammann, Robert, 540
- amplitudes de probabilidad, 300-308, 374
- ángulo de inclinación, células sensibles a, 480
- antipartículas, 365
- Apolonio, 211
- área de Broca, 468, 476
- área de Wernicke, 468-469, 470
- Argand, Jean-Robert, 129
- Arquímedes, 210
- Aspect, Alain, 360
- autoconsciencia, 508
- Bach, Johann Sebastian
- Bekenstein-Hawking, fórmula de, 426, 429
- Bell, teorema de, 356
- Berger, Robert, 182
- big bang, 404
 - naturaleza especial de, 425-437
 - teoría del, 202, 404-408, 409
 - y segunda ley de la termodinámica, 411-412
- big crunch, 407
- Bohm, David, 354, 356
- Bohr, Niels, 293, 354
- bola de fuego primordial, 408-411
- Boltzmann, constante de, 392, 393
- Bose, S. N., 572n
- Bose-Einstein, estadística de, 572n
- bosones, 334-335, 572n
- Broca, área de, 468, 476
- Brouwer, L. E. J., 154, 156
- caja de Hawking, 447-451
- cálculo lambda, 100-106
- cámara de niebla, observaciones en, 458
- campo vectorial, 201, 231
- Cantor, Georg, 95, 117, 118, 119, 137

- corte diagonal de, 95, 119
 teoría de números transfinitos, 117-119
 caos, 226, 237, 371
 Cardano, Girolamo, 133-134
 cardioide, 185, 185
 Carter, Brandon, 537
 causalidad relativista, 273-278
 ceguera cortical, 479-480
 ceguera negada, 479n
 censura cósmica, 278, 419n
 cerebelo, 463, 466-471, 472
 cerebro humano, 463-501
 modelos de computador, 486-491
 estructura de, 463-472
 cerebro (como parte del encéfalo), 463-464
 circuito hamiltoniano, 191-192, 191
 Colby, programa de ordenador de, 34-35
 combustibles fósiles, entropía de, 402
 composición musical, 551-552
 compuertas de potasio, 485, 485
 compuertas de sodio, 485, 485
 computabilidad, 80, 100, 106
 en el mundo de bolas de billar, 221-226
 en física clásica, 278-279
 en los objetos físicos, 194
 y ecuación de ondas, 243-244
 computadores cuánticos, 497-499
 computadores paralelos, 77-78, 493
 condiciones de contorno, 437
 conexión de neuronas, computador a base de, 486-490
 conjetura de Goldbach, 92
 conjuntos
 complementarios, 163
 conjunto de Mandelbrot, 12, 107-113
 construcción de, 129-131
 descubrimiento inicial, 132
 posible no-recursividad de, 168-174, 183-185, 566n
 enumerablemente recursivos, 161-168
 representación esquemática, 167
 no recursivos, propiedad de los, 168
 números definidos en términos de, 137-138
 recursivos, 163
 ¿es el conjunto de Mandelbrot uno de ellos?, 168-174
 representación esquemática, 166
 teoría de, 161
 conocimiento, 503, 504
 consciencia
 significado del término, 502-506, 508
 propósito de, 505-507
 bifurcación de, 53-54, 478
 localización de la sede, 472-476
 «unicidad» de, 493-495
 retardo temporal de, 544-548
 consciencia animal, 527-529
 constante cosmológica, 407-408, 408n
 constante de Planck, 292
 corteza cerebral, 463-468
 córtex
 auditivo, 465, 467
 motor, 466, 467
 olfativo, 465, 466
 somatosensorial, 466, 467, 546
 visual, 464, 464, 479
 procesamiento de la información en, 480-481
 cromodinámica cuántica, 201
 cuasicristales, 540, 541, 542
 cuerpo caloso, 471, 476
 curare, efectos del, 506
 curvas psi, 310-311
 Chandrasekhar, límite de, 414, 415
 Church, Alonzo, 79, 101
 cálculo lambda de, 100-106
 Church-Turing, tesis de, 76-79
 Dase, Johann Martin Zacharias, 28
 de Broglie, príncipe Louis, 293, 376
 propuesta de dualidad onda-partícula de, 293-294
 de Broglie-Bohm, modelo de, 355
 décimo problema de Hilbert, 61
 insolubilidad de, 91-98
 desorden manifiesto, 385, 388
 determinismo, 534-536
 relatividad general, 277
 relatividad especial, 276-277
 determinismo fuerte, 536
 Deutsch, David, 194
 Dicke, Robert, 537
 dinámica, 199, 211-218
 dinámica newtoniana, 216-217, 568n
 mundo mecanicista de la, 218-221
 Dirac, Paul A. M., 201, 294, 376, 522
 572n

- ecuación de Dirac para el electrón, 364-365, 522
- disco unidad, 170-171
- distribución maxwelliana, 389-390
- doble-rendija, experimento de la, 299-300, 307, 310-314
- dualismo, significado del término, 45
- ecuaciones de movimiento de Lorentz, 245-248
- ecuación de ondas, computabilidad en la, 243-244
- ecuación de Schrödinger, 227, 316-318, 363-365, 374, 435
- ecuaciones de Euler-Lagrange, 568n
- ecuaciones dinámicas, 437
- ecuaciones diofánticas, 174
- efecto de marea (de la gravedad), 264-265
- Einstein, Albert, 197, 198, 248-249, 353, 356, 408n, 525, 572n
 - ecuaciones de campo de, 272
 - relatividad general de, 198, 261-273
 - relación masa-energía, 258, 281
- Einstein-Podolsky-Rosen (EPR)
 - paradoja, 353-360, 553
- electrodinámica cuántica (QED), 201, 365
- energía, conservación de la, 214-215, 382
- energía-momento (cuadri-vector), 283
- enmascaramiento retroactivo, 271
- entropía
 - definición de, 387-393
 - incremento de, 381-387
 - baja, origen de la, 398-403
- Entscheidungsproblem*, 61, 79, 81, 90
- equivalencia topológica de variedades, 174-175
- Escher, Maurits C., 205
- escisión cerebral, experimentos de, 476-479
- esfera de estados de Riemann, 336-339, 343-344
- espacio de fases, 229-239, 388-389, 449
 - división de grano-grueso del, 389, 389, 449, 451
 - efecto de difusión en, 236-237, 238
- espacio de Hilbert, 325-339
 - vectores, 325
 - suma de, 326-327
- espacio sináptico, 483, 483, 492
- espacio vectorial, 325
- espacio-tiempo tetra-dimensional, 250
- espinas dendríticas, 492
- estadística
 - de Bose-Einstein, 572n
 - de Fermi-Dirac, 572n
- estado con dos picos, 319-321
- estados cuánticos, 308-314, 374
 - copia de un, 341
 - medibilidad de, 339-341
 - objetividad de, 340
 - salto entre, 333, 355-356
- estados de momento, 311
- estados de spin, 333-336, 374
- estados discretos, 500
- estática, 197, 210
- estrella de neutrones, 403, 415
- estrella enana blanca, 413, 413
- estrella gigante roja, 413, 413
- Eudoxo, 207 y n
- Euler, Leonhard, 115, 124
 - fórmula de, 124
- Euler-Lagrange, ecuaciones de
- Everet III, Hugh, 372
- expansiones decimales, 115
- experimentos con fotones, 360-362
- evaporación de Hawking, 428, 439, 450, 450, 451
- evolución temporal inversa en el tiempo, 381, 442, 443
- evolución unitaria, 435, 442
- Faraday, Michael, 240, 242
- fascículo arqueado, 468, 469
- Fermat, Pierre de, 91
 - último teorema de, 91, 140, 141
- Fermi, Enrico, 572n
- Fermi-Dirac, estadística de, 572n
- fermiones, 334, 572n
- Feynman, Richar P., 201, 365
- fibra nerviosa, estructura de la, 483-485
- figura de interferencia, 296
- física clásica, 195-285
 - computabilidad en, 278-280
 - límites experimentales de, 289-291
- física cuántica, 196-197, 198
- FitzGerald-Lorentz, contracción de, 249
- fluctuación, significado del término, 451
- formación de juicios, 36
- formación reticular, 471, 472, 474-475, 510-511, 512-513

- fotones, 251-252
 fotosíntesis, 399
 Fourier, Joseph, 312
 transformada de, 312, 316
 fracciones, 79-80, 114
 Fredkin-Toffoli, computador de bolas de
 billar, 222-223, 239, 499
 Frege, Gottlob, 138
 Friedmann-Robertson-Walker
 modelos de, 403, 422
 función de onda
 colapso de, 317
 evolución de, 317
 paquete de ondas, 316
 significado del término, 308
 función hamiltoniana, 363
 funcionalismo, 554
 funciones delta, 312, 315
- Galileo Galilei, 117, 211-214, 216
 Galton, Francis, 525
 Gamow, Georges, 404
 Gauss, Carl Freidrich, 60, 204n
 Gell-Mann-Zweig, modelo de quarks de
 geodésicas, 201, 267, 268-269
 geometría de Lobachevski, 204-206, 406
 geometría de Minkowski, 250-260,
 268-269
 geometría euclídea, 199, 204-211, 405
 Gödel, Kurt, 61, 140
 GQC, 438, 453, 544
 gravitación cuántica, 431
 razón para la, 433-436
 gravitón, 344n
 Gregory, James, 115
 Grey, Walter, «tortuga» de, 34, 36, 39
- habitación china de Searle, 42-45
 habla, 467-468, 476, 527
 HCN, 438, 453
 Hadamard, Jacques, 519, 522, 524,
 522-526
 Hamilton, William Rowan, 227, 568n
 hardware, significado del término, 48-49
 Heisenberg, Werner, 201, 294
 Hilbert, David, 61, 137, 139, 570n
 hipocampo, 471, 475
 Hipótesis de curvatura de Weyl, 431, 433,
 436-441
 hipótesis de la célula de la abuela, 481
- hipótesis del continuo, 120
 Hofstadter, Douglas, 46
 horizonte de sucesos, 417-419
 Hubble, Edwin, 404n
 Hubel, David, 480
- IA fuerte, 40-48
 consciencia descrita por, 475, 504-505
 cerebro humano descrito en términos de
 ¿enemigo ficticio?, 554
 ideas matemáticas infusas, 135, 154
 individualidad
 de los átomos, 49-52, 352-353
 punto de vista de la IA fuerte, 52, 54-55
 inspiración, 518-525
 inteligencia, significado del término,
 504-505
 inteligencia artificial (IA), 11, 13, 33-36
 aproximación al «placer» y al «dolor»,
 36-40
 fuerte, 11, 13, 40-48
 descripción del cerebro en términos
 de, 469
 consciencia descrita por, 475, 504-505
 ¿enemigo ficticio?, 554
 interferencia cuántica, 306
 interferencia destructiva, 297-298
 interpretación de los muchos-universos,
 372-373, 536
 intuición matemática, 147-153
 naturaleza no-algorítmica de, 512-513,
 516-518
 intuiciones, 71, 150
 intuicionismo, 153-158
 irreversibilidad temporal, 443-447, 453
- Kepler, Johannes, 218
 Kleene, Stephen C., 79, 101
 Kornhuber, H. H., 545, 545, 549-550
- Lagrange, Joseph C., 568n
 Landau-Oppenheimer-Volkov, límite de,
 415
 lenguaje, 475-476
 ley de radiación de Planck, 292, 293, 311
 ley de la suma vectorial, 217
 Libet, Benjamin, 546, 547, 548
 libre albedrío, 221, 534
 líneas de universo, 250, 270
 líneas espectrales, 290, 291, 570n

- Liouville, Joseph, 235
- lista de referencias, 579-589
- Lobachevski, Nicolai Ivanovich, 204
- logaritmo natural, base de, 124n
- longitud de Planck, 434 y n
- Lorentz, Hendrick Antoon, 245, 249
- ecuaciones de movimiento de Lorentz, 245-248
- ley de fuerza de Lorentz, 249, 280
- Lorenz, Konrad, 528
- luz del Sol, 399, 400-401
- luz monocromática, 296
- Mandelbrot, Benoit, 12, 132
- máquina de teleportación, 53, 341, 478
- máquina de Turing, 60, 62-71
- algoritmo de Euclides como, 69, 71, 74
- cerebro como, 469, 490
- de sumar 1 a un número binario expandido, 75, 82, 83
- de sumar 1 a un número unario, 69, 70, 75
- de duplicar un número binario expandido, 75
- de duplicar un número unario, 76, 84
- numeración de, 81-84
- problema de la detención, 90-91, 96
- universal, 81-90, 222
- computadores de uso general como, 49, 89
- más-rápidas-que-la-luz, dispositivo de señales, 274-275
- masa de Planck, 458
- masa en reposo, 281, 284
- matemática no recursiva
- ejemplos de, 174-183
- ¿semejanza con el conjunto de Mandelbrot?, 183-187
- materia oscura, 408
- matrices densidad, 371
- Maxwell, James Clerk, 197, 241
- mecánica cuántica, actividad cerebral descrita por la, 497-499
- mecánica hamiltoniana, 326-229
- medidas mecano-cuánticas, 329-333
- mente, significado del término, 502-507
- Mermin, David, 359, 572n
- mielina, 486
- miniagujeros negros, 428
- Minkowski, Hermann, 250
- modelo de gas-en-una-caja, 390-391
- modelo del sistema sol (de un átomo), 290
- módulo al cuadrado, significado del término, 304
- momento angular, 216, 293, 346 y n
- movimiento de Poincaré, 259
- Mozart, Wolfgang Amadeus, 524, 551
- muchos-universos, interpretación de los mundo de bolas de billar, computabilidad en, 221-226
- mundo platónico de los conceptos matemáticos, 207, 530
- Neumann, John von, 570n, 577n
- neuronas, 482
- neurotransmisores, 486, 493
- Newton, Isaac, 196, 216-218, 567n, 568n
- niño, visión de, 554-556
- notación binaria expandida, 73
- números
- complejos, 122-128
- computables, 116, 120, 170
- enumerables, 118, 169
- imaginarios, 123-124
- irracionales, 81
- naturales, significado del término, 58, 114
- negativos, 79-80, 114
- no naturales, 79-81
- racionales, 114
- reales, 114-117, 207
- «realidad de», 121-122
- transfinitos, 117
- objetos con gran spin, 345-347
- ondas gravitatorias, 289
- ondas electromagnéticas, 241, 284
- circularmente polarizadas, 342-344
- plano polarizadas, 342-344
- Oppenheimer-Snyder, agujeros negros de orden lexicográfico, 145 y n, 159
- originalidad (del pensamiento), 523
- paradoja de Einstein-Podolski-Rosen, 353-360, 553
- paradoja de los dos-lugares-a-la-vez, 319-325
- paradoja de los gemelos, 256-257
- paradoja de Russell, 137, 138, 139n, 151
- paralelismo cuántico, 495

- partículas de prueba, 280
 partículas de spin $\frac{1}{2}$, 334-335, 357-358
 partículas desbocadas, 247
 partículas distinguibles, 350, 352
 Pauli, Wolfgang, 201, 572n
 Penfield, Wilder, 473
 Penrose, Roger, 182, 520-521
 pensamiento analítico, 476-477, 526, 527
 pensamiento filosófico, 520
 pensamiento geométrico, 476, 527
 pensamiento (no verbalidad del), 525-527
 pensamiento valorativo, 36, 510-511, 512-513
 percepción consciente, papel del tiempo en, 549-554
 pi (π), 115, 156
 «placer-dolor», aproximación IA al, 36-40
 plano de Argand, 125-128, 304, 305, 306
 plasticidad cerebral, 491-493, 542-544
 Platón, 134, 206-207, 530
 Platonismo, 134, 153-155, 206-207
 Podolski, Boris, 356
 Poincaré, Henri, 137, 248, 519
 polarización (de la luz), 342-345
 Pour-El-Richards, no computabilidad de fenómeno, 244, 277n
 principio antrópico, 440, 537-538
 principio de exclusión de Pauli, 351, 572n, 414, 415, 574n
 principio de equivalencia, 264
 principio de incertidumbre, 314-316, 318
 principio de reflexión, 150
 principio de la relatividad, punto de vista de Eusstein, 249
 principio de relatividad de Galileo, 213
 problema de las palabras, 175-178
 problemas NP, 190-193
 procedimientos de evolución, 316-317
 procedimiento sistemático, significado del término, 57
 programa de Hilbert para las matemáticas, 136-140
 puertas lógicas, 487-490
 puerta NO, 489
 puerta Y, 485
 Purkinje, células de, 491
 radiación de fondo de cuerpo negro (en el Universo), 404, 429
 radiación de Hawking, 450-451
 radio de Schwarzschild, 416-417
 Ramón y Cajal, 577n
 Rayleigh-Jeans, radiación de, 290-291, 292
 reacciones termonucleares, 401
 realidad física, 209, 286, 532-534, 570n
 recurrencia de Poincaré, 394n
 redes neurales, 493
 reducción del vector de estado, 317, 333, 435, 441, 455, 456
 asimetría temporal en la, 441-447
 factores que afectan, 457-462
reductio ad absurdum, método de, 92n, 119
 relación masa-energía, 258, 281
 relatividad
 teoría especial de la, 197-198, 230, 249-260
 teoría general de la, 198, 199, 206, 261-273
 renormalización, 365
 resultado de Turing, teorema tipo Gödel a partir de, 158-161
 retina, 480, 496
 RICCI, 266, 271, 422
 Ricci, Gregorio, 570n
 Riemann, Bernhard, 570n
 Robinson, Raphael, 180, 181, 182
 robótica, 33
 Rosen, Nathan, 356
 Russell, Bertrand, 137, 138-139
 Rutherford, Ernest, 290
 Schank, Roger, 41
 Schrödinger, Erwing
 Schrödinger, 368, 376
 ecuación de, 227, 316-318, 363-365, 374, 374, 435
 paradoja del gato de, 366-369
 Searle, John, 48
 secuencia computable, significado del término, 94
 selección natural, 502, 506-507, 513-515
 señales nerviosas, 482-486
 Shechtman, Dany
 simetría temporal en la teoría de gravitación cuántica, 441, 452-453
 simultaneidad, 260, 570n

- sinapsis, 482, 483
- sinapsis de Hebb, 493
- singularidad inicial, 436, 439
- singularidades del espacio-tiempo, 420-424, 430
- sistema binario de numeración, 67, 70-76
- sistema ptolemaico, 203, 207n
- sistemas de muchas partículas, 347-353
- sistemas expertos, 33
- sistemas matemáticos formales, 140-144
 - computabilidad de, 158-159
- software, significado del término, 48-49
- soplos de inspiración, 518-525
- Sperry, Roger, 476
- spin del fotón, 342-345
- Stern-Gerlach, aparato de, 571n
- Stokes, vector de, 345
- subrutina, significado del término, 59
- superposición lineal cuántica, 324, 326

- temperatura de Hawking, 427-428
- tensor de curvatura de Riemann, 271
- tensor de Ricci, 570n, 271, 422
- tensor de Weyl, 570n, 271, 422
- teorema
 - de Bell
 - de Gödel, 61, 140, 144-147, 517
 - punto de vista de Turing, 158-161
 - de Liouville, 235, 237, 452, 453
 - de Ptolomeo, 209
- teoremas de singularidad, 421
- teoría
 - cuántica
 - comienzos de la, 291-294
 - desarrollos más allá de la, 499-501
 - diversas actitudes hacia la, 369-373
 - discusión sobre el futuro de la, 373
 - naturaleza provisional de la, 353-354, 376
 - cuántica de campos, 365-366
 - de conjuntos, 161
 - de Gold (sobre el petróleo), 402, 574n
 - de la complejidad, 187-193
 - en los objetos físicos, 193-194
 - de twistors, 13, 568n, 577n
 - de variables ocultas, 354-355
 - de Yang-Mills, 201
 - del big bang, 202, 404-408, 409
 - del big bang caliente, 404, 409
 - electromagnética de Maxwell, 197, 239-243, 245, 364
 - física, status de, 195-203
 - tercera ley del movimiento de Newton, 216-217
 - termodinámica
 - primera ley de la, 382-383
 - segunda ley de la, 381, 386, 393-398
 - explicación por la teoría del big bang, 411-412
 - teselación (del plano), 179-183
 - teselación cuasi-periódica, 184, 540
 - teselaciones no periódicas, 178-183
 - teselaciones periódicas, 178-180
 - teselas de Penrose, 14, 184, 522
 - test de Turing, 26-33
 - tiempo
 - de la percepción consciente, 549-554
 - flujo del, 378-380, 550
 - «Tor'Bled-Nam», 107-114
 - Valle de los Caballos de Mar en, 109-111
 - tronco cerebral superior, 473
 - Turing, Alan, 27, 60

- U, 317, 364, 379, 435, 441, 570n
- Universo en expansión, 405-408
- Universo participatorio, 372

- variables ocultas, teorías de, 354-355
- variedad, equivalencia topológica de variedades riemannianas, 570n
- velocidad de escape, 416
- verbalización, 528
- verdad matemática, 136, 149, 150
- versátil (forma de tesela), 178, 181

- Wallis, John, 115, 125
- Wang, Hao, 180
- Wernicke, área de, 468-469, 476
- Wessel, Caspar, 125
- WEYL**, 264, 271, 422, 436
- Weyl, Hermann, 570n
- Wheeler, John A., 372, 578n
- Whitehead, Alfred North, 139
- Wiesel, Torsten, 480
- Wigner, Eugene P. 371
- Wilson, cámara de niebla de, 458, 459
- Wilson, Donald, 477

X

La nueva mente del emperador

DUE DATE ON RECEIPT

DUE DATE ON RECEIPT

DEMCO

BM 9

Carlos García Gual
LECTURAS Y FANTASÍAS MEDIEVALES

BM 10

Serge Guilbaut
DE CÓMO NUEVA YORK ROBÓ LA IDEA DE ARTE
MODERNO

BM 11

Norman Malcolm
LUDWIG WITTGENSTEIN

BM 12

Hans Sedlmayr
LA REVOLUCIÓN DEL ARTE MODERNO

BM 13

Ralf Dahrendorf
EL CONFLICTO SOCIAL MODERNO

BM 14

Maria Gondolo della Riva
CHOPIN

BM 15

Paul Veyne
LA SOCIEDAD ROMANA

BM 16

Cristóbal Zaragoza
CERVANTES

BM 17

Bronislaw Geremek
LA ESTIRPE DE CAÍN

BM 18

Carlos García Gual
FIGURAS HELÉNICAS
Y GÉNEROS LITERARIOS

BM 19

José Antonio Maravall
ESTUDIOS DE LA HISTORIA
DEL PENSAMIENTO ESPAÑOL (S. XVIII)

BM 20

Roger Penrose
LA NUEVA MENTE DEL EMPERADOR

D

urante décadas los defensores de la inteligencia artificial han mantenido que los ordenadores harán pronto todo lo que la mente humana puede hacer. En su favor se puede utilizar, por ejemplo, el que ya hay máquinas que juegan al ajedrez como los grandes maestros. Ahora bien, ¿comprenden el juego como lo hacemos nosotros? En este libro, ROGER PENROSE, probablemente el especialista en la teoría general de la relatividad más prestigioso del mundo y una de las mentes analíticas más originales de la actualidad, sostiene que existen facetas del pensamiento humano que nunca serán emuladas por un ordenador. Para defender esa tesis, Penrose recurre a una amplia gama de conocimientos científicos, que van desde la máquina de Turing hasta la estructura del cerebro, pasando por otros como pueden ser el teorema de la indecidibilidad de Gödel, los agujeros negros y los blancos, la radiación de Hawking, la entropía, o la mecánica cuántica. Entre los numerosos estudios existentes dedicados a la relación entre la mente y el cuerpo (un problema sempiterno de la filosofía), esta ambiciosa obra sobresale tanto por su lucidez y claridad como por su rigor y profundidad.

ISBN 84-397-1786-5



9 788439 717867