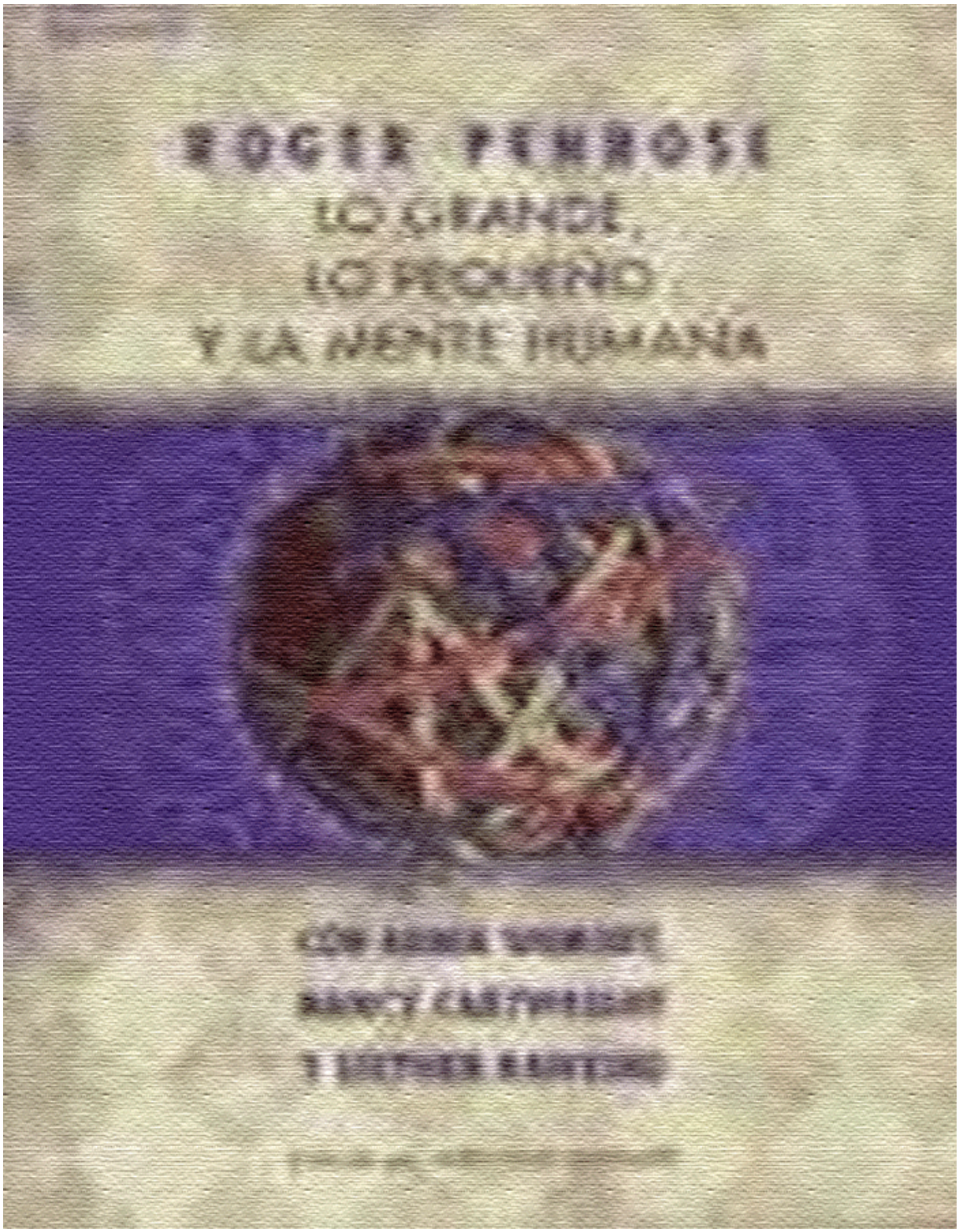
Lo grande, lo pequeño y la mente humana

ROGER PENROSE



LIBROdot.com
http://www.librodot.com

Lo grande, lo pequeño y la mente humana

ROGER PENROSE
Rouse Ball Professor de Matemáticas, Universidad de Oxford

*

con
ABNER SHIMONY
Catedrático Emérito de Filosofía y Física, Universidad de Boston

NANCY CARTWRIGHT
Catedrática de Filosofía, Lógica y Método Científico,
London School of Economics and Political Science

y
STEPHEN HAWKING
Catedrático Lucasiano de Matemáticas, Universidad de Cambridge

*

Edición de Malcolm Longair

*

Traducción de Javier García Sanz

CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press agradece la cooperación del Presidente y los miembros de la Junta Rectora de Clare Hall, de Cambridge, bajo cuyos auspicios se celebraron las «Conferencias Tanner sobre los Valores Humanos» de 1995, de las cuales se deriva este libro.

Indice

Prólogo por Malcolm Longair
Capítulo 1 Espacio-tiempo y cosmología
Capítulo 2 Los misterios de la física cuántica
Capítulo 3 La física y la mente
Capítulo 4 Sobre mentalidad, mecánica cuántica y la actualización de potencialidades *Abner Shimony*
Capítulo 5 ¿Por qué física? *Nancy Cartwright*
Capítulo 6 Las objeciones de un reduccionista descarado *Stephen Hawking*
Capítulo 7 Respuesta *Roger Penrose*
Apéndice 1
Apéndice 2

The Emperor's New Mind, R. Penrose, 1989, Oxford, Oxford University Press: 1.6, 1.8, 1.11, 1.12, 1.13, 1.16.a, l.l6.b, l.l6.c, 1.18, 1.19, 1.24, 1.25, 1.26, 1.28.a, 1.28.b, 1.29, 1.30, 2.2, 2.5.a y 3.20.
Shadows of the Mind, R. Penrose, 1994, Oxford, Oxford University Press: 1.14. 2.3, 2.4, 2.5.b, 2.6, 2.7, 2.19, 2.20, 3.7, 3.8, 3.10, 3.11, 3.12, 3.13, 3.14, 3.16, 3.17 y 3.18.
High Energy Astrophysics, Vol. 2, M.S. Longair, 1994, Cambridge, Cambridge University Press: 1.15 y 1.22.
Cortesía de Cordón Art-Baarn-Holand © 1989: 1.17 y 1.19.

Prólogo por
Malcolm Longair

Uno de los sucesos más alentadores de la última década ha sido la publicación de varios libros escritos por científicos eminentes en los que intentan comunicar la esencia y lo apasionante de su ciencia al lector profano. Algunos de los ejemplos más sorprendentes incluyen el extraordinario éxito de *Una historia del tiempo* de Stephen Hawking, que ahora forma parte de la historia de la edición; el libro *Caos* de James Gleick, que muestra de qué manera tan acertada un tema intrínsecamente difícil puede convertirse en una historia de detectives; y *El sueño de una teoría final* de Steven Weinberg, que hace notablemente accesibles y convincentes la naturaleza y los objetivos de la física de partículas contemporánea.

En este alud de divulgación, el libro *La nueva mente del emperador* de Roger Penrose, publicado en 1989, destaca como algo particularmente diferente. Mientras que otros autores se proponían comunicar el contenido y la emoción de la ciencia contemporánea, el libro de Roger presentaba una visión sorprendente y original acerca de los muchos aspectos, en apariencia dispares, de la física, las matemáticas, la biología, las ciencias del cerebro e incluso la filosofía, que podían ser subsumidos dentro de una nueva, aunque todavía no definida, teoría de los procesos fundamentales. No es de extrañar que *La nueva mente del emperador* provocase una gran controversia; y, en 1994, Roger publicó un segundo libro, *Las sombras de la mente*, en el cual intentaba responder a diversas críticas a sus argumentos y ofrecía nuevas intuiciones y desarrollos de sus ideas. En sus «Conferencias Tanner» de 1995, presentó una revisión de los temas centrales discutidos en sus dos libros y después participó en una discusión sobre ellos con Abner Shimony, Nancy Cartwright y Stephen Hawking. Las tres conferencias, reproducidas en los capítulos 1, 2 y 3 de este libro, proporcionan una somera introducción a las ideas expuestas con mucho más detalle en los dos libros anteriores; y las aportaciones de los tres participantes, en los capítulos 4, 5 y 6, plantean muchas de las cuestiones que aquellas habían suscitado. Finalmente, Roger tiene la oportunidad de comentar estas cuestiones en el capítulo 7.

Los capítulos de Roger hablan por sí mismos de manera elocuente, pero unas palabras de introducción podrían fijar el escenario para el enfoque particular que Roger sigue en relación con algunos de los problemas más profundos de la ciencia moderna. Ha sido reconocido internacionalmente como uno de los matemáticos contemporáneos mas dotados, pero sus investigaciones siempre han estado firmemente situadas en un escenario físico real. El trabajo por el que es más famoso en astrofísica y cosmología se refiere a dos teoremas relacionados con las teorías relativistas de la gravedad, tarea que, en parte, fue llevada a cabo conjuntamente con Stephen Hawking. Uno de estos teoremas demuestra que. según las teorías relativistas clásicas de la gravedad, dentro de un agujero negro debe existir inevitablemente una singularidad física, es decir, una región en la que la curvatura del espacio o, de forma equivalente, la densidad de materia, se hace infinitamente grande. El segundo teorema afirma que, de acuerdo con las teorías relativistas clásicas de la gravedad, existe inevitablemente una singularidad física similar en el origen en los modelos cosmológicos de Gran Explosión. Estos resultados indican que, en cierto sentido, esas teorías no están acabadas, puesto que las singularidades físicas deberían ser evitadas en cualquier teoría físicamente significativa.

Este es, sin embargo, solo un aspecto de un enorme abanico de aportaciones hechas en muchas áreas diferentes de las matemáticas y de la física matemática. El procedimiento de Penrose explica de qué modo las partículas pueden extraer energía de la energía rotacional de los agujeros negros. Los diagramas de Penrose se utilizan para estudiar el comportamiento de la materia en la vecindad de los agujeros negros. Subyacente a buena parte de su enfoque, hay un sentido geométrico muy fuerte, casi pictórico, que está presente a lo largo de los capítulos 1, 2 y 3. El público en general está más familiarizado con este aspecto de su trabajo a través de las imágenes *imposibles* de M.C. Escher y por las teselas de Penrose. Resulta fascinante que fuera el artículo de Roger y de su padre, L.S. Penrose, el que proporcionara la inspiración para varios grabados *imposibles* de Escher. Además, las figuras del *Límite Circular* de Escher se utilizan para ilustrar el entusiasmo de Roger por las geometrías hiperbólicas en el capítulo 1. Los teselados de Penrose son construcciones geométricas notables en los que un plano infinito puede ser completamente cubierto mediante teselas de unas pocas formas diferentes. Los ejemplos más sorprendentes de estos teselados son aquellos que pueden cubrir por completo un plano infinito pero en los cuales no hay repetición; en otras palabras, no es posible encontrar el mismo patrón de teselas en ninguna región del plano infinito. Este tema vuelve a aparecer en el capítulo 3 en relación con la cuestión de si conjuntos

específicos de procedimientos matemáticos exactamente definidos pueden o no ser llevados a cabo por un ordenador.

Roger aporta así un formidable arsenal de armas matemáticas además de extraordinarios logros en esta disciplina y en física en algunos de los más trascendentales problemas de la física moderna. La realidad y la importancia de las cuestiones que aborda están fuera de discusión. Los cosmólogos tienen buenas razones para estar firmemente convencidos de que la Gran Explosión proporciona la imagen más convincente que tenemos para comprender las características a gran escala de nuestro Universo. Esta imagen es, sin embargo, incompleta en varios aspectos. La mayoría de los cosmólogos están convencidos de que tenemos una buena comprensión de la física básica necesaria para explicar las propiedades globales del Universo desde, aproximadamente, el instante en que tenía una milésima de segundo hasta el momento presente. Pero la imagen solo resulta correcta si fijamos las condiciones iniciales de una forma muy cuidadosa. El gran problema reside en que, al tratar el Universo cuando era significativamente más joven que una milésima de segundo, nos estamos apartando de la física ensayada y comprobada, y por ello tenemos que basarnos en extrapolaciones razonables de las leyes conocidas de la física. Sabemos bastante bien cuáles debieron de haber sido estas condiciones iniciales, pero el por qué fueron así es un asunto especulativo. Todos están de acuerdo en que estos problemas figuran entre los más importantes de la cosmología contemporánea.

Se ha desarrollado un marco estándar para tratar de resolver estos problemas, este modelo se conoce como la imagen inflacionaria del Universo primitivo. Incluso en este modelo se supone que algunos rasgos de nuestro Universo se originaron en los instantes más tempranos, en lo que se conoce como la *era de Planck*, para la que se hace necesario comprender la gravedad cuántica. Este periodo transcurrió cuando el Universo tenía, aproximadamente, solo 10^{-43} segundos; lo que puede parecer algo excesivo, pero sobre la base de lo que conocemos hoy tenemos que tomar muy en serio lo que sucedió en esos instantes iniciales.

Roger acepta la imagen convencional de la Gran Explosión, hasta donde ésta alcanza, pero rechaza la imagen inflacionaria de sus etapas primitivas. Más bien, cree que nos faltan algunos elementos de la física y que estos deben estar asociados con una adecuada teoría cuántica de la gravedad, una teoría que todavía no tenemos pese al hecho de que los teóricos han estado tratando de resolver este problema durante muchos años. Roger argumenta que se han enfrentado al problema equivocado. Una parte de su interés está relacionada con la entropía del Universo en conjunto. Puesto que la entropía o, para decirlo de forma más sencilla, el desorden, aumenta con el tiempo, el Universo debe haber empezado en un estado altamente ordenado con muy poca entropía. La probabilidad de que esto sucediera por azar es prácticamente nula. Roger señala que este problema debería quedar resuelto como parte de la teoría correcta de la gravedad cuántica.

La necesidad de cuantización conduce a la discusión de los problemas de la física cuántica en el capítulo 2. La mecánica cuántica y su extensión relativista en la teoría cuántica de campos han sido tremendamente acertadas para dar cuenta de muchos resultados experimentales en física de partículas y para explicar las propiedades de los átomos y de las partículas, aunque se necesitaron muchos años antes de que fuera apreciada en su totalidad la importancia física de la teoría. Como Roger ilustra de forma muy bella, la teoría contiene como parte de su estructura rasgos determinantes no intuitivos, que no poseen equivalente en la física clásica. Por ejemplo, el fenómeno de la no localidad significa que, cuando se produce un par de partículas materia-antimateria, cada partícula conserva un *recuerdo* del proceso de creación, es decir, que no pueden considerarse completamente independientes una de otra. Tal como Roger lo expresa: «El enmarañamiento cuántico es un fenómeno muy extraño. Es algo intermedio entre objetos que están separados y que están comunicados entre si». La mecánica cuántica nos permite también obtener información acerca de procesos que podrían haber sucedido pero no sucedieron. El ejemplo más sorprendente que plantea es el extraordinario problema de la comprobación de bombas de Elitzur-Vaidman que refleja las profundas diferencias entre la mecánica cuántica y la física clásica.

Estos aspectos no intuitivos son parte de la estructura de la física cuántica, pero existen problemas más profundos. Aquellos en los que se centra Roger se refieren al modo en que relacionamos los fenómenos que ocurren en el nivel cuántico con el nivel macroscópico en el que hacemos las observaciones de sistemas cuánticos. Esta es una área controvertida. La mayoría de los físicos profesionales utilizan simplemente las reglas de la mecánica cuántica como una herramienta computacional que proporciona respuestas de una precisión extraordinaria. Si aplicamos las reglas correctamente, obtendremos las respuestas correctas. Esto incluye, sin embargo, un proceso poco elegante para traducir fenómenos desde el sencillo mundo lineal en el nivel cuántico al mundo del experimento real. Ese proceso implica lo que se conoce como *colapso de la función de onda* o *reducción del vector de estado.* Roger cree que faltan

algunas piezas físicas fundamentales en la imagen convencional de la mecánica cuántica. Argumenta que es necesaria una teoría completamente nueva que incorpore lo que él llama la *reducción objetiva de la función de onda* como parte integral de la teoría. Esta nueva teoría debe reducirse a la mecánica cuántica convencional y a la teoría cuántica de campos en el límite apropiado, pero es probable que introduzca nuevos fenómenos físicos. En ello podrían estar las soluciones al problema de la cuantización de la gravedad y la física del Universo primitivo.

En el capítulo 3, Roger intenta descubrir las características comunes a las matemáticas, la física y la mente humana. Es sorprendente que la más rigurosamente lógica de las ciencias, las matemáticas abstractas, a menudo, no puedan ser programadas en un ordenador, por muy preciso que este sea y por mucha capacidad que tenga su memoria. Un ordenador no puede descubrir teoremas matemáticos tal como lo hacen los matemáticos humanos. Esta conclusión sorprendente se deriva de una variante de lo que se denomina el *teorema de Gödel.* La interpretación de Roger es que esto significa que la construcción del pensamiento matemático y, por extensión, todo pensamiento y comportamiento consciente, se lleva a cabo por medios *no computacionales.* Esta es una clave fructífera porque nuestra intuición nos dice que la enorme diversidad de nuestras percepciones conscientes es también *no computacional.* Debido a la importancia capital de este resultado para su argumento general, Roger dedicaba más de la mitad de su libro *Las sombras de la mente* a demostrar que su interpretación del teorema de Gödel era irrebatible.

Roger considera que los problemas de la mecánica cuántica y los de la comprensión de la conciencia están relacionados de varias maneras. La no localidad y la coherencia cuántica sugieren, en principio, modos en los que grandes áreas del cerebro podrían actuar coherentemente. Cree que los aspectos no computacionales de la conciencia pueden estar relacionados con los procesos no computacionales que podrían estar involucrados en la reducción objetiva de la función de onda observable en niveles macroscópicos. No contento con enunciar simplemente principios generales, intenta identificar las diferentes estructuras internas del cerebro que podrían ser capaces de sostener semejantes tipos de nuevos procesos físicos.

Este resumen apenas hace justicia a la originalidad y fecundidad de estas ideas y a la brillantez con que se desarrollan en este libro. A lo largo de la exposición, varios temas subyacentes tienen una gran importancia para determinar la dirección de su pensamiento.

Quizá el más importante sea la notable capacidad de las matemáticas para describir procesos fundamentales en el mundo natural. Tal como Roger lo expresa, el mundo físico emerge en cierto sentido del mundo platónico de las matemáticas. Sin embargo, nosotros no obtenemos las nuevas matemáticas a partir de la necesidad de describir el mundo, o de hacer que el experimento y la observación encajen en reglas matemáticas. La comprensión de la estructura del mundo puede venir de amplios principios generales y de las propias matemáticas.

No debe sorprender que ambas propuestas hayan sido tema de controversia. Las contribuciones de los participantes dan una idea de muchas de las preocupaciones manifestadas por expertos con formaciones intelectuales muy diferentes. Abner Shimony coincide con Roger en varios de sus objetivos: en que existe cierta insuficiencia en la formulación estándar de la mecánica cuántica, según las mismas líneas subrayadas por Roger, y en que los conceptos mecanocuánticos son relevantes para la comprensión de la mente humana. Afirma, sin embargo, que Roger «es un alpinista que ha tratado de escalar la montaña equivocada» y sugiere formas alternativas de considerar las mismas áreas de interés de una manera constructiva. Nancy Cartwright plantea la cuestión básica de si la física es o no el punto de partida correcto para entender la naturaleza de la conciencia. Plantea también el espinoso problema de si las leyes que gobiernan diferentes disciplinas científicas pueden ser obtenidas realmente unas a partir de otras. El más crítico de todos es Stephen Hawking, viejo amigo y colega de Roger. En muchos aspectos la posición de Hawking es la más próxima a la que podría denominarse la posición típica del físico *medio*, y le desafía a que desarrolle una teoría detallada de la reducción objetiva de la función de onda. Niega que la física tenga mucho de valor que decir sobre el problema de la conciencia. Todas estas son preocupaciones justificables, pero Roger defiende su posición en el último capítulo de este libro.

Roger ha creado con éxito una visión o manifiesto acerca del desarrollo que podría tener la física matemática en el Siglo XXI. A lo largo de los capítulos 1, 2 y 3, construye una narrativa que sugiere cómo cada parte de la historia podría encajar en una imagen coherente de un tipo de física completamente nuevo que lleva incorporadas sus preocupaciones fundamentales de no computabilidad y la reducción objetiva de la función de onda. La prueba de estos conceptos dependerá de la capacidad de Roger y de otros para hacer realidad este nuevo tipo de teoría física. E incluso si este programa no tuviera un éxito inmediato, ¿son

fructíferas las ideas inherentes en el concepto general para el desarrollo futuro de la física teórica y las matemáticas? En verdad sería muy sorprendente que la respuesta fuera «no».

1
Espacio-tiempo y cosmología

El título de este libro es *Lo grande, lo pequeño y la mente humana,* y el tema de este primer capítulo es *lo grande.* Los capítulos 1 y 2 están dedicados a nuestro Universo físico, que voy a representar muy esquemáticamente mediante la *esfera* de la figura 1.1. Sin embargo, en ellos no se va a describir con detalle qué es lo que hay en uno u otro lugar de nuestro Universo, sino más bien quiero concentrarme en la comprensión de las leyes reales que rigen el comportamiento del mundo. Una de las razones por las que he decidido dividir las descripciones de las leyes físicas en dos capítulos: *lo grande* y *lo pequeño,* es que las leyes que determinan el comportamiento del mundo a gran escala y las que rigen su comportamiento a pequeña escala parecen muy diferentes. El hecho de que parezcan ser tan diferentes, y de que quizá tengamos algo que ver con esta aparente discrepancia, es fundamental para el tema del capítulo 3, y aquí es donde interviene la mente humana.

Dado que voy a hablar acerca del mundo físico en términos de las teorías físicas que subyacen a su comportamiento, también tendré que decir algo sobre otro mundo, el mundo platónico de las ideas y, concretamente, sobre su papel como mundo de la verdad



Figura 1.1.

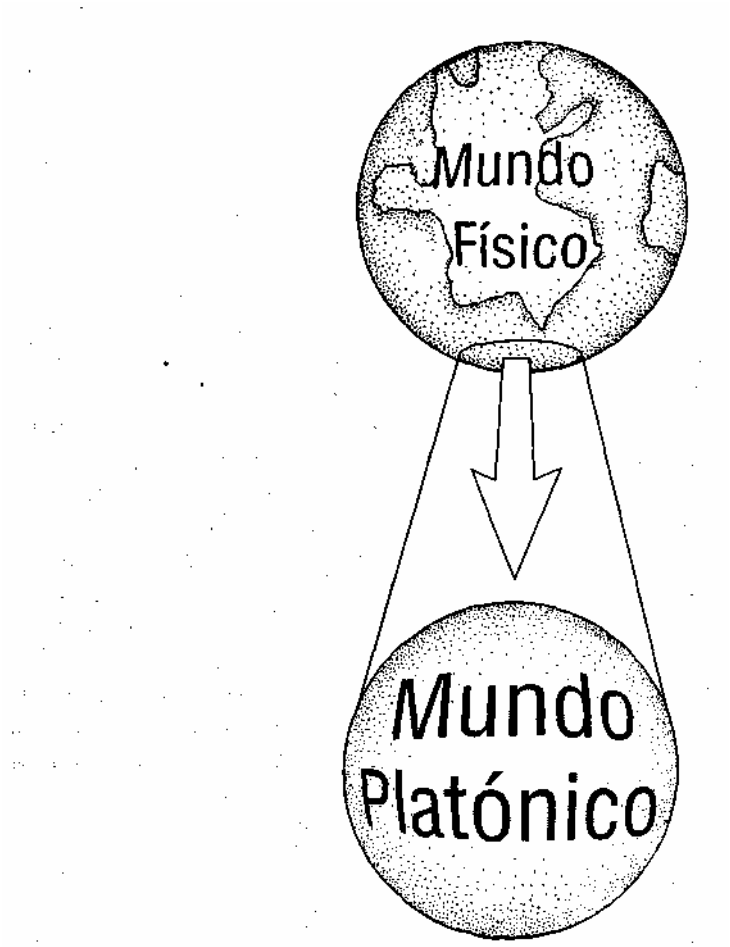


Figura 1.2.

matemática. Uno puede muy bien adoptar el punto de vista de que el *mundo platónico* contiene otras ideas, tales como la *Bondad y* la *Belleza*, pero aquí solo me interesaré por los conceptos platónicos de las matemáticas. Para algunas personas resulta difícil concebir que este mundo tenga una existencia independiente. Preferirán considerar los conceptos matemáticos como meras idealizaciones de nuestro mundo físico y, desde esta perspectiva, el mundo matemático se concebiría como algo que emerge del mundo de los objetos físicos (figura 1.2).

Pero no es así como yo concibo las matemáticas, ni creo que la mayoría de los matemáticos y los físicos matemáticos tengan esa idea del mundo. Lo conciben de un modo bastante diferente, como una estructura gobernada de manera precisa y de acuerdo con leyes matemáticas intemporales. Por eso encuentran más apropiado considerar el mundo físico como algo que emerge del *intemporal* mundo de las matemáticas, tal como se ilustra en la figura 1.3. Esta imagen tendrá importancia en el tema del capítulo 3, y también subyace a mucho de lo que diré en los capítulos 1 y 2.



Figura 1.3.

Uno de los aspectos más importantes sobre el comportamiento del Universo es que parece estar basado en las matemáticas hasta un grado de precisión extraordinario. Cuanto mejor entendemos el mundo físico, y más profundamente sondeamos en las leyes de la naturaleza, más nos parece que la realidad física se evapora hasta que nos quedamos solo con las matemáticas. Cuanto mejor entendemos las leyes de la física, más nos vemos abocados a este ámbito de las matemáticas y de los conceptos matemáticos.

Observemos las escalas con las que tenemos que tratar en el Universo y también el papel de nuestro lugar en él. Puedo resumir todas estas escalas en un único diagrama (figura 1.4). En el lado izquierdo se muestran las escalas de tiempo y en el lado derecho están las correspondientes escalas de distancias. En la parte inferior del diagrama, a la izquierda, se encuentra la escala de tiempo más corta que tiene sentido físico. Esta escala de tiempo es de, aproximadamente, 10^{-43} segundos y suele conocerse como *escala de tiempo de Planck* o *cronón.* Este intervalo de tiempo es mucho más corto que cualquier otro fenómeno experimentado en la física de partículas. Por ejemplo, las partículas de vida más corta, llamadas *resonancias,* existen durante unos $1O^{-23}$ segundos. Más arriba en el diagrama, a la izquierda, se muestran el día y el año y, en la parte superior, la edad actual del Universo.

En el lado derecho del gráfico aparecen las distancias correspondientes a estas escalas de tiempo. La longitud que corresponde a la escala de tiempo de Planck (un cronón) es la unidad fundamental de longitud, llamada *longitud de Planck.* Los conceptos

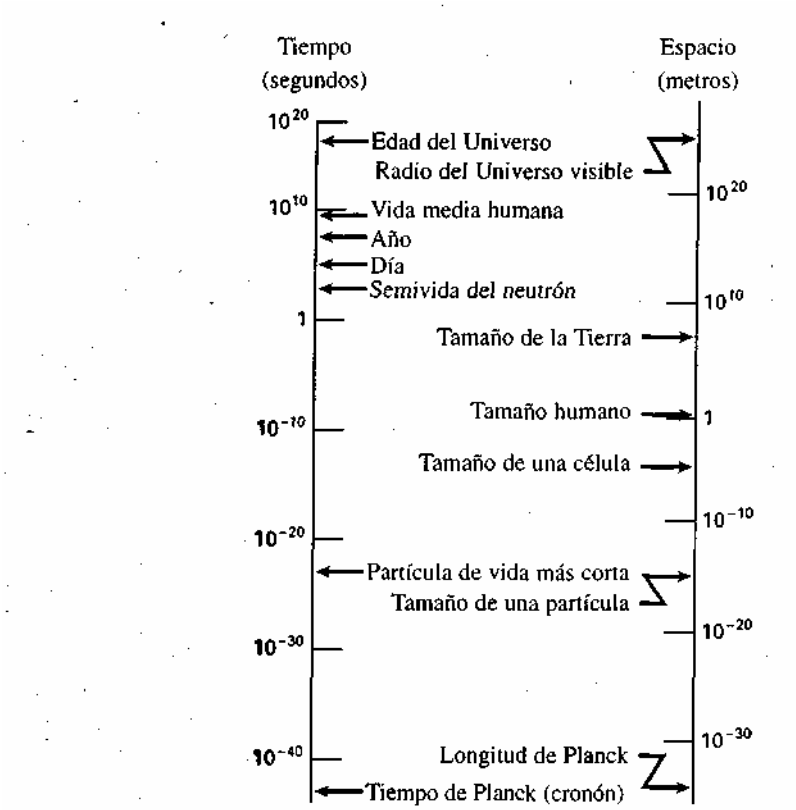


Figura 1.4. Escalas de tamaño y de tiempo en el Universo.

de tiempo y de longitud de Planck aparecen de forma natural cuando tratamos de combinar las teorías físicas que describen lo grande y las que definen lo pequeño, o, lo que es lo mismo, combinar la teoría general de la relatividad de Einstein, que se ocupa de la física de lo muy grande, con la mecánica cuántica, que describe la física de lo muy pequeño. Cuando se unen estas teorías, estas longitudes y tiempos de Planck resultan ser fundamentales. La traducción desde el eje izquierdo hasta el eje derecho del diagrama se hace mediante la velocidad de la luz, de modo que los tiempos pueden traducirse en distancias preguntando a qué distancia podría viajar una señal luminosa en ese intervalo temporal.

Los tamaños de los objetos físicos representados en el diagrama abarcan desde 10^{-15} metros (el tamaño característico de las partículas) hasta 10^{27} metros (el radio del Universo observable en la actualidad, que es aproximadamente la edad del Universo multiplicada por la velocidad de la luz). Resulta curioso advertir el lugar donde nos encontramos *nosotros* en el diagrama, la escala humana. Puede verse que, con respecto a las dimensiones espaciales, estamos más o menos en el centro de la gráfica. Somos enormes comparados con la longitud de Planck; incluso si se nos compara con el tamaño de las partículas somos muy grandes. Pese a todo, en la escala de distancias del Universo observable somos minúsculos. De hecho, somos mucho más pequeños comparados con esta escala que grandes comparados con las partículas. Por otro lado, con respecto a las dimensiones temporales, ¡el tiempo de una vida humana es casi tan largo como el del Universo! Es habitual hablar de la naturaleza efímera de la existencia pero, cuando se considera la duración de una vida humana tal como se muestra en el diagrama, puede comprobarse que esta no es en absoluto breve: ¡vivimos más o menos tanto como el propio Universo! Por supuesto, esto solo es cierto mirando a *escala logarítmica*, pero eso es lo más práctico cuando estamos interesados en rangos tan enormes. Para decirlo de otro modo, el número de veces que la duración de una vida humana iguala la edad del Universo es muchísimo menor que el número de tiempos de Planck, o incluso de vidas medias de las partículas de vida más corta, que equivalen a una vida humana. Así pues, somos estructuras muy estables en el Universo. Por lo que se refiere a los tamaños espaciales, estamos prácticamente en el medio: no experimentamos directamente ni la física de lo muy grande ni la de lo muy pequeño. Estamos exactamente entre ambas. De hecho, en una escala logarítmica, todos los seres vivos, desde las simples células hasta los árboles y las ballenas, están aproximadamente en el mismo tamaño intermedio.

¿Qué tipo de física se aplica en estas escalas diferentes? Permítanme introducir el diagrama que resume el conjunto de la física (figura 1.5). Por supuesto, he tenido que dejar fuera algunos detalles, como son ¡todas las ecuaciones! Pero están indicadas las teorías básicas esenciales que utilizan los físicos.

El punto clave es que, en física, utilizamos dos tipos de procedimientos muy diferentes. Para describir el comportamiento a pequeña escala utilizamos la mecánica cuántica, que he descrito como el nivel cuántico en la figura 1.5. Se acusa a la mecánica cuántica de ser poco nítida e indeterminista, pero esto no es cierto. Mientras permanezcamos en este nivel, la

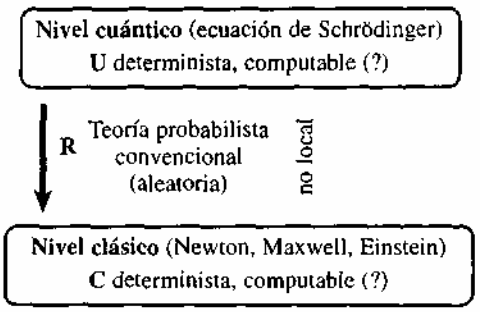


Figura 1.5.

teoría cuántica es determinista y precisa. En su forma más familiar, la mecánica cuántica implica el uso de la ecuación conocida como *ecuación de Schrödinger,* que rige el comportamiento del estado físico de un sistema cuántico -lo que se denomina su *estado cuántico- y* esta es una ecuación determinista. He utilizado la letra U para describir esta actividad en el nivel cuántico. La indeterminación en mecánica cuántica aparece solo cuando realizamos lo que se denomina *hacer una medida y* eso implica amplificar un suceso desde el nivel cuántico al nivel clásico. Posteriormente, en el capítulo 2, ampliaré estas ideas.

A gran escala utilizamos la física clásica, que es completamente determinista; estas leyes clásicas incluyen las leyes del movimiento de Newton; las leyes de Maxwell para el campo electromagnético, que incorporan la electricidad, el magnetismo y la luz; y las teorías de la relatividad de Einstein, la *teoría especial,* que trata con velocidades grandes, y la *teoría general,* que trata con campos gravitatorios grandes. Estas leyes se aplican de forma muy exacta a gran escala.

Solo como una nota a la figura 1.5, se puede ver que he incluido un comentario sobre *computabilidad* en física cuántica y clásica. Esto no tiene relevancia en este capítulo ni en el capítulo 2, pero sí la tendrá en el capítulo 3, y allí volveré a la cuestión de la computabilidad.

En el resto de este capítulo me interesaré principalmente por la teoría de la relatividad de Einstein: en concreto, por el funcionamiento de la teoría, su extraordinaria precisión y algo sobre su elegancia como teoría física. Pero consideremos primero la teoría newtoniana. La física de Newton permite, como la relatividad, utilizar una descripción espacio-temporal. Esta fue formulada de forma precisa por primera vez por Cartan para la teoría de la gravedad de Newton, algún tiempo después de que Einstein hubiera presentado su teoría general de la relatividad. La física de Galileo y Newton se representa en un espacio-tiempo para el que existe una coordenada de tiempo global, que aquí se muestra dirigida hacia la parte superior del diagrama (figura 1.6). Para cada valor constante del tiempo, existe una sección espacial que es un espacio tridimensional euclidiano, aquí representado como un plano horizontal. Una característica

esencial de la imagen del espacio-tiempo newtoniano es que estos cortes espaciales, a través del diagrama, representan momentos de simultaneidad.

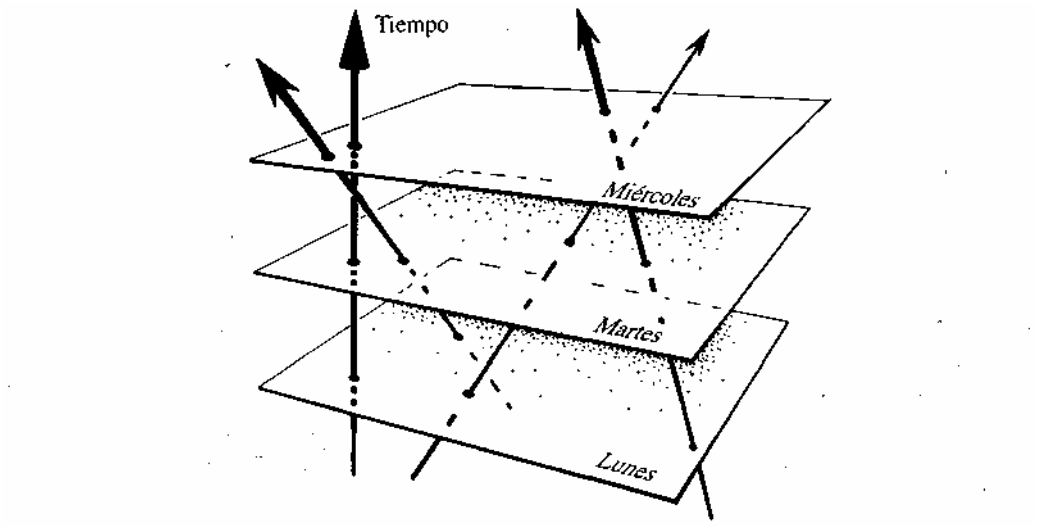


Figura 1.6. Espacio-tiempo galileano: las partículas en movimiento uniforme se representan por líneas rectas.

Así pues, todo lo que ocurre el lunes a mediodía yace en un corte horizontal a través del diagrama espacio-temporal; todo lo que sucede el martes a mediodía yace en el siguiente corte mostrado en el diagrama, y así sucesivamente. El tiempo cruza el diagrama espacio-temporal y las secciones euclidianas se suceden una tras otra a medida que el tiempo avanza. Todos los observadores, independientemente de cómo se muevan a través del espacio-tiempo, pueden ponerse de acuerdo acerca del instante en el que ocurren los sucesos, porque todos utilizan los mismos cortes temporales para medir el paso del tiempo.

En la teoría especial de la relatividad de Einstein tenemos que adoptar una imagen diferente. En dicha teoría, la imagen espacio-temporal es absolutamente esencial: la diferencia clave es que el tiempo ya no tiene el carácter universal que posee en la teoría newtoniana. Para apreciar la diferencia entre ambas teorías es necesario entender una parte esencial de la teoría de la relatividad, a saber, aquellas estructuras conocidas como *conos de luz.*

¿Qué es un cono de luz? Un cono de luz está dibujado en la figura 1.7. Imaginemos un destello de luz que tiene lugar en cualquier punto y en algún instante -es decir, en un *suceso* en el espacio-tiempo- y las ondas luminosas que viajan hacia fuera a partir de este suceso, la fuente del destello, a la velocidad de la luz. En una imagen puramente espacial (figura l.7.b), podemos representar las trayectorias de las ondas luminosas en el espacio como una esfera que se expande a la velocidad de la luz. Ahora podemos traducir este movimiento de las ondas luminosas en un diagrama espacio-temporal (figura 1.7.a) en el que el tiempo corre hacia la parte superior del diagrama y las coordenadas espaciales se refieren a desplazamientos horizontales, lo mismo que sucedía en la situación newtoniana de la figura 1.6. Por desgracia, en la imagen espacio-temporal completa (figura 1.7.a) solo podemos representar dos dimensiones espaciales horizontalmente en el diagrama, porque el espacio-tiempo de nuestra imagen es solo tridimensional. Ahora vemos que el destello está representado por un punto (suceso) en el origen y que las trayectorias subsiguientes de los rayos luminosos (ondas) intersectan los planos *espaciales* horizontales en círculos cuyos radios aumentan hacia la parte superior del diagrama a la velocidad de la luz. Puede verse que las trayectorias de los rayos luminosos forman conos en el diagrama espacio-temporal. El cono de luz representa así la historia de este destello luminoso: la luz se propaga desde el origen a lo largo del cono de luz, lo que significa a la velocidad de la luz, hacia el futuro. Los rayos luminosos también pueden llegar al origen a lo largo del cono de luz del pasado -esa parte del cono de luz se conoce como el *cono de luz pasado* y toda información llevada al observador por las ondas luminosas llega al origen a lo largo de este cono-.

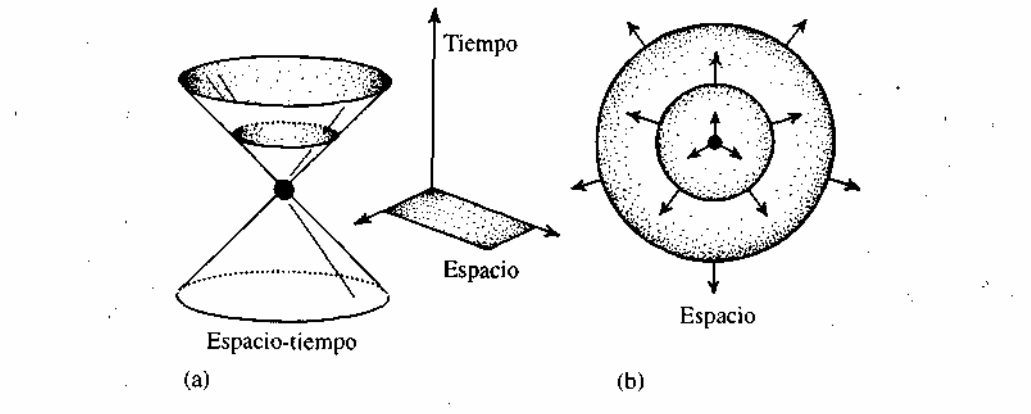


Figura 1.7. Representación de la historia de un destello luminoso en términos de su propagación en (*a*) el espacio-tiempo y (*b*) el espacio.

Los conos de luz describen las estructuras más importantes del espacio-tiempo. En particular, representan los límites de la influencia causal. La historia de una partícula en el espacio-tiempo se representa por una línea que viaja hacia arriba en el diagrama espacio-temporal, y esta línea tiene que estar dentro del cono de luz (figura 1.8). Esta es simplemente otra forma de decir que una partícula material no puede viajar más rápido que la luz. Ninguna señal puede viajar desde el interior al exterior del cono de luz futuro y, por eso, el cono de luz representa los límites de la causalidad.

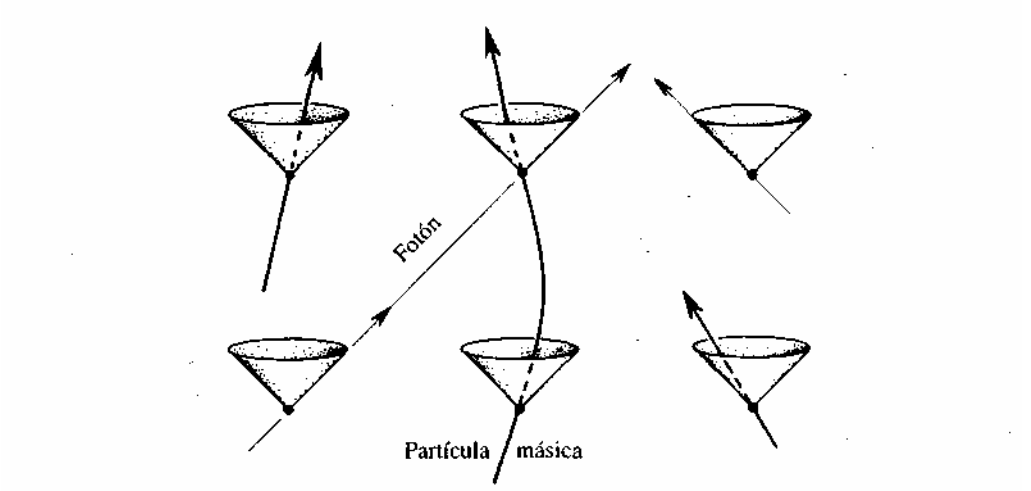


Figura 1.8. Ilustración del movimiento de una partícula en el espacio-tiempo de la relatividad especial, que se conoce como espacio-tiempo de Minkowski o geometría de Minkowski. Los conos de luz en los diferentes puntos del espacio-tiempo están alineados y las partículas solo pueden viajar en el interior de sus conos de luz futuros.

Existen algunas propiedades geométricas significativas en relación con los conos de luz. Consideremos dos observadores que se mueven a velocidades diferentes en el espacio-tiempo. A diferencia del caso de la teoría newtoniana, donde los planos de simultaneidad son los mismos para todos los observadores, no hay simultaneidad absoluta en la relatividad. Los observadores que se mueven a velocidades diferentes dibujan sus propios planos de simultaneidad como secciones diferentes en el espacio-tiempo, como se ilustra en la figura 1.9. Hay una forma perfectamente definida de pasar desde un plano a otro mediante lo que se conoce como una *transformación de Lorentz*, y estas transformaciones constituyen lo que se denomina el *grupo de Lorentz.* El descubrimiento de este grupo fue un ingrediente esencial en el hallazgo de la teoría especial de la relatividad de Einstein. El grupo de Lorentz puede entenderse como un grupo de transformaciones (lineales) espacio-temporales que dejan invariante un cono de luz.

También podemos apreciar el grupo de Lorentz desde un punto de vista ligeramente diferente. Como he resaltado, los conos de luz son las estructuras fundamentales del espacio-tiempo. Imaginen un observador situado en algún lugar en el espacio, mirando al Universo. Lo que ese observador ve son los

rayos de luz que proceden de las estrellas y entran en sus ojos. Según el punto de vista espacio-temporal, los sucesos que observa son las intersecciones de las *líneas-de-universo* de las estrellas con su cono de luz pasado, como se ilustra en la figura 1.10.a. Observen a lo largo de su cono de luz pasado

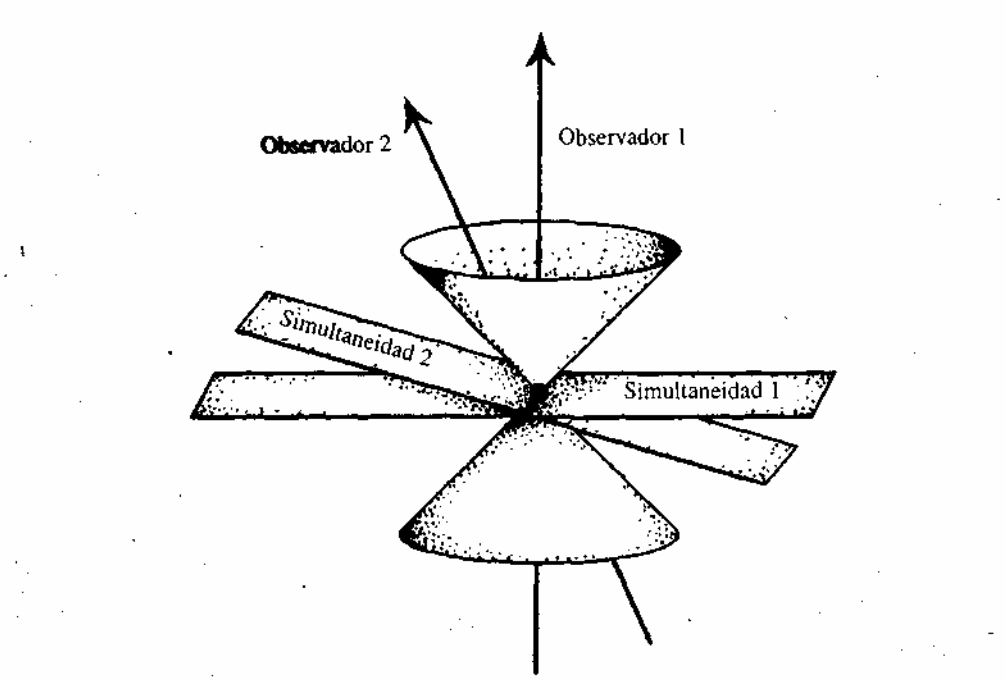


Figura 1.9. Ilustración de la relatividad de la simultaneidad según la teoría especial de la relatividad de Einstein. Los observadores 1 y 2 están en movimiento uno con respecto al otro en el espacio-tiempo. Sucesos que son simultáneos para el observador 1 no son simultáneos para el observador 2 y *viceversa*.

las posiciones de las estrellas en puntos concretos. Dichos puntos parecen estar situados en la esfera celeste que le rodea. Situemos ahora otro observador, que se mueve a alta velocidad con respecto al primero, y que pasa muy cerca de él en el momento en que ambos miran al cielo. Este segundo observador percibe las mismas estrellas que el primero, pero encuentra que están situadas en posiciones diferentes en la esfera celeste (figura l.l0.b) -este es el efecto conocido como *aberración-*. Existe un conjunto de transformaciones que nos permite calcular la relación entre lo que cada uno de estos observadores ve en su esfera celeste. Individualmente, estas transformaciones hacen corresponder una esfera a otra esfera. Pero es una transformación de un tipo muy especial. Transforma círculos exactos en círculos exactos y conserva los ángulos. Así, si una figura en el cielo parece un círculo para el primer observador, entonces también debe parecer un círculo para el otro observador.

Existe una manera muy bella de describir cómo funciona esto y la ilustro para mostrar que hay una elegancia particular en las matemáticas que suelen subyacer a la física en el nivel más fundamental. La figura 1.10.c muestra una esfera con un plano dibujado que pasa por su ecuador. Podemos dibujar figuras en la superficie de la esfera

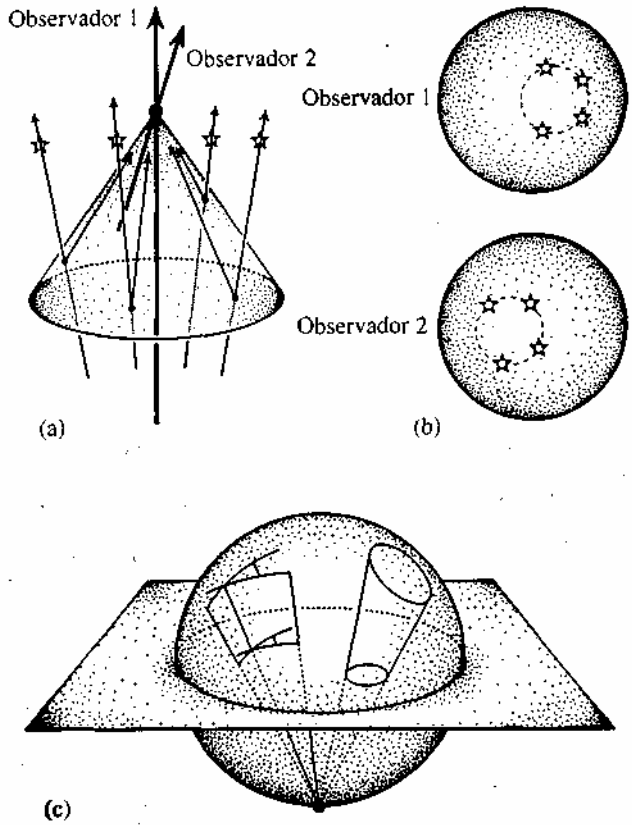


Figura 1.10. Ilustración de las observaciones celestes hechas por los observadores 1 y 2. (*a*) Los observadores 1 y 2 ven estrellas a lo largo del cono de luz pasado. Los puntos donde las estrellas cortan el cono de luz se indican por puntos negros. Las señales luminosas se propagan desde las estrellas hasta los observadores a lo largo del cono de luz tal como se ilustra. El observador 2 se mueve en el espacio-tiempo a cierta velocidad con respecto al observador 1. (*b*) Ilustración de la posición de las estrellas en el cielo tal como las ven el observador 1 y el observador 2 cuando coinciden en un mismo punto del espacio-tiempo. (*c*) Una buena manera de representar la transformación del cielo de un observador a otro es mediante una proyección estereográfica: los círculos se transforman en círculos, y los ángulos se conservan.

y luego examinar cómo se proyectan en el plano ecuatorial desde el polo sur, tal como se ilustra. Este tipo de proyección se conoce como proyección estereográfica y tiene algunas propiedades bastante extraordinarias. Círculos en la esfera se proyectan en círculos exactos en el plano, y los ángulos que forman entre sí curvas en la esfera se proyectan en ángulos exactamente iguales en el plano. Como discutiré más extensamente en el capítulo 2 (compárese con la figura 2.4), esta proyección nos permite etiquetar los puntos de la esfera mediante números complejos (números que incluyen la raíz cuadrada de -1), números que también se utilizan para etiquetar los puntos del plano ecuatorial, junto con el *infinito*, para conferir a la esfera la estructura conocida como *esfera de Riemann.*

Para aquellos que estén interesados, la transformación de aberración es:

$$u \to u' = \frac{\alpha u + \beta}{\gamma u + \delta}$$

y, como es bien conocido para los matemáticos, esta función transforma círculos en círculos y conserva los ángulos. Las funciones de este tipo se conocen como *transformaciones de Möbius.* Para nuestros propósitos presentes, simplemente necesitamos señalar la admirable sencillez de la fórmula (de aberración) de Lorentz cuando se escribe en términos de un parámetro complejo *u.*

Un punto sorprendente acerca de esta forma de considerar dichas transformaciones es que, según la teoría especial de la relatividad, la fórmula es muy sencilla, mientras que, si expresáramos la correspondiente transformación de aberración según la mecánica newtoniana, la fórmula sería mucho más complicada. A menudo resulta que, cuando descendemos a los fundamentos y desarrollamos una teoría más exacta, las matemáticas resultan ser más sencillas, incluso si la apariencia formal es más complicada en primera instancia. Este punto importante queda ejemplificado en el contraste entre la teoría de la relatividad de Galileo y la de Einstein.

Así pues, en la teoría especial de la relatividad tenemos una teoría que, en muchos aspectos, es más sencilla que la mecánica newtoniana. Desde el punto de vista de las matemáticas, y en particular desde el punto de vista de la teoría de grupos, es una estructura mucho más amable. En la teoría especial de la relatividad, el espacio-tiempo es plano y todos los conos de luz están alineados regularmente, tal como se ilustra en la figura 1.8. Ahora bien, si damos un paso más allá hasta la teoría general de la relatividad de Einstein, es decir, la teoría del espacio-tiempo que tiene en cuenta la gravedad, la imagen parece a primera vista bastante retorcida: los conos de luz son un caos (figura 1.11). Con todo, he venido señalando que, a medida que desarrollamos teorías cada vez más evolucionadas, las matemáticas se hacen más simples. Pero veamos lo que ha sucedido aquí: yo tenía una porción muy elegante de las matemáticas que se ha hecho horriblemente complicada. Este tipo de cosas sucede, ustedes tendrán que seguir conmigo algún tiempo hasta que reaparezca la simplicidad.

Permítanme recordarles los ingredientes fundamentales de la teoría de la gravedad de Einstein. Un ingrediente básico se denomina *principio de equivalencia de Galileo.* En la figura 1.12.a muestro a Galileo inclinado en lo alto de la Torre de Pisa y dejando caer piedras grandes y pequeñas. Ya hiciera o no realmente este experimento, Galileo comprendió muy bien que, si se ignoran los efectos de la resistencia del aire, las dos piedras llegarán al suelo en el mismo instante. Si uno de ustedes estuviera sentado en

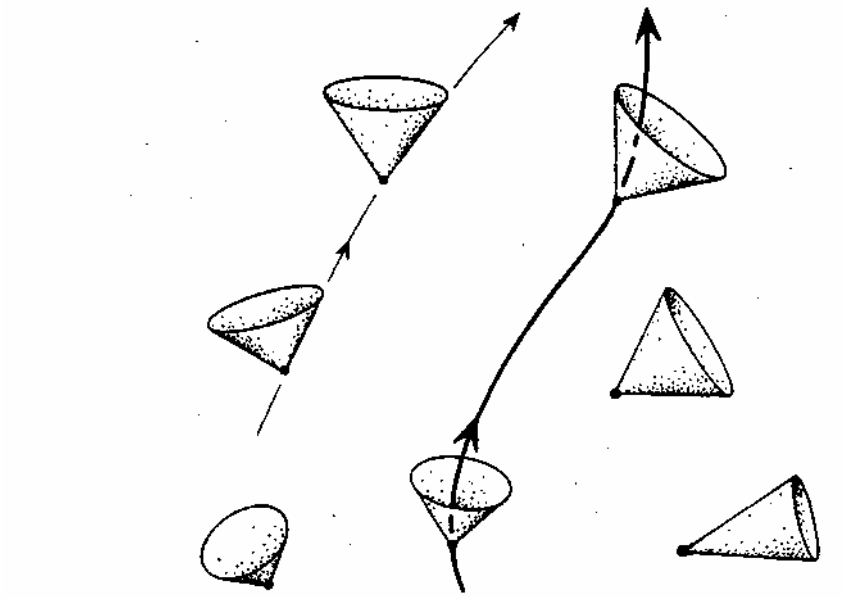


Figura 1.11. Una imagen del espacio-tiempo curvo.

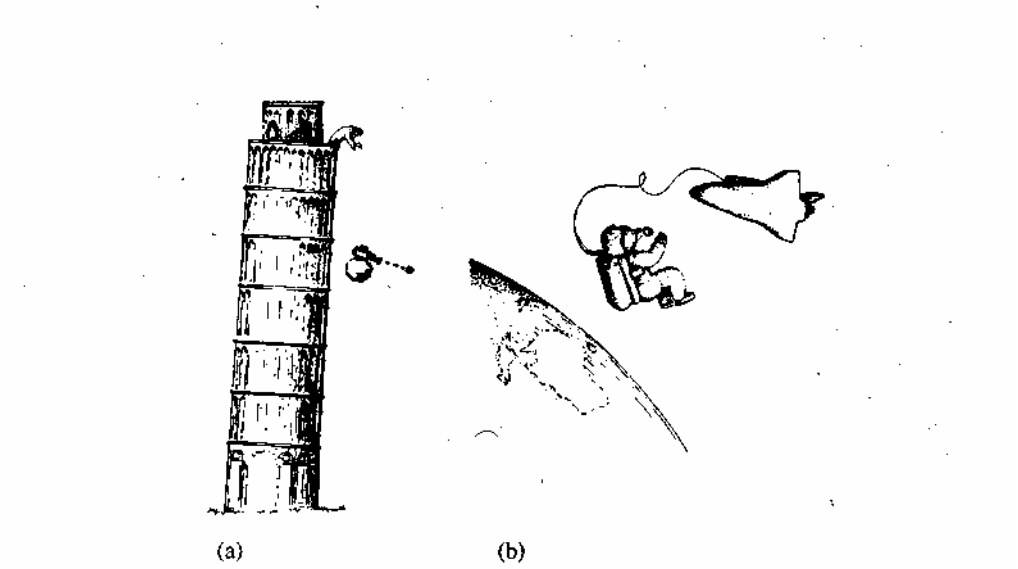


Figura 1.12. (*a*) Galileo arrojando dos piedras (y una videocámara) desde la Torre inclinada de Pisa. (*b*) El astronauta observa que el vehículo espacial se mantiene frente a él, aparentemente sin ser afectado por la gravedad.

una de estas piedras mirando a la otra mientras caen juntas, observaría que la otra piedra se mantiene frente a él (he situado una videocámara fijada a una de las piedras para realizar la observación). Hoy día, con los viajes espaciales, este es un fenómeno muy familiar -recientemente hemos visto un astronauta de origen inglés paseando por el espacio- y, como sucede con la piedra grande y la piedra pequeña, la nave espacial se mantiene junto al astronauta -este es exactamente el mismo fenómeno que el principio de equivalencia de Galileo-.

Así, si se considera la gravedad de la forma correcta, es decir, en un sistema de referencia en caída, esta parece desaparecer directamente ante nuestros ojos. Esto es, de hecho, correcto; pero la teoría de Einstein *no* nos dice que la gravedad desaparezca: solo nos dice que la *fuerza* de la gravedad desaparece. Queda algo, y ese algo es el efecto de marea de la gravedad.

Permítanme introducir algunos elementos matemáticos más, pero no muchos. Tenemos que describir la curvatura del espacio-tiempo y esto se consigue mediante un objeto conocido como un *tensor* que yo he denominado **Riemann** en la siguiente ecuación. En realidad se denomina *tensor de curvatura de Riemann*, pero yo no les voy a decir qué es, solo que está representado por una R mayúscula con varios índices adheridos a su pie, que se indican por los puntos. El tensor de curvatura de Riemann consta de dos partes. Una de las partes se denomina *curvatura de Weyl* y la otra, *curvatura de Ricci*, y tenemos la ecuación (esquemática):

$$\text{Riemann} = \text{Weyl} + \text{Ricci}$$

$$R.... = C....+ R..g..$$

Formalmente, C.... y R.. son los tensores de curvatura de Weyl y Ricci, respectivamente, y g.. es el tensor métrico.

La curvatura de Weyl mide eficazmente el efecto de marea. ¿Qué es el efecto *de marea?* Recordemos que, desde el punto de vista del astronauta, parece como si la gravedad hubiera sido abolida, aunque esto no es del todo cierto. Imaginemos que el astronauta está rodeado por una esfera de partículas que están inicialmente en reposo con respecto a él. En un primer momento, las partículas se mantienen aproximadamente en sus posiciones pero pronto empezarán a acelerarse debido a las ligeras diferencias en la atracción gravitatoria que ejerce la Tierra en los diferentes puntos de la esfera. (Nótese que estoy describiendo el efecto en lenguaje newtoniano, pero por ahora es suficiente.) Estas ligeras diferencias hacen que la esfera original de partículas se distorsione hasta mostrar una configuración elíptica, como se ilustra en la figura 1.13.a.

Esta distorsión se debe en parte a la atracción ligeramente mayor que ejerce la Tierra sobre aquellas partículas más próximas a ella y la menor atracción que ejerce sobre aquellas que están más alejadas, y en parte también a que, en los lados izquierdo y derecho de la esfera, la atracción de la Tierra actúa ligeramente hacia dentro. Esto

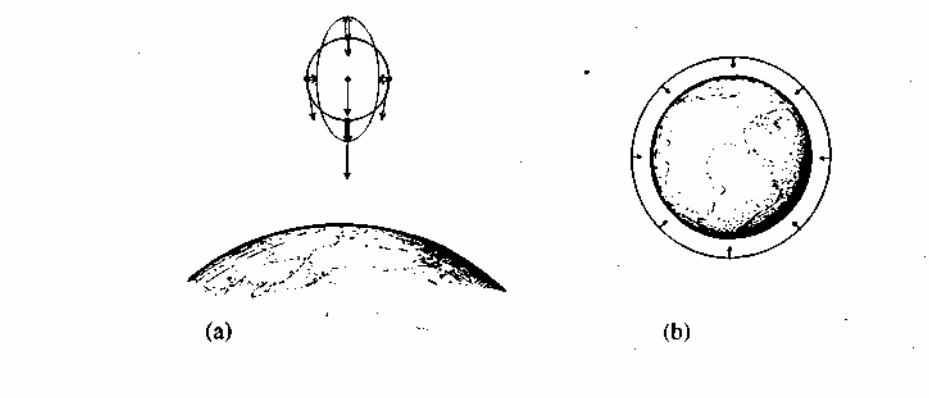


Figura 1.13. (*a*) El efecto de marea. Las flechas dobles muestran la aceleración relativa. (*b*) Cuando la esfera rodea un cuerpo material (aquí rodea la Tierra) hay una aceleración neta hacia dentro.

hace que la esfera se distorsione y se convierta en un elipsoide. Se denomina *efecto de marea* por la muy buena razón de que si se reemplaza la Tierra por la Luna y la esfera de partículas por la Tierra con sus océanos, entonces la Luna produce el mismo efecto gravitacional sobre la superficie de los océanos que el que la Tierra produce en la esfera de partículas: la superficie del mar más próxima a la Luna es atraída hacia ella, mientras que la que está en el lado opuesto de la Tierra es, de hecho, empujada lejos de ella. El efecto hace que la superficie del mar se abombe en los lados próximo y lejano de la Tierra y es la causa de las dos mareas altas que ocurren cada día.

Los efectos de la gravedad, desde el punto de vista de Einstein, consisten simplemente en este efecto de marea. Está definido esencialmente por la curvatura de Weyl, es decir, la parte C.... de la curvatura de Riemann. Esta parte del tensor de curvatura conserva el volumen: es decir, si se calculan las aceleraciones iniciales de las partículas de la esfera, el volumen de la esfera y el volumen del elipsoide en el que se transforma son inicialmente iguales.

La parte restante de la curvatura se conoce como *curvatura de Ricci* y tiene un efecto reductor del volumen. A partir de la figura 1.13.b puede verse que, si en lugar de estar en la parte inferior del diagrama, la Tierra estuviera dentro de la esfera de partículas, el volumen de la esfera de partículas se reduciría a medida que las partículas se aceleraran hacia dentro. La cantidad de esta reducción de volumen es una medida de la curvatura de Ricci. La teoría de Einstein nos dice que esta curvatura en un punto del espacio está determinada por la cantidad de materia presente dentro de una pequeña esfera centrada aproximadamente en dicho punto. En otras palabras, la densidad de materia, adecuadamente definida, en un punto del espacio nos dice cómo se aceleran las partículas hacia dicho punto. La teoría de Einstein es casi la misma que la de Newton cuando se expresa de este modo.

Así es como Einstein formuló su teoría de la gravedad: se expresa en términos de los efectos de marea, que son medidas de la curvatura espacio-temporal local. Es crucial que tengamos que pensar en términos de la curvatura de un espacio-tiempo tetra-dimensional. Esto se mostraba esquemáticamente en la figura 1.11: consideramos las líneas que representan las líneas-de-universo de partículas y las formas en que estas trayectorias son distorsionadas como una medida de la curvatura del espacio-tiempo. Así pues, la teoría de Einstein es esencialmente una teoría geométrica del espacio-tiempo tetradimensional -es una teoría de extraordinaria belleza desde una perspectiva matemática-.

La historia del descubrimiento por Einstein de la teoría general de la relatividad contiene una moraleja importante. Fue formulada en su totalidad por primera vez en 1915. No estaba motivada por ninguna necesidad de describir o explicar ciertas observaciones sino por diversos desiderata estéticos, geométricos y físicos. Los ingredientes clave eran el principio de equivalencia de Galileo, ejemplificado por la caída de piedras de diferentes masas (figura 1.12), y las ideas de la geometría no euclidiana, que es el lenguaje natural para describir la curvatura del espacio-tiempo. No existía un trabajo de observación en 1915. Una vez que la relatividad general estuvo formulada en su forma final, se advirtió que había tres pruebas derivadas de la observación que debían apoyar la teoría. El perihelio de la órbita de Mercurio avanza, o gira, de una forma que no podía ser explicada por la influencia gravitatoria newtoniana de los

otros planetas: la teoría general de la relatividad predice exactamente el avance observado. Las trayectorias de los rayos luminosos son curvadas por el Sol, y esta fue la razón de la famosa expedición para observar el eclipse de 1919, dirigida por Arthur Eddington y que encontró un resultado coherente con la predicción de Einstein (figura 1.14.a). La tercera prueba era la predicción de que los relojes se hacen más lentos en un potencial gravitatorio: es decir, un reloj próximo al suelo atrasa con respecto a un reloj situado en lo alto de una torre. Este efecto también ha sido medido experimentalmente. Estas nunca fueron, sin embargo, pruebas muy impresionantes: los efectos eran siempre muy pequeños y varias teorías diferentes podrían haber dado los mismos resultados.

Ahora la situación ha cambiado drásticamente: en 1993, Hulse y Taylor recibieron el Premio Nobel por una extraordinaria serie de observaciones. La figura 1.15.a muestra el pulsar binario conocido como PSR 1913+16: consiste en un par de estrellas de neutrones, cada una de las cuales, de enorme densidad, tiene una masa aproximadamente igual a la del Sol pero solo unos pocos kilómetros de diámetro. Las estrellas de neutrones describen órbitas extremadamente elípticas alrededor de su centro de gravedad común. Una de ellas genera un campo magnético muy intenso que hace que las partículas que son arrastradas a su alrededor emitan una radiación intensa que viaja hasta la Tierra, a unos 30.000 años-luz de distancia, donde es observada como una serie de pulsos bien definidos. Se han realizado todo tipo de observaciones muy precisas de los tiempos de llegada de dichos pulsos. En particular, han podido calcularse todas las propiedades de las órbitas de las dos estrellas de neutrones así como todas las correcciones minúsculas debidas a la teoría general de la relatividad.

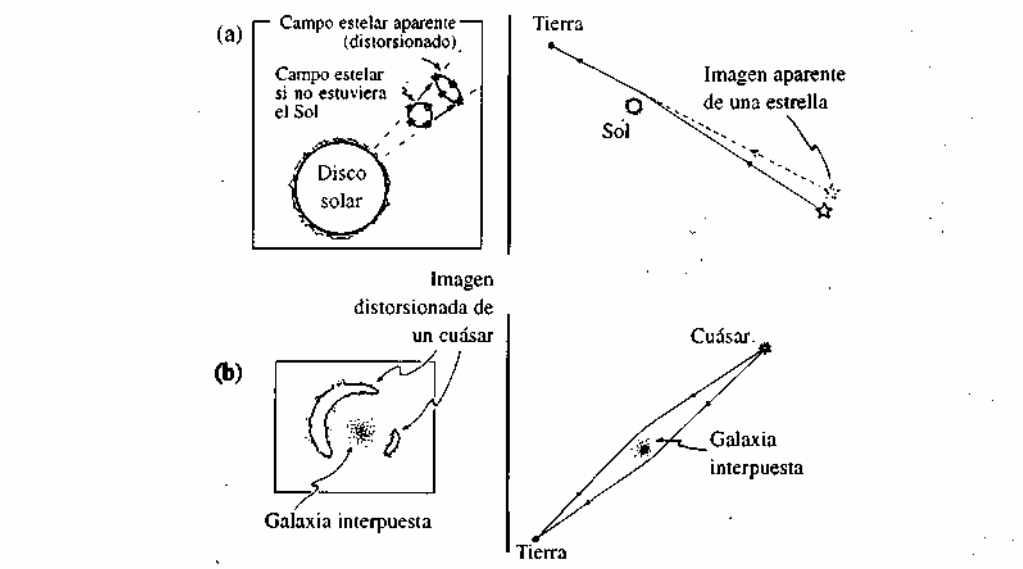
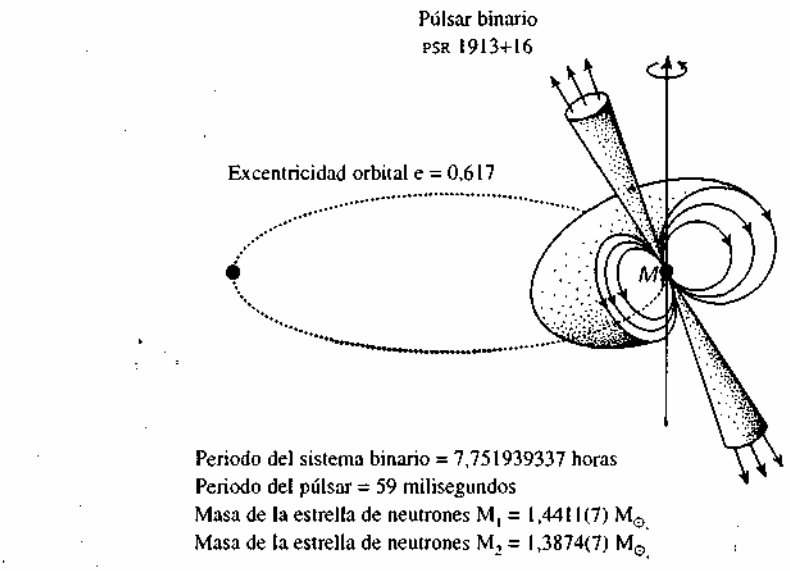


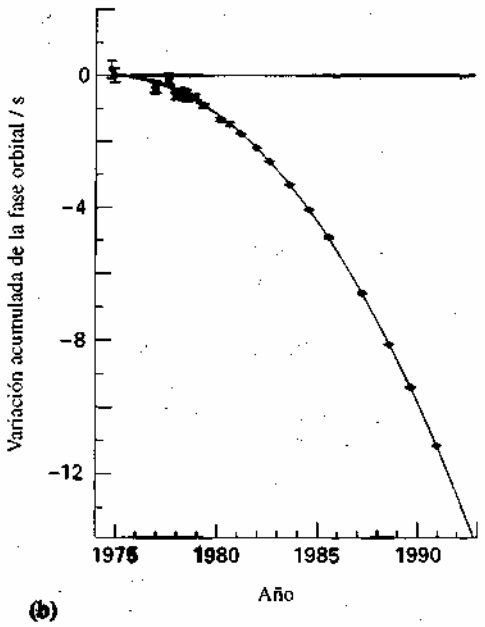
Figura 1.14. (*a*) Efectos observacionales directos de la gravedad sobre la luz según la relatividad general. La curvatura de Weyl del espacio-tiempo se manifiesta como una distorsión del campo estelar lejano, en este caso debida al efecto de curvatura de la luz en el campo gravitatorio del Sol. Una figura circular de estrellas se distorsionaría en una figura elíptica. (*b*) El efecto de curvatura de la luz de Einstein es ahora una herramienta importante en la astronomía observacional. La masa de la galaxia interpuesta puede ser estimada a partir de la medida de la distorsión que provoca en la imagen de un cuásar lejano.

Existe, además, una característica que es absolutamente exclusiva de la relatividad general, y que no está presente en ningún caso en la teoría newtoniana de la gravedad: los objetos que están en órbita uno alrededor del otro irradian energía en forma de ondas gravitatorias. Estas son similares a ondas de luz, aunque son ondulaciones en el espacio-tiempo en lugar de ser ondulaciones en el campo electromagnético. Estas ondas absorben energía del sistema a un ritmo que puede ser exactamente calculado según la teoría de Einstein, y este ritmo de pérdida de energía del sistema de estrellas de neutrones binarias coincide muy exactamente con las observaciones, como se ilustra en la figura 1.15.b, que muestra la aceleración del período orbital de las estrellas de neutrones, medida durante 20 años de observación. Estas señales pueden cronometrarse de forma tan precisa que, al cabo de 20 años, se sabe que la teoría es correcta con una precisión equivalente a una parte en 10^{14}. Esto hace de la relatividad general la teoría científica comprobada con más exactitud.

Hay una moraleja en esta historia: las motivaciones de Einstein para dedicar ocho o más años de su vida al desarrollo de la teoría general de la relatividad no eran ni producto de la observación ni de la experimentación. A veces, se argumenta: «Los físicos buscan pautas en sus resultados experimentales y luego encuentran alguna bonita



(a)



(b)

Figura 1.15. (*a*) Representación esquemática del púlsar binario PSR 1913+16. Una de las estrellas de neutrones es un púlsar de radio. La señal de radio se emite a lo largo de los polos del dipolo magnético que no está alineado con el eje de rotación de la estrella de neutrones. Se observan pulsos muy precisos cuando el estrecho haz de radiación barre la visual del observador. Las propiedades de las dos estrellas de neutrones han sido obtenidas a partir de un cronometraje muy preciso de los tiempos de llegada de los pulsos utilizando (y verificando) efectos que solo están presentes en la teoría general de la relatividad de Einstein. (*b*) El cambio de la fase de los tiempos de llegada de los pulsos procedentes del púlsar binario PSR 1913+16, comparado con el cambio esperado debido a la emisión de radiación gravitatoria por el sistema binario de estrellas de neutrones (línea continua).

teoría que coincide con ellas. Quizá esto explica por qué las matemáticas y la física trabajan tan bien juntas».

Pero, en este caso, las cosas no fueron así. La teoría fue desarrollada originalmente sin tener que responder a ninguna observación: la teoría matemática es muy elegante y está físicamente muy bien motivada. La cuestión es que la estructura matemática está precisamente allí, en la naturaleza, y la teoría existe realmente allí, en el espacio: no ha sido impuesta a la naturaleza por nadie. Este es uno de los puntos esenciales de este capítulo. Einstein revelaba algo que ya estaba presente. Más aún, no era simplemente

algún elemento menor de física lo que descubrió: es lo más fundamental que tenemos en nuestro Universo, la propia naturaleza del espacio y del tiempo.

He aquí un caso muy claro, remite de nuevo a mi diagrama original concerniente a la relación entre el mundo de las matemáticas y el mundo físico (figura 1.3). En la teoría general de la relatividad tenemos cierto tipo de estructura que realmente subyace al comportamiento del mundo físico de una forma extraordinariamente precisa. Estas características fundamentales de nuestro mundo no suelen descubrirse observando cómo se comporta la naturaleza, aunque esto es obviamente muy importante. Uno tiene que estar preparado para desechar teorías que pudieran atraer por muchas otras razones pero que no encajan con los hechos. En esta tenemos una teoría que sí encaja con los hechos con una precisión extraordinaria. La exactitud corresponde aproximadamente a un número de cifras doble que el que se obtiene en la teoría newtoniana; en otras palabras, se sabe que la teoría general de la relatividad es correcta hasta una parte en 10^{14}, mientras que la teoría newtoniana solo es exacta hasta una parte en 10^{7}. Esta mejora equivale al aumento que tuvo lugar entre el siglo XVII y nuestros días de la precisión con que se sabía que era correcta la teoría newtoniana. Newton sabía que su teoría era válida hasta, aproximadamente, una parte en 1.000, mientras que ahora se sabe que es exacta hasta una parte en 10^{7}.

La teoría de la relatividad general de Einstein es solo una teoría, por supuesto. ¿Qué sucede con la estructura del mundo real? Existen tres tipos de modelo estándar que se deducen de la teoría de Einstein y estos tipos están definidos por un parámetro, que es, de hecho, el denotado por *k* en la figura 1.16. Existe otro parámetro que aparece a veces en discusiones cosmológicas y que se conoce como la constante cosmológica. Einstein consideraba que la introducción de la constante cosmológica en las ecuaciones de la relatividad general había sido el mayor error de su vida, y por eso yo también la dejaré fuera. Si nos vemos obligados a traerla de nuevo..., bien, entonces tendremos que convivir con ella.

Suponiendo que la constante cosmológica es cero, los tres tipos de Universo que están descritos por la constante *k* se ilustran en la figura 1.16. En el diagrama, *k* toma los valores 1, 0 y -1, porque todas las demás propiedades de los modelos se han escrito en la escala conveniente. Una forma más apropiada hubiera sido hablar de la edad o la escala del Universo, y entonces tendríamos un parámetro continuo; pero, cualitativamente, los tres modelos diferentes pueden considerarse definidos por la curvatura de las secciones espaciales del Universo. Si las secciones espaciales son planas, tienen curvatura cero y $k = 0$ Figura 1.16.a). Si las secciones espaciales están curvadas positivamente, lo que significa

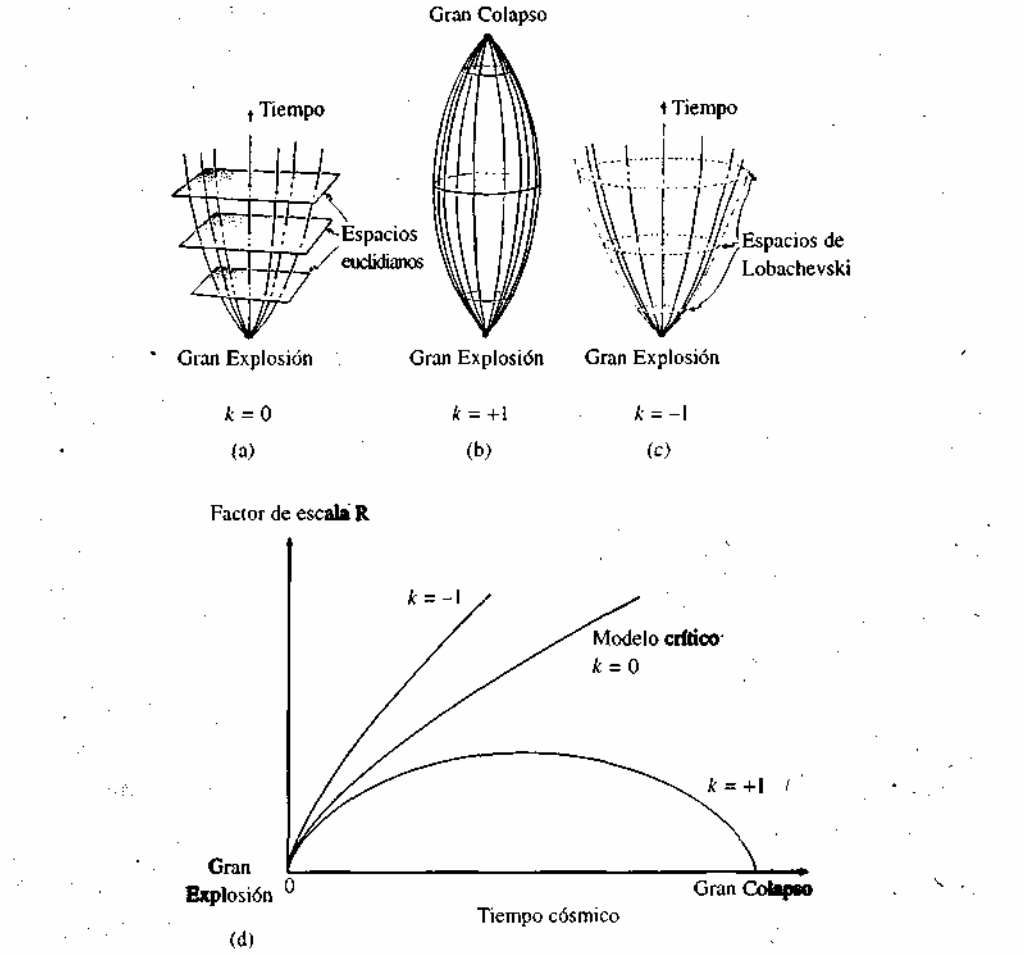


Figura 1.16. (*a*) Imagen espacio-temporal de un Universo en expansión con secciones espaciales euclidianas (se muestran dos dimensiones espaciales): $k = 0$. (*b*) Como en (*a*), pero para un Universo en expansión (y posterior contracción) con secciones espaciales esféricas: $k = +1$. (*c*) Como en (*a*) pero para un Universo en expansión con secciones espaciales de Lobachevski: $k = -1$. (*d*) Dinámica de los tres diferentes tipos de modelos de Friedman.

que el Universo se cierra sobre sí mismo, entonces $k = +1$ (figura 1.16.b). En todos estos modelos, el Universo tiene un estado inicial singular, la Gran Explosión, que marca su comienzo. Pero en el caso $k = +1$, el Universo se expande hasta un tamaño máximo y luego se comprime de nuevo hasta un Gran Colapso. Como alternativa, existe el caso $k = -1$, en el que el Universo se expande para siempre (figura 1.16.c). El caso $k = 0$ es la frontera límite entre los casos $k = 1$ y $k = -1$. He mostrado las relaciones radio-tiempo para estos tres tipos de Universo en la figura 1.16.d. El radio puede considerarse como cierta escala típica en el Universo, y puede verse que solo en el caso $k = +1$ se concentra hasta un Gran Colapso, mientras que, en los otros dos, el Universo se expande indefinidamente.

Quiero considerar el caso $k = -1$ con algo más de detalle: es quizá el más difícil de entender de los tres. Existen dos razones para estar interesado particularmente en esta situación. Una razón es que, si se toman al pie de la letra las observaciones existentes en este momento, es el modelo preferido. Según la teoría general de la relatividad, la curvatura del espacio está determinada por la cantidad de materia presente en el Universo y no parece haber suficiente para cerrar la geometría del Universo. Ahora bien, pudiera ser que haya un montón de materia oscura y oculta que todavía no conocemos. En este caso, el Universo podría ser uno de los otros modelos; pero si no hay una gran cantidad de materia extra, mucha más de la que creemos que debe estar presente dentro de las imágenes ópticas de las galaxias, entonces el Universo tendría $k = -1$. La otra razón es que, ¡este es el modelo que yo prefiero! Las propiedades de las geometrías $k = -1$ son particularmente elegantes.

¿Qué aspecto tienen los universos $k = -1$. Sus secciones espaciales tienen lo que se conoce como *geometría hiperbólica* o *de Lobachevski.* Para obtener una imagen de una geometría de Lobachevski lo

mejor es observar uno de los grabados de Escher. De su serie denominada *Límites circulares:* el *Límite circular IV* se muestra en la figura 1.17. Esta es la descripción del Universo de Escher: ¡comprueben que está llena de ángeles y demonios! Un aspecto a señalar es que parece como si las figuras se amontonasen hacia el borde del límite circular. Esto ocurre debido a que esta representación del espacio hiperbólico está dibujada en una hoja de papel plano ordinario; en otras palabras, en el espacio euclidiano. Lo que ustedes tienen que imaginar es que todos los demonios son en realidad de la misma forma y tamaño, de modo que, si ustedes vivieran en este Universo cerca del borde del diagrama, les parecerían exactamente iguales que los que están en el centro. Esta imagen proporciona alguna idea de lo que sucede en la geometría de Lobachevski: a medida que caminan desde el centro hacia el borde, ustedes tienen que imaginar que, debido al modo en que ha tenido que distorsionarse la imagen geométrica, la geometría real allí es exactamente la misma que en el centro, de modo que la geometría que les rodea sigue siendo la misma independientemente de cómo se muevan.

Este es quizá el ejemplo más sorprendente de una geometría bien definida. Pero la geometría euclidiana es, a su manera, igual de notable. La geometría euclidiana proporciona una maravillosa ilustración de la relación entre las matemáticas y la física. Esta geometría es una parte de las matemáticas pero los griegos la consideraban también una descripción del modo de ser del mundo. En realidad, resulta ser una descripción extraordinariamente precisa de la forma real del mundo: no absolutamente exacta, porque la teoría de Einstein nos dice que el espacio-tiempo está ligeramente curvado de diversas formas, pero, en cualquier caso, es una descripción del mundo extremadamente fiel. En otro tiempo, la gente se preocupaba por saber si otras geometrías eran posibles o no.

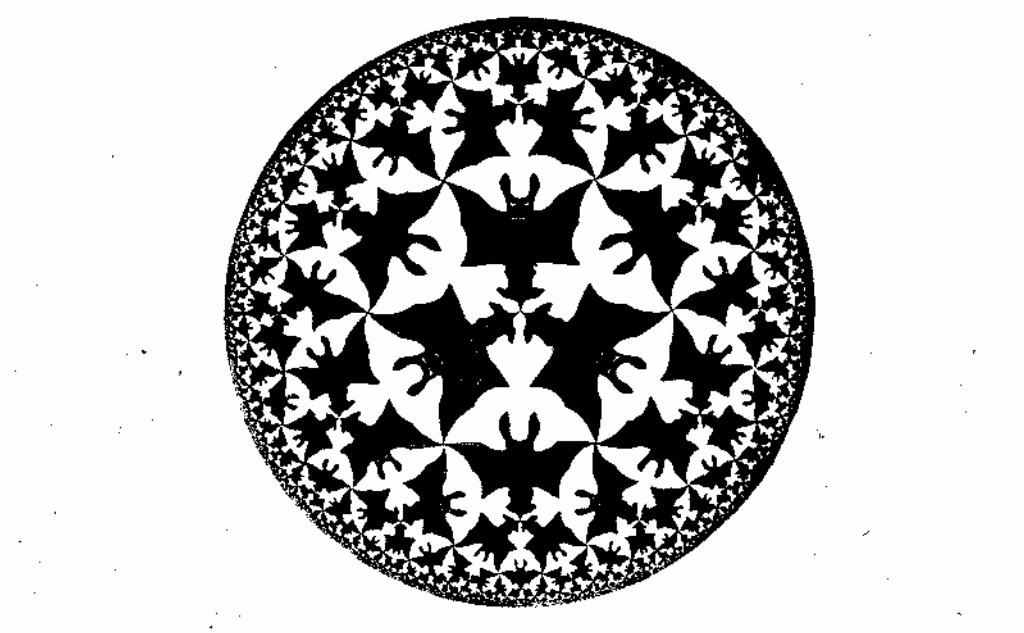


Figura 1.17. *Límite circular IV* de M.C. Escher. © 1998 Cordon Art-Baarn-Holland. Todos los derechos reservados. (Representación de un espacio de Lobachevski.)

En particular, se interesaban por lo que se conoce como *quinto postulado de Euclides,* que puede reformularse como el enunciado según el cual, si existe una línea recta en un plano y hay un punto exterior a esa recta, entonces existe una única paralela a dicha línea que pasa por ese mismo punto. Se pensaba que esto quizá podría demostrarse a partir de los otros axiomas más obvios de la geometría euclidiana; pero resulta que no es posible, y de ello surgió la noción de geometría no euclidiana.

En las geometrías no euclidianas, los ángulos de un triángulo no suman 180°. Este es otro ejemplo en el que uno piensa que las cosas van a hacerse más complicadas porque, en la geometría euclidiana, los ángulos de un triángulo sí suman 180° (figura 1.18.a). Pero luego se ve que, en la geometría no euclidiana, si se toma la suma de los ángulos de un triángulo y se resta de 180°, esta diferencia es proporcional al área del triángulo. En la geometría euclidiana, el área de un triángulo es algo complicada de escribir en términos de ángulos y longitudes. En la geometría no euclidiana de Lobachevski existe esta fórmula maravillosamente simple, debida a Lambert, que hace posible calcular el área del triángulo (figura 1.18.b).

De hecho, Lambert obtuvo su fórmula antes de que se descubriera la geometría no euclidiana, ¡y esto es algo que nunca he acabado de entender!

Hay aquí otro punto muy importante que se refiere a los números reales. Estos son absolutamente fundamentales para la geometría euclidiana. Fueron introducidos

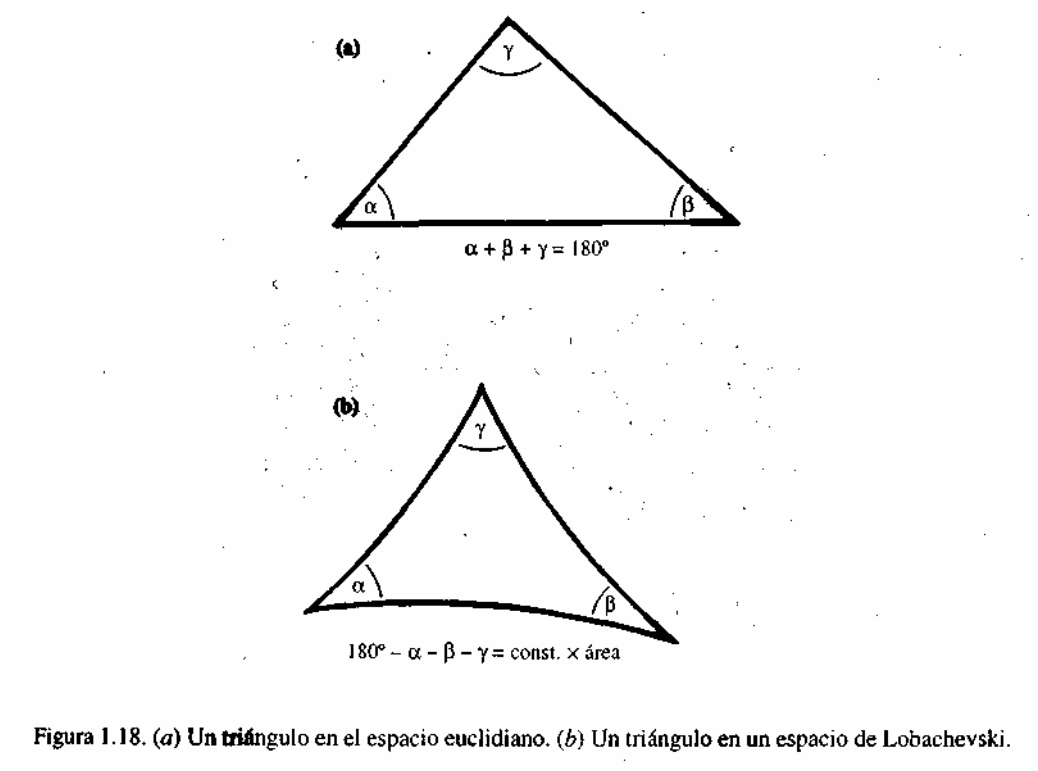


Figura 1.18. (*a*) **Un triángulo en el espacio euclidiano.** (*b*) **Un triángulo en un espacio de Lobachevski.**

esencialmente por Eudoxo en el siglo IV a. de C. y aún están con nosotros. Son los números que describen toda nuestra física. Como veremos más adelante, los números complejos también son necesarios, pero están basados en los números reales.

Examinemos otro de los grabados de Escher para ver cómo funciona la geometría de Lobachevski. La figura 1.19 es incluso más bonita que la figura 1.17 para entender esta geometría, porque las *líneas rectas* son más obvias, están representadas por arcos de círculo que cortan los límites en ángulos rectos. Por ello, si ustedes fueran *personas lobachevskianas y* vivieran en esta geometría, lo que ustedes considerarían una línea recta sería uno de estos arcos. Podemos verlas claramente en la figura 1.19: algunas son líneas rectas euclidianas que pasan por el centro del diagrama pero todas las demás son arcos curvos. Algunas de estas *líneas rectas* se muestran en la figura 1.20. En este diagrama he marcado un punto que no se encuentra en la línea recta (diámetro) que cruza el diagrama. La población lobachevskiana puede trazar dos (y más) líneas independientes paralelas al diámetro que pasan por dicho punto, tal como he indicado. Así pues, el postulado de las paralelas se viola en esta geometría. Además, ustedes pueden dibujar triángulos y calcular la suma de los ángulos de los triángulos para calcular sus áreas. Esto puede darles alguna idea de la naturaleza de la geometría hiperbólica.

Déjenme poner otro ejemplo. Dije que yo prefiero la geometría hiperbólica lobachevskiana. Una de las razones es que su grupo de simetrías es exactamente el mismo

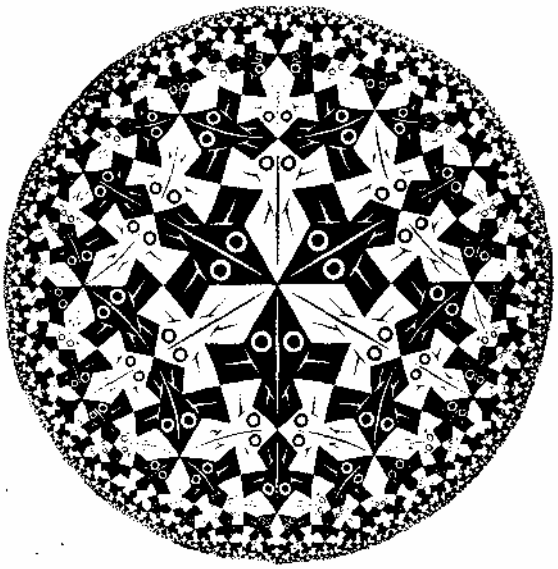


Figura 1.19. ***Límite circular I*** de M.C. Escher. © 1998 Cordon Art-Baarn-Holland. Todos los derechos reservados.

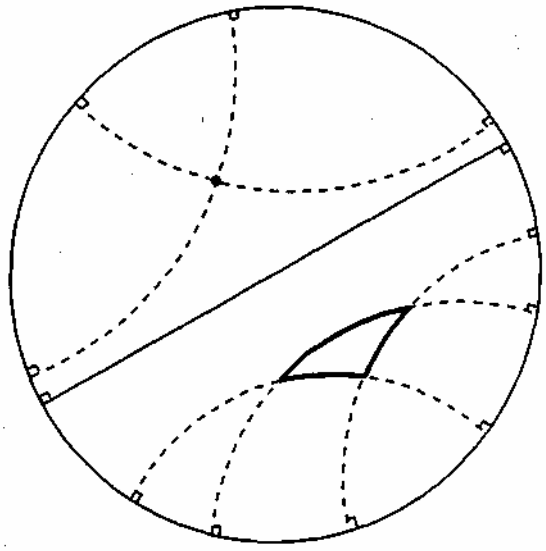


Figura 1.20. Aspectos de la geometría de un espacio de Lobachevski (hiperbólico) como el ilustrado por *Límite circular I*.

que el que ya hemos encontrado, a saber, el grupo de Lorentz, el grupo de la relatividad especial, o el grupo de simetría de los conos de luz de la relatividad. Para ver qué es esto, he dibujado un cono de luz en la figura 1.21 pero con algunas piezas extra. He tenido que suprimir una de las dimensiones espaciales para representarlo en el espacio tridimensional. El cono de luz está descrito por la ecuación usual mostrada en el diagrama:

$$t^2-x^2-y^2 = 0$$

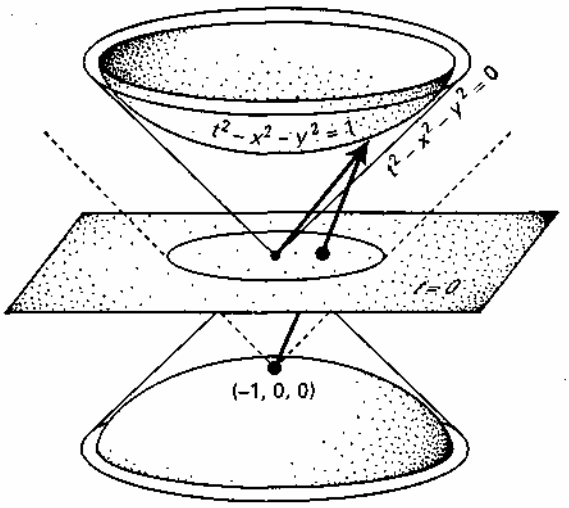


Figura 1.21. Espacio de Lobachevski insertado como una rama hiperboloide en el espacio-tiempo de Minkowski. La proyección estereográfica le hace corresponder el disco de Poincaré, cuya frontera es el círculo dibujado en el plano $t = 0$.

Las superficies en forma de cuenco que se muestran arriba y abajo están situadas a una *distancia unidad* del origen en esta geometría minkowskiana. *(Distancia* en la geometría minkowskiana es realmente *tiempo,* el tiempo propio que es medido físicamente por relojes en movimiento.) Así pues, estas superficies representan la superficie de una *esfera* para la geometría minkowskiana. Resulta que la geometría intrínseca de la *esfera* es realmente una geometría lobachevskiana (hiperbólica). Si consideramos una esfera ordinaria en el espacio euclidiano, podemos girarla y el grupo de simetrías es el grupo de rotaciones de la esfera. En la geometría de la figura 1.21, el grupo de simetrías es el asociado con la superficie mostrada en el diagrama; en otras Palabras, con el grupo de Lorentz de rotaciones. Este grupo de simetrías describe cómo se transforman el espacio y el tiempo cuando se mantiene fijo un punto concreto en el espacio-tiempo: rotando el espacio-tiempo de diferentes maneras. Vemos ahora, con esta representación, que el grupo de simetrías del espacio lobachevskiano es esencialmente idéntico al grupo de Lorentz.

La figura 1.21 ilustra una versión minkowskiana de la proyección estereográfica mostrada en la figura l.l0.c. El equivalente al polo sur es ahora el punto (-1, 0, 0), y proyectamos puntos de la superficie superior en forma de cuenco en la superficie plana en $t = 0$, que es la análoga al plano ecuatorial en la figura 1.10.c. Mediante este procedimiento, proyectamos todos los puntos de la superficie superior en el plano $t = 0$. Todos los puntos proyectados están dentro de un disco en el plano $t = 0$, que se suele denominar *disco de Poincaré.* Así es precisamente como surgen los diagramas del límite circular de Escher: la superficie hiperbólica (lobachevskiana) ha sido aplicada en el disco de Poincaré. Además, esta aplicación se comporta del mismo modo que la proyección de la figura l.l0.c: conserva ángulos y círculos, y todo resulta geométricamente de una forma muy bella. Bien, quizá me he dejado llevar por mi entusiasmo, ¡me temo que esto es lo que les pasa a los matemáticos cuando se encariñan con algo!

El punto intrigante es que, cuando uno se deja llevar por un tema como la geometría del problema anterior, el análisis y los resultados tienen una brillantez que los sostiene, mientras que los análisis que no poseen esta elegancia matemática se quedan en nada. Hay algo particularmente elegante en la geometría hiperbólica. Sería muy bonito, al menos para las personas como yo, que el Universo estuviese construido también de este modo. Déjenme decir que tengo otras razones para creer que esto es así. A muchas otras personas no les gustan estos universos hiperbólicos abiertos: frecuentemente, prefieren universos cerrados, tales como los ilustrados en la figura 1.16.b, que son bonitos y confortables. Está bien; los universos cerrados siguen siendo muy grandes. Por otro lado, a muchos otros les gustan los modelos planos (figura 1.16.a) porque existe un cierto tipo de teoría del Universo primitivo, la *teoría inflacionaria,* que sugiere que la geometría del Universo tendría que ser plana. Debería decir que yo no creo en estas teorías.

Los tres tipos estándar de modelo del Universo se conocen como *modelos de Friedman y* se caracterizan porque son muy, muy simétricos. Son inicialmente modelos en expansión, pero en todo instante el Universo es perfectamente uniforme en todas partes. Esta hipótesis está incorporada en la estructura de los modelos de Friedman y se conoce como *principio cosmológico.* Dondequiera que uno esté, el universo de Friedman tiene la misma apariencia en todas direcciones. Resulta que nuestro Universo

actual es así hasta un grado notable. Si las ecuaciones de Einstein son correctas, y yo he mostrado que la teoría concuerda con la observación hasta un grado extraordinario, entonces nos vemos obligados a tomar en serio los modelos de Friedman. Todos estos modelos tienen este rasgo molesto, conocido como la Gran Explosión, donde todo falla, exactamente, en el comienzo. El Universo es infinitamente denso, infinitamente caliente y así sucesivamente; algo ha ido fallando en la teoría. De todas maneras, si se acepta que esta fase muy caliente y muy densa tuvo lugar realmente, podemos hacer predicciones sobre cuál debería ser hoy el contenido térmico del Universo y una de estas afirma que, en el momento actual, dicho contenido debería consistir en un fondo uniforme de radiación de cuerpo negro a nuestro alrededor. Precisamente este tipo de radiación fue descubierto por Penzias y Wilson en 1965. Las observaciones más recientes del espectro de esta radiación, que se conoce como *radiación de fondo cósmico de microondas*, hechas por el satélite COBE, muestran que tiene un espectro de cuerpo negro con extraordinaria precisión (figura 1.22).

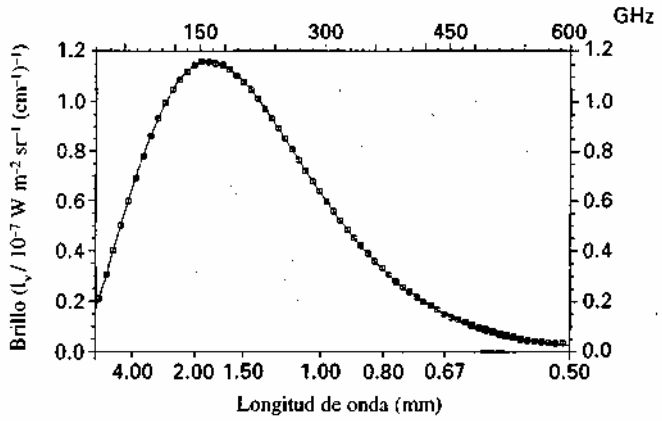
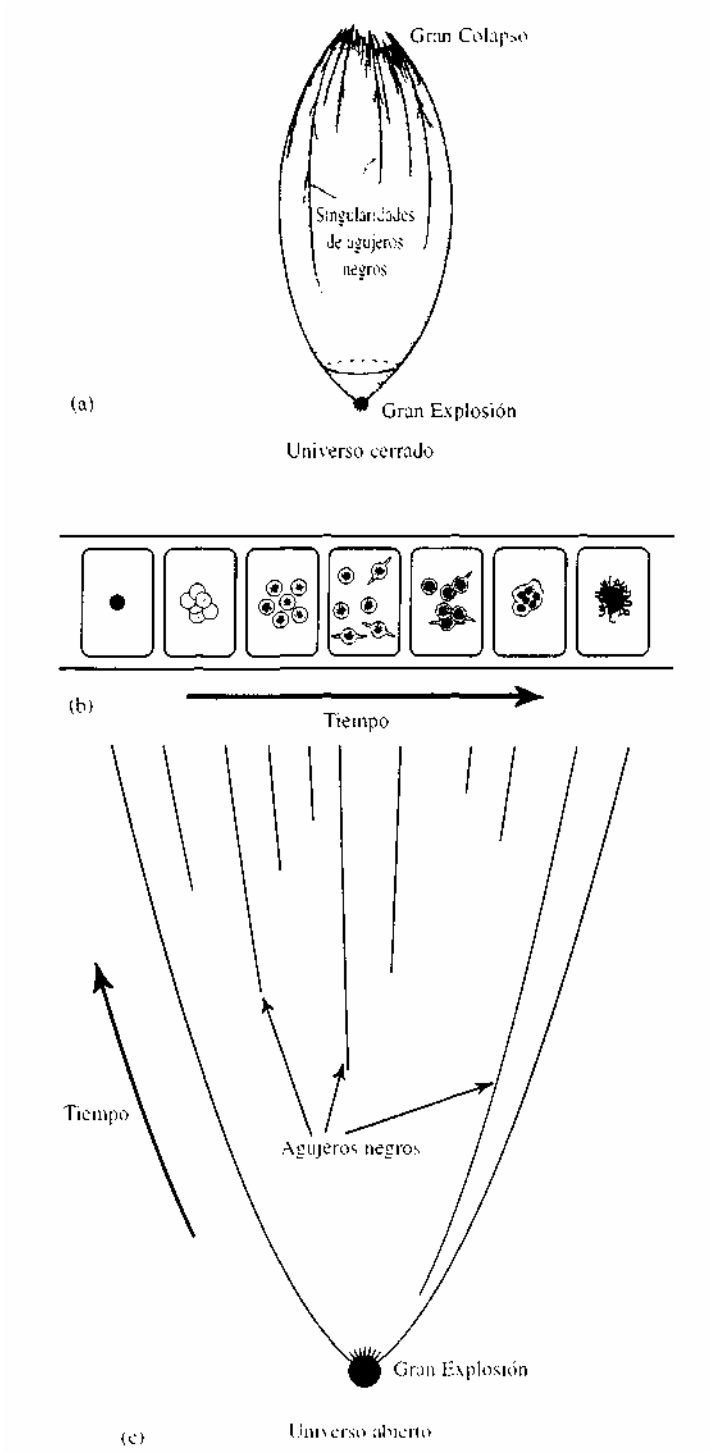


Figura 1.22. Acuerdo exacto entre las medidas de COBE del espectro de la radiación de fondo cósmico de microondas y la naturaleza *térmica* esperada de la radiación de la Gran Explosión (línea continua).

Todos los cosmólogos interpretan la existencia de esta radiación como evidencia de que nuestro Universo pasó por una fase densa y caliente. Esta radiación nos está diciendo, por consiguiente, algo sobre la naturaleza del Universo primitivo; no nos está diciendo todo, sino que algo parecido a la Gran Explosión tuvo lugar. En otras palabras, el Universo debe de haber sido muy parecido a los modelos ilustrados en la figura 1.16.

Hay otro descubrimiento muy importante hecho por el satélite COBE, y es que, aunque la radiación de fondo cósmico de microondas es notablemente uniforme y todas sus propiedades pueden ser explicadas de forma matemáticamente muy bella, el Universo no es en absoluto uniforme. Existen minúsculas pero reales irregularidades en la distribución de la radiación en el cielo. De hecho, esperábamos que estas minúsculas irregularidades estuvieran presentes en el Universo primitivo: al fin y al cabo, nosotros estamos aquí para observar el Universo y, ciertamente, no somos solo una mancha uniforme. El Universo es probablemente más parecido a las imágenes ilustradas en la figura 1.23. Para mostrar mi imparcialidad, estoy utilizando como ejemplos tanto un universo abierto como uno cerrado.

En el Universo cerrado, las irregularidades se desarrollarán para formar estructuras observables reales -estrellas, galaxias y demás- y, al cabo de un tiempo, aparecerán agujeros negros a partir del colapso de estrellas, mediante la acumulación de masa en



los centros de las galaxias y así sucesivamente. Todos estos agujeros negros tienen centros singulares, muy parecidos a una Gran Explosión invertida. Sin embargo, no es tan simple como eso. Según la imagen que hemos desarrollado, la Gran Explosión inicial es un estado bonito, simétrico y uniforme, pero el punto final del modelo cerrado es un caos horrible, en el que todos los agujeros negros llegan eventualmente a juntarse para producir una increíble confusión en el Gran Colapso final (figura 1.23.a). La evolución de este modelo cerrado está ilustrada esquemáticamente en las secuencias mostradas en la figura 1.23.b. En el caso de un modelo de universo abierto, los agujeros negros siguen formándose: sigue habiendo una singularidad inicial y aparecen singularidades en los centros de los agujeros negros (figura 1.23.c).

Resalto estas características de los modelos de Friedman para mostrar que hay una gran diferencia entre lo que parece que vemos en el estado inicial y lo que esperamos encontrar en el futuro remoto. Este problema está relacionado con una ley fundamental de la física conocida como *segunda ley de la termodinámica.*

Podemos entender esta ley mediante un ejemplo sencillo de la vida cotidiana. Imaginemos un vaso de vino situado en el borde de una mesa: podría caer de la mesa, romperse en pedazos y quedar el vino salpicado sobre la alfombra (figura 1.24). No hay nada en la física newtoniana que nos diga que no puede suceder el proceso inverso. Sin embargo, esto no se observa nunca: nunca vemos vasos de vino que se recomponen por sí solos y que el vino sea extraído de la alfombra y entre en el vaso recompuesto. Por lo que concierne a las leyes detalladas de la física, un sentido del tiempo es exactamente tan bueno como el otro. Para entender esta diferencia necesitamos la segunda ley de la termodinámica, que nos dice que la entropía del sistema aumenta con el tiempo. Esta magnitud, llamada *entropía,* es menor cuando el vaso está sobre la mesa que cuando está hecho pedazos en el suelo. De acuerdo con la segunda ley de la termodinámica, la entropía del sistema ha aumentado. Hablando en términos muy generales, la entropía es una medida del desorden de un sistema. Para expresar este concepto de forma más precisa tenemos que introducir el concepto de *espacio de fases.*

Un espacio de fases es un espacio de un enorme número de dimensiones, y cada punto de este espacio multidimensional describe las posiciones y los momentos de todas las partículas que constituyen el sistema bajo consideración. En la figura 1.25 hemos seleccionado un punto concreto en este enorme espacio de fases, un punto que representa dónde están situadas todas las partículas y cómo se están moviendo. Conforme evoluciona el sistema de partículas, el punto se desplaza hacia algún otro lugar en el espacio de fases y yo lo he mostrado serpenteando de un lugar a otro en este espacio.

Figura 1.23. (a) Evolución de un modelo cerrado del mundo con la formación de agujeros negros, a medida que objetos de diferentes tamaños alcanzan los puntos finales de su evolución. Puede preciarse que se espera que haya un horrible caos en el Gran Colapso. La secuencia de sucesos en *(a)* [se] muestra también como una *película* en *(b). (c)* Evolución de un modelo abierto que muestra la formación de agujeros negros en diferentes instantes.

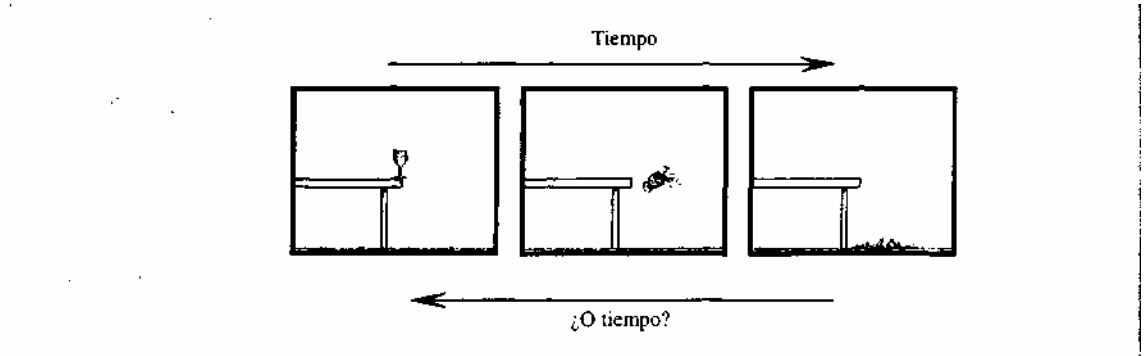


Figura 1.24. Las leyes de la mecánica son reversibles con respecto al tiempo; pese a todo, la ordenación temporal de una escena desde el fotograma derecho al izquierdo es algo que nunca se experimenta, mientras que desde el fotograma izquierdo al derecho resulta perfectamente normal.

Esta línea serpenteante representa la evolución ordinaria del sistema de partículas. Aún no hay entropía allí. Para tener entropía tenemos que dibujar pequeñas burbujas alrededor de regiones en donde amontonamos aquellos estados diferentes que uno no puede distinguir. Esto puede parecer algo confuso: ¿qué es lo que se entiende por *no puede distinguir?* ¿Depende de quién está mirando y con qué cuidado mira? Bien, una de las cuestiones algo delicadas de la física teórica consiste en decir exactamente lo que se entiende por entropía. Esencialmente, lo que se quiere decir es que uno tiene que agrupar estados de acuerdo con lo que se conoce como *granulado grueso,* es decir, de acuerdo con aquellas cosas que uno no puede distinguir. Uno toma todos los estados que, digamos, yacen en esta región del espacio de fases, los amontona, considera el

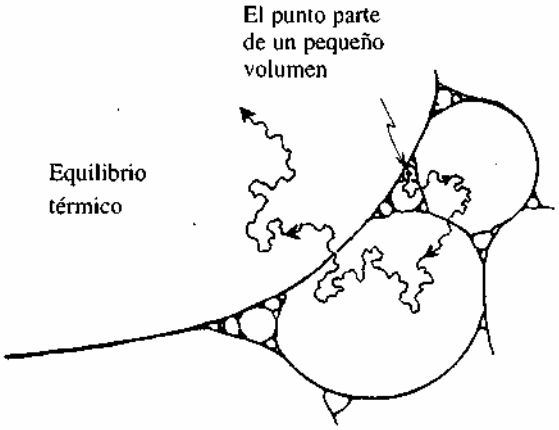


Figura 1.25. La segunda ley de la termodinámica en acción: a medida que el tiempo evoluciona, el punto en el espacio de fases entra en compartimentos de volumen cada vez mayor. En consecuencia, la entropía aumenta continuamente.

volumen de dicha región del espacio de fases, toma el logaritmo del volumen y lo multiplica por la constante conocida como constante de Boltzmann, y eso es la entropía. Lo que nos dice la segunda ley de la termodinámica es que la entropía aumenta, realmente algo bastante estúpido: todo consiste en que, si el sistema empieza en una caja minúscula y se le permite evolucionar, se acomodará a cajas cada vez más grandes. Es muy probable que esto suceda porque, si uno considera el problema cuidadosamente, las cajas más grandes son extraordinariamente más grandes que las cajas pequeñas vecinas. Por ello, si uno se encuentra en una de las cajas grandes, no hay prácticamente ninguna posibilidad de que vuelva a una caja más pequeña. Eso es todo lo que hay. El sistema simplemente vaga por el espacio de fases en cajas cada vez mayores. Esto es lo que nos afirma la segunda ley. ¿O no es así?

En realidad, esta es solo la mitad de la explicación. La ley nos dice que, si conocemos el estado del sistema ahora, podemos prever su estado más probable en el futuro. Pero nos dará una respuesta completamente errónea si tratamos de utilizar el mismo argumento hacia atrás. Supongamos que el vaso está en el borde de la mesa. Podemos preguntar: ¿Cuál es la forma más probable en la que pudo haber llegado allí? Si ustedes utilizan el argumento que acabamos de presentar hacia atrás, concluirán que lo más probable es que empezara como un gran amasijo en la alfombra y que luego saltó por sí solo desde la alfombra y se recompuso en la mesa. Evidentemente no es esta la explicación correcta: la explicación correcta es que alguien lo puso allí. Y dicha persona lo puso allí por alguna razón, que a su vez era debida a alguna otra razón y así sucesivamente. La cadena de razonamiento se remonta cada vez más atrás hasta estados cada vez menores de entropía en el pasado. La curva física correcta es la *real* ilustrada en la figura 1.26 (no la *retrodicha):* la entropía disminuye, y disminuye, y disminuye en el pasado.

La razón de que la entropía aumente en el futuro se explica por la acomodación a cajas cada vez mayores; la razón de que se reduzca en el pasado es algo completamente

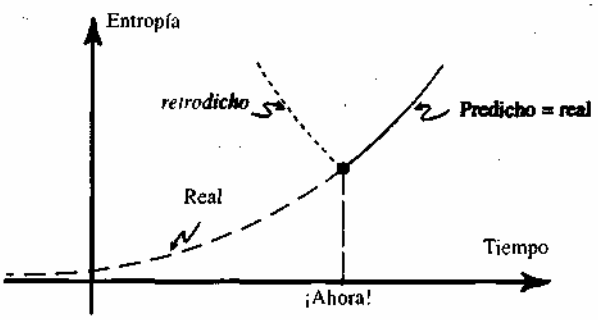


Figura 1.26. Si utilizamos el argumento representado en la figura 1.25 en el sentido contrario del tiempo, *retrodecimos* que la entropía también debería aumentar en el pasado, a partir de su valor actual. Esto está en franca contradicción con la observación.

diferente. Debe de haber habido algo que la hizo disminuir en el pasado. ¿Qué pudo ser? A medida que retrocedemos en el tiempo, la entropía se hace cada vez más pequeña hasta que, finalmente, terminamos en la Gran Explosión.

Debe de haber habido algo muy, muy especial en la Gran Explosión, pero qué fue exactamente es una cuestión controvertida. Una teoría popular, en la que dije que no creía pero con la que simpatiza mucha gente, es la idea del universo inflacionario. Esta idea consiste en que el Universo es tan uniforme a gran escala debido a algo que se supone que ha tenido lugar en las fases muy tempranas de su expansión. Se supone que una tremenda expansión tuvo lugar cuando el Universo tenía solo unos 10^{-36} segundos y la idea es que, independientemente del aspecto que tuviera en aquellas etapas muy tempranas, si se expande en un factor enorme de aproximadamente 10^{60}, entonces, parecerá plano. De hecho, esta es una razón que justifica la popularidad del Universo plano.

Pero, tal como está, el argumento no hace lo que se le supone: lo que cabría esperar en este estado inicial, si fuera escogido al azar, sería una horrible mescolanza y, si se expande esta mescolanza en este factor enorme, sigue siendo una completa mescolanza. De hecho, parece cada vez peor cuanto más se expande (figura 1.27).

Por eso, el argumento por sí mismo no explica por qué el Universo es tan uniforme. Necesitamos una teoría que describa cómo fue realmente la Gran Explosión. No sabemos cuál es esta teoría pero sabemos que tiene que involucrar una combinación de física a gran escala y a pequeña escala. Tiene que involucrar tanto la física cuántica como la física clásica. Además, afirmaré que entre las implicaciones de la teoría debe estar que la Gran Explosión era tan uniforme como observamos que lo es ahora. Quizá semejante teoría termine produciendo un universo hiperbólico de Lobachevski, corno la imagen que yo prefiero, pero no insistiré en ello.

i

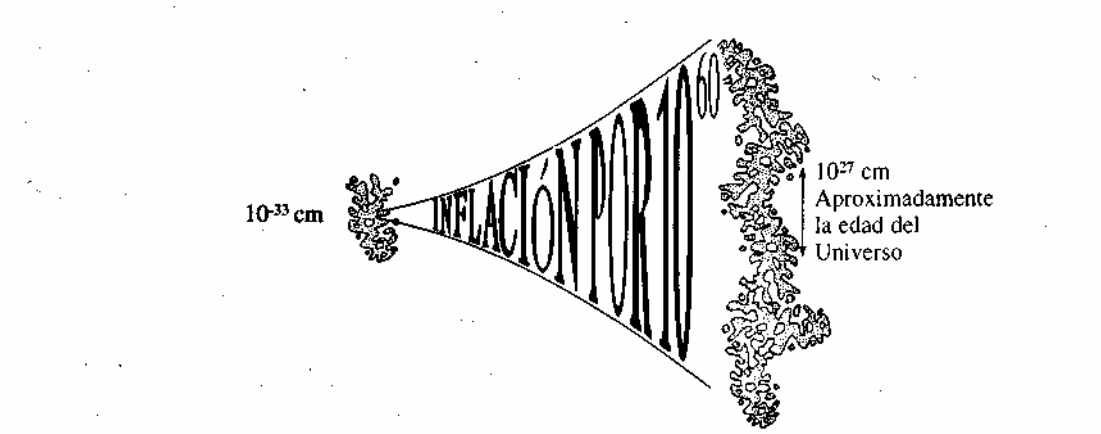


Figura 1.27. Ilustración del problema de la inflación de irregularidades ***genéricas*** **en el Universo primitivo.**

Volvamos de nuevo a las imágenes de los universos cerrado y abierto (figura 1.28). Ahora he incluido una imagen de la formación de un agujero negro, que será bien conocida para los expertos. La materia que se colapsa en un agujero negro produce una singularidad, y eso es lo que representan las líneas oscuras en los diagramas espacio-temporales del Universo.

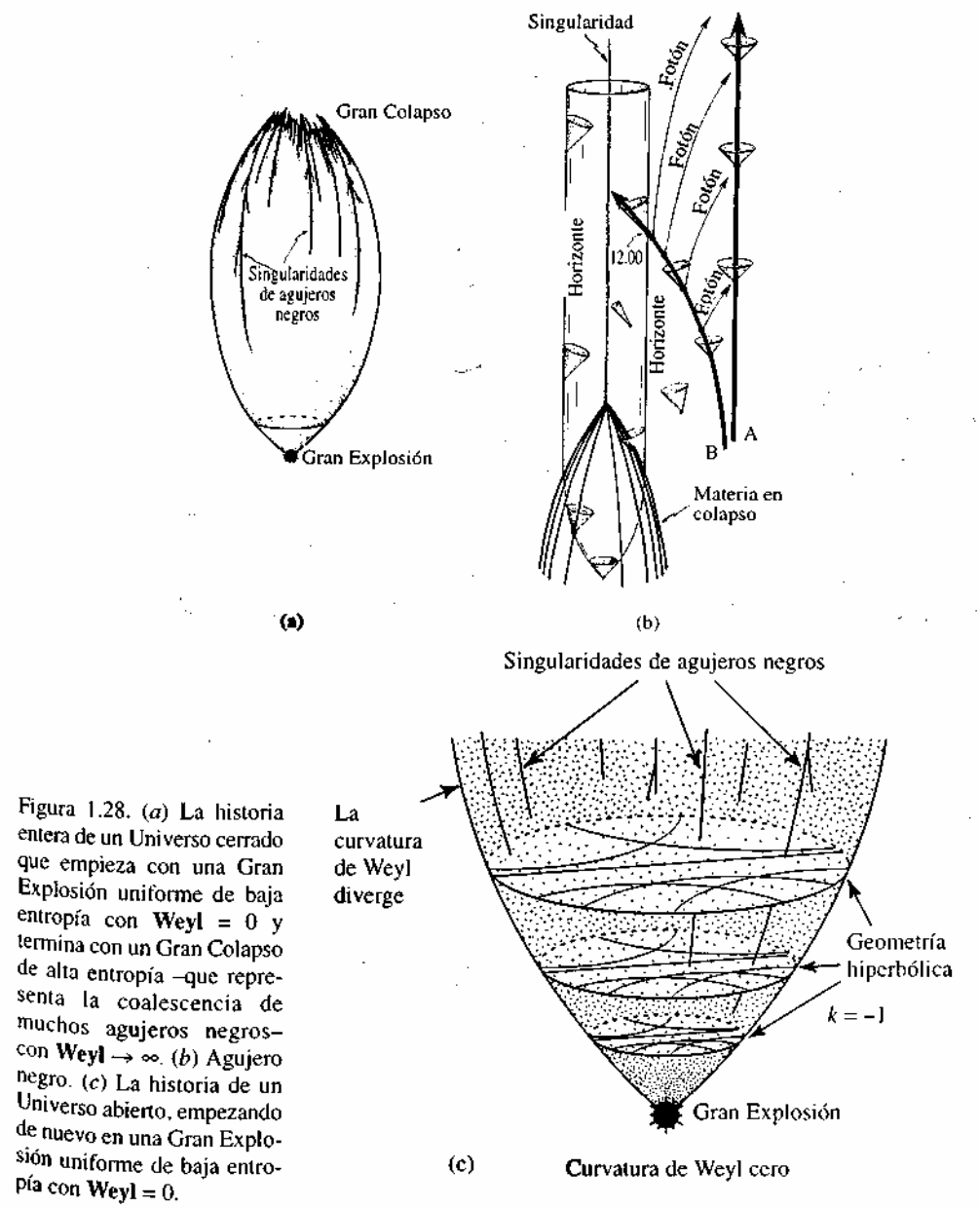


Figura 1.28. (*a*) La historia entera de un Universo cerrado que empieza con una Gran Explosión uniforme de baja entropía con **Weyl** = 0 y termina con un Gran Colapso de alta entropía –que representa la coalescencia de muchos agujeros negros– con **Weyl** $\to \infty$. (*b*) Agujero negro. (*c*) La historia de un Universo abierto, empezando de nuevo en una Gran Explosión uniforme de baja entropía con **Weyl** = 0.

Quiero introducir una hipótesis que yo llamo la *hipótesis de curvatura de Weyl.* Esto no es una consecuencia de ninguna teoría conocida. Como he dicho, no sabemos cuál es la teoría, porque no sabemos cómo combinar la física de lo muy grande y de lo muy pequeño. Cuando descubramos dicha teoría, una de sus consecuencias debería ser esta característica que yo he llamado hipótesis de curvatura de Weyl. Recordemos que la curvatura de Weyl es esa pequeña parte del tensor de Riemann que provoca distorsiones y efectos de marea. Por alguna razón que todavía no entendemos, en las inmediaciones de la Gran Explosión, la combinación apropiada de teorías debe dar como resultado que el tensor de Weyl sea esencialmente cero, o más bien que esté limitado por un valor muy pequeño.

Esto nos daría un Universo como el que se muestra en la figura 1.28.a o 1.28.C, y no como el de la figura 1.29. La hipótesis de curvatura de Weyl tiene asimetría temporal y se aplica solo a las singularidades de tipo pasado y no a las singularidades futuras. Si la misma flexibilidad de permitir que el tensor de Weyl sea *general*, que he aplicado al futuro, se aplicara también al pasado del Universo en el modelo cerrado, terminaríamos con un Universo de apariencia horrible, con tanta mezcla en el pasado como en el futuro (figura 1.29). Esto no se parece nada al Universo en que vivimos.

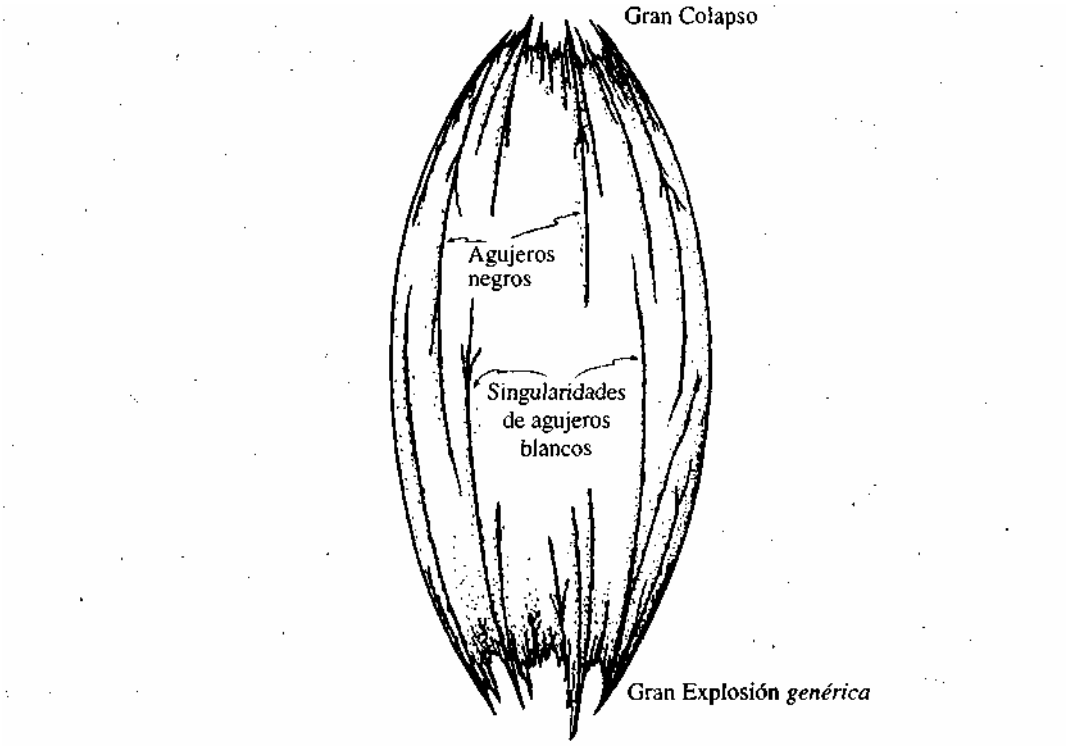


Figura 1.29. Si se elimina la ligadura **Weyl** = 0, entonces tenemos también una Gran Explosión con alta entropía, con **Weyl** → ∞. Un universo semejante estaría plagado de agujeros blancos, y no habría segunda ley de la termodinámica, lo que está en franca contradicción con la experiencia.

¿Cuál es la probabilidad de que, puramente por *azar*, el Universo tuviera una singularidad inicial que se pareciera siquiera remotamente a esto? La probabilidad es menor que una parte en $10^{10(123)}$. ¿De dónde procede esta estimación? Se deriva de una fórmula de Jacob Beckenstein y Stephen Hawking relativa a la entropía de los agujeros negros y, si se aplica en este contexto particular, se obtienen numerosas respuestas. Depende de lo grande que sea el Universo pero, si se adopta mi Universo favorito, el número es, de hecho, infinito.

¿Qué dice esto sobre la precisión que debe estar implicada en la puesta en marcha de la Gran Explosión? Es realmente muy, muy extraordinaria. He ilustrado la probabilidad en un dibujo del Creador, que apunta con una aguja minúscula al punto del espacio de fases que representa las condiciones iniciales a partir de las que debe haber evolucionado nuestro Universo si tiene que parecerse remotamente al Universo en que vivimos (figura 1.30). Para apuntar, el Creador tiene que localizar dicha punta en el espacio de fases con una precisión de una parte en $10^{10(123)}$. Si yo pusiera un cero en cada partícula elemental del Universo, seguiría sin poder escribir todo el número. Es un número inimaginable.

He estado hablando de precisión, de cómo encajan las matemáticas y la física con precisión extraordinaria. He hablado también sobre la segunda ley de la termodinámica, que a menudo se considera como una ley bastante difusa -se refiere a aleatoriedad y a azar- y, pese a todo, hay algo muy preciso oculto bajo esta ley. Tal como se aplica al Universo, tiene que ver con la precisión con la que fue fijado el estado inicial. Esta precisión debe de tener algo que ver con la unión de la teoría cuántica y la relatividad general, una teoría que no tenemos. En el próximo capítulo, sin embargo, diré algo sobre el tipo de cosas que deberían estar implicadas en una teoría semejante.

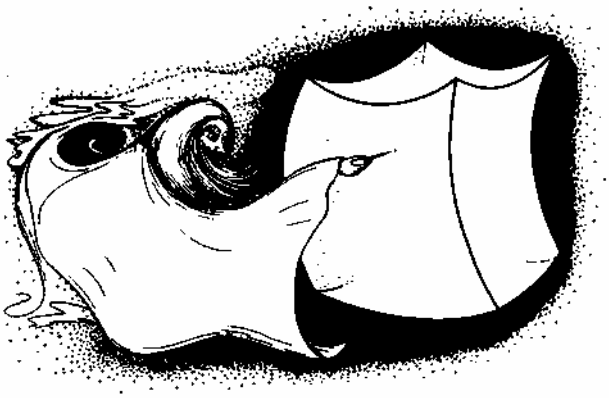


Figura 1.30. Para producir un universo parecido al Universo en que vivimos, el Creador tendría que apuntar a un volumen absurdamente minúsculo del espacio de fases de universos posibles: como mucho a $1/10^{10^{123}}$ del Universo entero. (¡La aguja y el punto señalado no están dibujados a escala!)

2 Los misterios de la física cuántica

En el primer capítulo expliqué que la estructura del mundo físico depende de forma muy precisa de las matemáticas, como se ilustraba simbólicamente en la figura 1.3. Es notable lo extraordinariamente precisas que son las matemáticas para describir los aspectos fundamentales de la física. En una famosa conferencia, Eugene Wigner (1960) se refirió a esto como: «La irrazonable efectividad de las matemáticas en las ciencias físicas». La relación de sus logros es impresionante:

- *La geometría euclidiana* es exacta con un margen de error menor que el diámetro de un átomo de hidrógeno sobre un rango de un metro. Como se discutió en el primer capítulo, no es totalmente exacta debido a los efectos de la relatividad general pero, en cualquier caso, para la mayoría de los propósitos prácticos, la geometría euclidiana es suficientemente exacta.

- *La mecánica newtoniana* es exacta, según se sabe, hasta, aproximadamente, una parte en 10^7, pero no es totalmente exacta. Una vez más, necesitamos la teoría de la relatividad para obtener resultados más precisos.

- *La electrodinámica de Maxwell* es válida en un enorme rango de escalas que va desde los tamaños de las partículas fundamentales, donde se utiliza en combinación con la mecánica cuántica, hasta las dimensiones de las galaxias distantes, lo que corresponde a un rango de escalas de 10^{35} o más.

- *La teoría de la relatividad de Einstein*, como se discutió en el primer capítulo, puede decirse que es exacta hasta casi una parte en 10^{14}, aproximadamente el doble de cifras significativas exactas que en la mecánica newtoniana, por lo que se considera que la mecánica newtoniana queda incluida dentro de la teoría de Einstein.

- *La mecánica cuántica* es el tema de este capítulo y es también una teoría extraordinariamente exacta. En la teoría cuántica de campos, que es la combinación de la mecánica cuántica con la electrodinámica de Maxwell y la teoría especial de la relatividad de Einstein, existen efectos que pueden ser calculados con una precisión aproximada de una parte en 10^{11}. En concreto, en un sistema de unidades conocidas como *unidades de Dirac*, el valor teórico calculado para el momento magnético del electrón es 1,001159652(46), frente al valor determinado experimentalmente que es 1,0011596521(93).

Hay algo muy importante que señalar con respecto a las teorías matemáticas: no solo son extraordinariamente efectivas y exactas en su descripción de nuestro mundo físico, sino que también son tremendamente fructíferas como matemáticas propiamente dichas. Muy a menudo encontramos que algunos de los conceptos matemáticos más fecundos se han basado en conceptos que procedían de teorías físicas. He aquí algunos ejemplos de modelos matemáticos que han sido estimulados por las necesidades de la física:

- los números reales;
- la geometría euclidiana;
- el cálculo infinitesimal y las ecuaciones diferenciales;
- la geometría simpléctica;

- las formas diferenciales y las ecuaciones en derivadas parciales;
- las geometrías de Riemann y de Minkowski;
- los números complejos;
- los espacios de Hilbert; etc.

Uno de los ejemplos más sorprendentes lo constituye el descubrimiento del cálculo infinitesimal, que fue desarrollado por Newton, entre otros, para proporcionar las bases matemáticas de lo que ahora llamamos mecánica newtoniana. Cuando estas diversas herramientas matemáticas fueron aplicadas posteriormente a la solución de problemas puramente matemáticos, también resultaron ser extraordinariamente fructíferas como matemáticas *per se.*

En el capítulo 1 examinamos las escalas de los objetos, que empiezan en la longitud y el tiempo de Planck, las unidades fundamentales de longitud y tiempo; continúan por los tamaños más pequeños encontrados en la física de partículas, que son, aproximadamente, 10^{20} veces mayores que la escala de Planck; siguen por las escalas de longitud y tiempo humanas, que muestran que somos estructuras extraordinariamente estables en el Universo; hasta llegar a la edad y el radio de nuestro Universo físico. Mencioné el hecho más bien inquietante de que en nuestra descripción de la física fundamental utilizamos dos formas completamente diferentes de describir el mundo, dependiendo de si estamos hablando de objetos en el extremo superior o en el extremo inferior de la escala. La figura 2.1 (que es una repetición de la figura 1.5) ilustra la utilización de la mecánica cuántica para describir el pequeño nivel cuántico de actividad y la física clásica para describir los fenómenos a gran escala. He denotado estos niveles de actividad con U, que significa *Unitario,* para el nivel cuántico y con C para el nivel *Clásico.* Discutí la física a gran escala en el capítulo 1 y resalté el hecho de que parece que tenemos leyes muy diferentes en la escala grande y en la escala pequeña.

Creo que la opinión general de los físicos es que si realmente entendiéramos adecuadamente la física cuántica, entonces podríamos deducir a partir de ella la física clásica. Yo quiero presentar un argumento diferente. En la práctica, no es eso lo que uno hace: uno utiliza, *bien* el nivel clásico, *bien* el nivel cuántico. Esto es, de un modo inquietante, parecido a la manera en que los griegos de la Antigüedad consideraban el mundo. Para ellos, había un conjunto de leyes que se aplicaba a la Tierra y otro conjunto diferente que se aplicaba a los Cielos. Fue necesaria la fuerza del punto de vista

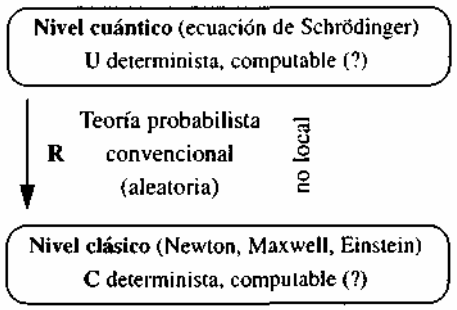


Figura 2.1.

galileano-newtoniano para unificar estos dos conjuntos de leyes y ver que podían entenderse a partir de una misma física. Parece que ahora volvemos a un tipo de situación griega, con un conjunto de leyes que se aplican en el nivel cuántico y otro conjunto para el nivel clásico.

Debería aclarar un posible malentendido existente en relación con la figura 2.1. He colocado los nombres Newton, Maxwell y Einstein en la caja etiquetada «nivel clásico», junto con la palabra *determinista.* No quiero decir que ellos creyeran, por ejemplo, que el Universo se comporta de un modo determinista. Es bastante razonable suponer que Newton y Maxwell no sostuvieron dicha opinión, aunque Einstein aparentemente sí lo hizo. Las notas «determinista, computable (?)» se refieren solo a sus teorías y no a lo que estos científicos creían sobre el mundo real. En la caja etiquetada «nivel cuántico» he incluido las palabras «ecuación de Schrödinger» pese a que estoy seguro de que Schrödinger no creía que toda la física esté descrita por la ecuación que lleva su nombre. Volveré a este punto más adelante. En otras palabras, las personas y las teorías que llevan su nombre son cosas completamente independientes.

Pero, ¿son en realidad distintos estos dos niveles ilustrados en la figura 2.1? Ciertamente podríamos plantear la pregunta: ¿Está el Universo exactamente gobernado solo por las leyes mecano-cuánticas?

¿Podemos explicar el Universo entero en términos de mecánica cuántica? Para abordar estas cuestiones tendré que añadir algo sobre la mecánica cuántica. Permítanme primero una breve lista de algunas de las cosas que la mecánica cuántica puede explicar.

- *La estabilidad de los átomos.* Antes del descubrimiento de la mecánica cuántica no se entendía por qué los electrones, situados dentro de los átomos, no caían en espiral hacia sus núcleos, como deberían hacerlo de acuerdo con una descripción enteramente clásica. No deberían existir átomos clásicos estables.

- *Líneas espectrales.* La existencia de niveles de energía *cuantizados* en átomos y las transiciones entre ellos dan lugar a las líneas de emisión que observamos con longitudes de onda exactamente definidas.

- *Fuerzas químicas.* Las fuerzas que mantienen unidas las moléculas son de naturaleza enteramente mecano-cuántica.

- *Radiación de cuerpo negro.* El espectro de la radiación de cuerpo negro solo puede entenderse si la propia radiación está cuantizada.

- *La fiabilidad de la herencia.* Esta depende de la mecánica cuántica en la escala molecular del ADN.

- *Láseres.* El funcionamiento de los láseres depende de la existencia de transiciones cuánticas estimuladas entre estados mecanocuánticos de moléculas y de la naturaleza cuántica (de Bose-Einstein) de la luz.

- *Superconductores y superfluidos.* Estos son fenómenos que se dan a temperaturas muy bajas y están asociados con correlaciones cuánticas de largo alcance entre electrones (y otras partículas) en diversas sustancias.

En otras palabras, la mecánica cuántica está omnipresente incluso en la vida cotidiana y se encuentra en el corazón de muchas áreas de alta tecnología, incluyendo los ordenadores. La *teoría cuántica de campos,* la combinación de la mecánica cuántica con la teoría especial de la relatividad de Einstein, es también esencial para entender la física de partículas. Como se mencionó más arriba, se sabe que la teoría cuántica de campos es exacta hasta una parte en 10^{11}. Esta lista nos ha mostrado precisamente lo maravillosa y potente que es la mecánica cuántica.

Permítanme decir algo sobre lo qué es la mecánica cuántica. El experimento cuántico arquetípico se muestra en la figura 2.2. Según la mecánica cuántica, la luz consiste ; en partículas denominadas *fotones, y* la figura muestra una fuente de fotones que suponemos los emite de uno en uno. Hay dos rendijas *s* e *i* y una pantalla situada detrás de ellas. Los fotones llegan a la pantalla como si fuesen individuales, y allí son detectados por separado, igual que si fueran partículas ordinarias. El curioso comportamiento cuántico se manifiesta del modo siguiente. Si solo estuviera abierta la rendija *s* y la otra estuviera cerrada, el fotón podría llegar a todos los puntos de una cierta área de la pantalla. Si ahora cierro la rendija *s* y abro la rendija *i*, encuentro de nuevo que el fotón podría llegar a todos los puntos de un área de la pantalla. Pero si abro ambas rendijas, y si he escogido cuidadosamente mi punto en la pantalla, puedo encontrar ahora que el fotón no

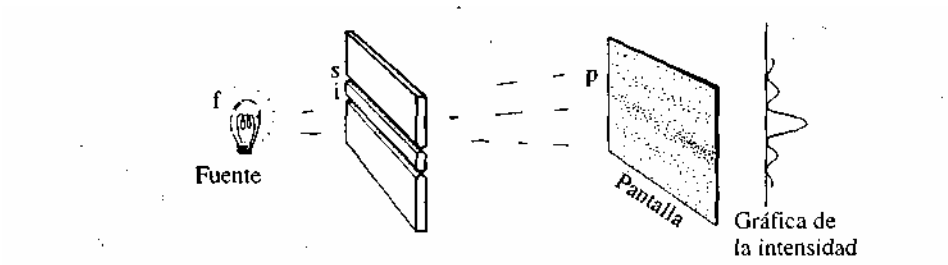


Figura 2.2. El experimento de la doble rendija, con fotones individuales de luz monocromática.

puede llegar a dicho punto, aunque sí podía hacerlo cuando solo una rendija estaba abierta. De algún modo, las dos cosas posibles que el fotón *podría* hacer se cancelan mutuamente. Este tipo de comportamiento no tiene lugar en física clásica, bien sucede una cosa, bien sucede la otra: uno no tiene dos fenómenos posibles que podrían suceder pero que conspiran de alguna manera para cancelarse mutuamente.

La forma en que entendemos el resultado de este experimento en teoría cuántica consiste en decir que, cuando el fotón está *en camino* desde la fuente a la pantalla, el estado del fotón no es el correspondiente a haber atravesado una rendija o haber atravesado la otra, sino que es alguna misteriosa

combinación de los dos, ponderada por *números complejos.* Es decir, podemos escribir el estado del fotón como:

w x (alternativa A) + z x (alternativa B)

donde w y z son números complejos. (Aquí «alternativa A» podría simbolizar el camino *fsp* tomado por el fotón, en la figura 2.2, y «alternativa B» representaría el camino *fip*.) Ahora bien, es importante advertir que los números que multiplican las dos alternativas son números complejos: esta es la razón por la que ocurran las cancelaciones. Ustedes podrían pensar que sería posible calcular el comportamiento del fotón en términos de la probabilidad de que hiciera una cosa u otra, y entonces w y z serían probabilidades reales ponderadas. Pero esta interpretación no es correcta, porque w y z son números complejos. Esto es lo importante en mecánica cuántica. Uno no puede explicar la naturaleza ondulatoria de las partículas cuánticas en términos de *ondas de probabilidad* de alternativas: ¡son *ondas complejas* de alternativas! Ahora bien, los números complejos son entidades que incluyen la raíz cuadrada de menos uno, i = V—1, además de los números reales ordinarios. Pueden representarse en una gráfica bidimensional donde los números puramente reales se encuentran a lo largo del eje x, el eje real, y los números puramente imaginarios a lo largo del eje Y, el eje imaginario, como se ilustra en la figura 2.3.a. En general, un número complejo es una cierta combinación de números puramente reales y puramente imaginarios, tal como 2 + 3v(-l) = 2 + 3i, y puede representarse por un punto en la gráfica de la figura 2.3.a, conocida como *diagrama de Árgana* (también *plano de Wessel* o *plano de Gauss).*

Cada número complejo puede representarse como un punto en la figura 2.3.a y existen diversas reglas acerca de cómo sumarlos, multiplicarlos y demás operaciones. Por ejemplo, para sumarlos se utiliza simplemente la regla del paralelogramo, que equivale a sumar las partes reales y las partes imaginarias por separado, como se ilustra en la figura 2.3.b. También pueden multiplicarse utilizando la regla de los triángulos semejantes, como se ilustra en la figura 2.3.c. Cuando ustedes se familiarizan con diagramas como los de la figura 2.3, los números complejos se convierten en entidades mucho más concretas, en lugar de objetos abstractos. El hecho de que estos números estén incorporados en los fundamentos de la teoría cuántica hace que la gente tenga a menudo la sensación de que la teoría es más bien abstracta y un tipo de cosa incognoscible, pero una vez que uno se acostumbra a los números complejos, particularmente después de jugar

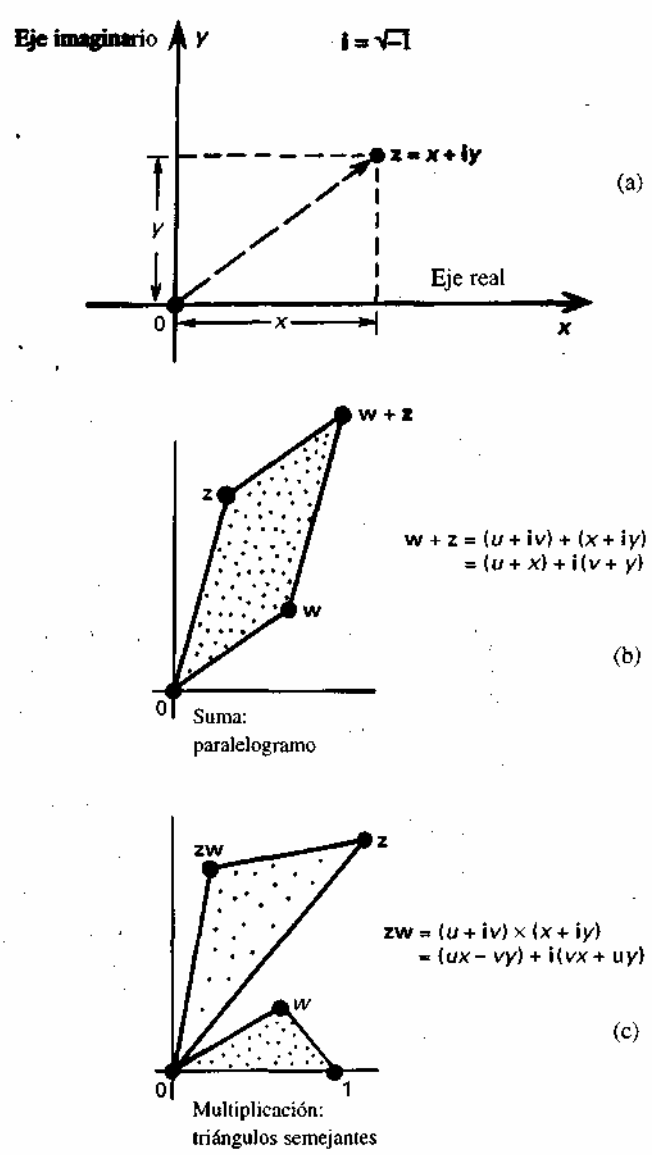


Figura 2.3. (*a*) Representación de un número complejo en el plano complejo (de Wessel-Argand-Gauss). (*b*) Descripción geométrica de la suma de números complejos. (*c*) Descripción geométrica de la multiplicación de números complejos.

con ellos en el diagrama de Argand, estos se hacen objetos muy concretos y uno ya no se preocupa tanto por ellos.

Hay en la teoría cuántica, sin embargo, algo más que la simple superposición de estados ponderados por números complejos. Hasta aquí hemos permanecido en el nivel cuántico, donde se aplican las reglas que llamo U. En este nivel, el estado del sistema viene dado por una superposición de todas las alternativas posibles ponderadas por números complejos. La evolución temporal del estado cuántico se denomina *evolución unitaria,* o evolución de Schrödinger -que es lo que representa U-. Una propiedad importante de U es que es *lineal.* Esto significa que una superposición de dos estados siempre evoluciona de la misma forma que lo haría cada uno de los dos estados por separado, pero superpuestos con ponderaciones complejas que permanecen *constantes en el tiempo.* Esta linealidad es una característica fundamental de la ecuación de Schrödinger. En el nivel cuántico, estas superposiciones ponderadas por números complejos se mantienen siempre.

Sin embargo, cuando se amplifica algo hasta el nivel clásico, *las reglas cambian.* Por amplificar hasta el nivel clásico entiendo el paso del nivel superior U al nivel inferior C de la figura 2.1. Esto es lo que sucede físicamente, por ejemplo, cuando observamos un punto en la pantalla. Un suceso cuántico a pequeña escala desencadena algo más grande que es realmente visible en el nivel clásico. Lo que uno hace en teoría cuántica estándar es sacar del armario algo que a la gente no le gusta mencionar demasiado. Es lo que se denomina el *colapso de la función de onda* o la *reducción del vector de estado* (utilizo la letra R para este proceso). Lo que uno hace ahora es algo completamente diferente de la evolución unitaria. En una superposición de dos alternativas, uno considera los dos números complejos y toma los cuadrados de sus módulos -eso significa tomar los cuadrados de las distancias desde el origen a los dos puntos en el plano , de Argand- y estos dos módulos al cuadrado se convierten en las razones de las probabilidades de las dos alternativas. Pero esto solo sucede cuando se *realiza una medida,* o se *hace una observación.* Podemos pensar en ello como el proceso de amplificar fenómenos desde el nivel U hasta el nivel C de la figura 2.1. Con este proceso, uno cambia las reglas -ya no mantiene estas superposiciones lineales-. Súbitamente, las razones de estos módulos al cuadrado se convierten en probabilidades. Es solo al pasar del nivel U al C cuando se introduce indeterminismo. Este indeterminismo llega con R. Todas las cosas en el nivel U son deterministas: la mecánica cuántica solo se hace indeterminista cuando uno lleva a cabo lo que se denomina *hacer una medida.*

Así pues, este es el esquema que se utiliza en mecánica cuántica estándar. Es un esquema de un tipo muy singular para una teoría fundamental. Quizá sea solo una aproximación a alguna otra teoría más general, que pudiera adquirir más sentido, ¡pero este procedimiento híbrido es considerado por todos los profesionales como una teoría fundamental!

Permítanme decir un poco más sobre estos números complejos. A primera vista Parecen objetos muy abstractos que pululan hasta que uno toma el cuadrado de sus módulos y entonces se convierten en probabilidades. De hecho, a menudo tienen un carácter fuertemente geométrico. Quiero darles un ejemplo en el que su significado puede apreciarse más claramente. Antes de hacerlo, permítanme decir algo más sobre *la* mecánica cuántica. Utilizaré esos paréntesis de aspecto divertido, conocidos como *paréntesis de Dirac.* Son simplemente una abreviatura para describir el estado del sistema -cuando escribo |A> quiero decir que el sistema está en el estado cuántico A-. Lo que hay dentro del paréntesis es cierta descripción del estado cuántico. Con frecuencia, el estado mecano-cuántico global del sistema se escribe como Ψ, que es cierta superposición de otros estados y que podríamos escribir, para el experimento de la doble rendija. en la forma:

$$|\Psi\rangle = w\,|A\rangle + z\,|B\rangle$$

Ahora bien, en mecánica cuántica no estamos tan interesados en los tamaños de los propios números como en su cuociente. Existe una regla en mecánica cuántica por la que se puede multiplicar el estado por un número complejo y ello no cambia la situación física (siempre que el número complejo no sea cero). En otras palabras, es solo el cuociente de estos números complejos el que tiene significado físico directo. Cuando interviene R estamos considerando probabilidades, y entonces son los cuocientes de los módulos al cuadrado los que se necesitan; pero si nos quedamos en el nivel cuántico, también cabe interpretar los cuocientes de estos propios números complejos, incluso antes de tomar sus módulos. La esfera de Riemann es una forma de representar números complejos en una esfera (figura l.l0.c). Para ser más correctos, no estamos tratando solamente con números complejos sino con *cuocientes* de números complejos. Tenemos que ser cuidadosos con los cuocientes porque lo que hay en el denominador podría resultar ser cero, en cuyo caso el cuociente se hace infinito -tenemos que tratar también este caso-. Podemos situar todos los números complejos, junto con el infinito, en una esfera mediante esta proyección, en la que el plano de Argand es ahora el plano ecuatorial, que corta a la esfera en el círculo unidad, donde este es el ecuador de la esfera (figura *2.4).* Evidentemente, podemos proyectar cada

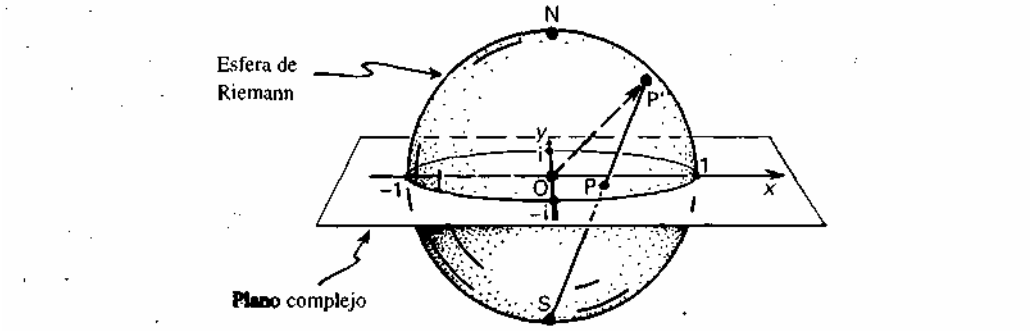


Figura 2.4. La esfera de Riemann. El punto P, que representa u = z / w en el plano complejo, es proyectado desde el polo sur S en un punto P' de la esfera. La dirección OP', desde el centro O de la esfera, es la dirección del eje del espín para el estado superpuesto de las dos partículas de espín 1/2.

punto del plano ecuatorial en la esfera de Riemann, proyectando desde su polo sur. Como puede verse en el diagrama, el polo sur de la esfera de Riemann correspondería, en esta proyección, al *punto del infinito* en el plano de Argand.

Si un sistema cuántico tiene dos estados alternativos, los diferentes estados que pueden construirse combinando ambos se representan por una esfera -una esfera abstracta en esta etapa- pero hay circunstancias en las que ustedes pueden visualizarla realmente. Tengo mucho apego al siguiente ejemplo. Si tenemos una partícula de espín 1/2, tal como un electrón, un protón o un neutrón, entonces las diversas combinaciones de sus estados de espín pueden realizarse geométricamente. Las partículas de espín 1/2 pueden exhibir dos estados de espín, uno con el vector de rotación apuntando hacia arriba (el estado *arriba) y* el otro con el vector de rotación apuntando hacia abajo (el estado *abajo).* La superposición de los dos estados puede representarse simbólicamente por la ecuación:

$$\left|\nearrow\right\rangle = w\left|\uparrow\right\rangle + z\left|\downarrow\right\rangle$$

Las diferentes combinaciones de estos estados de espín dan como resultado una rotación alrededor de algún otro eje y, si ustedes quieren saber cuál es dicho eje, basta con que tomen la razón de los números complejos w y z, que es otro número complejo: u = z/w. Sitúan este nuevo número u en la esfera de Riemann y la dirección que va desde el centro a dicho número complejo es la dirección del eje de espín. Pueden ver así que los números complejos de la mecánica cuántica no son tan abstractos como podría parecer al principio. Tienen un significado bastante concreto -a veces este significado es un poco más difícil de extraer, pero en el caso de la partícula de espín 1/2 el significado es manifiesto-.

Este análisis de las partículas de espín 1/2 nos dice algo más. No hay nada especial en el *espín-arriba* y en el *espín-abajo.* Podría haber escogido cualquier otro eje que hubiera querido, digamos, izquierda o derecha, hacia delante o hacia atrás -ello no supone ninguna diferencia-. Esto ilustra que no hay nada especial en los dos estados de los que uno parta (excepto que los dos estados de espín escogidos deben ser mutuamente opuestos). Según las reglas de la mecánica cuántica, cualquier otro estado de espín es tan bueno como cualquiera de los dos con los que hemos empezado.

La mecánica cuántica es una teoría bella y precisa. Sin embargo, también oculta muchos misterios. En muchos aspectos diversos es una teoría enigmática o paradójica. Quiero resaltar que existen *dos tipos diferentes* de misterios. Yo los llamo misterios z y misterios x.

Los misterios z son los misterios *puzzle:* son fenómenos que están ciertamente ahí, en el mundo físico. Es decir, existen buenos experimentos que nos muestran que la mecánica cuántica se comporta de estas misteriosas maneras. Quizá alguno de estos efectos no ha sido completamente verificado pero hay muy pocas dudas de la corrección de la mecánica cuántica. Estos misterios incluyen fenómenos tales como la *dualidad onda-corpúsculo* (a la que me referí antes), las *medidas nulas* (de las que hablaré en un momento), el *espín* (del que acabo de hablar), y los *efectos no locales* (de los que hablaré de inmediato). Estos fenómenos son genuinamente enigmáticos pero pocas personas discuten su realidad -son ciertamente parte de la naturaleza-.

Existen otros problemas, sin embargo, a los que yo llamo misterios x. Estos son los misterios *paradójicos* (paradox). Estos, a mi modo de ver, son indicios de que la teoría es incompleta, errónea o alguna otra cosa: necesita alguna atención adicional. El misterio x esencial concierne al *problema de la*

medida, que he discutido más arriba -a saber, el hecho de que las reglas cambian de U a R cuando salimos del nivel cuántico y entramos en el nivel clásico-. ¿Podríamos entender la aparición de este procedimiento R quizá como una aproximación, o ilusión, si comprendiéramos mejor la complejidad del comportamiento de los sistemas cuánticos? La más famosa de las paradojas x se refiere al *gato de Schrödinger.* En este experimento -un experimento mental, advierto, ya que Schrödinger era una persona de gran humanidad- el gato está en un estado de muerte y de vida al mismo tiempo. Ustedes no ven realmente gatos como este. Diré más sobre este problema en un momento.

Mi opinión es que debemos aprender a convivir con los misterios z pero que los misterios x deberían quedar eliminados cuando dispongamos de una teoría mejor. Hago hincapié en que esta es solo mi opinión muy particular sobre los misterios x. Muchos otros ven las (¿aparentes?) paradojas de la teoría cuántica bajo una luz diferente o, diría yo, ¡bajo *muchas* luces *diferentes!*

Permítanme decir algo sobre los misterios z antes de que aborde los problemas más serios de los misterios X. Discutiré dos de los misterios z más sorprendentes. Uno de estos es el problema de la *no localidad cuántica* o, como algunos prefieren, del *enmarañamiento cuántico.* Es un fenómeno realmente extraordinario. La idea procedía originalmente de Einstein y sus colegas, Podolsky y Rosen, y se conoce como el experimento EPR. La versión probablemente más fácil de entender es la que dio David Bohm. Tenemos una partícula de espín 0 que se descompone en dos partículas de espín 1/2; por ejemplo, un electrón y un positrón, que salen en direcciones opuestas. Entonces medimos los espines de las partículas salientes en puntos A y B muy separados. Existe un teorema muy famoso, debido a John Bell que afirma que aparece un conflicto entre las expectativas de la mecánica cuántica con respecto a las probabilidades conjuntas de los resultados de medidas en los puntos A y B y cualquier modelo *realista local.* Por modelo *realista local* entiendo cualquier modelo en el que el electrón es un objeto que está en A y el positrón es otro objeto que está en B, y estos dos objetos son independientes uno de otro: no existe ninguna conexión entre ellos. Entonces, esta hipótesis proporciona, para las probabilidades conjuntas de posibles medidas a realizar en A y B, unos resultados que están en conflicto con la mecánica cuántica. John Bell dejó esto muy claro. Es un resultado muy importante y experimentos posteriores, tales como los realizados por Alain Aspect en París, han confirmado esta predicción de la mecánica cuántica. El experimento se ilustra en la figura 2.5 y ; concierne a los estados de polarización de pares de fotones emitidos en sentidos opuestos desde una fuente central.

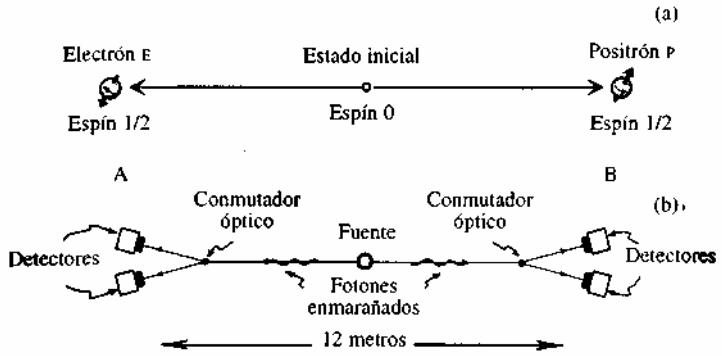


Figura 2.5. (*a*) Una partícula de espín 0 se desintegra en dos partículas de espín 1/2, un electrón E y un positrón P. La medida del espín de una de las partículas de espín 1/2 fija de forma aparentemente *instantánea* el estado de espín de la otra partícula. (*b*) El experimento EPR de Alain Aspect y colegas. Pares de fotones son emitidos desde la fuente en un estado enmarañado. La decisión acerca de en qué dirección se mide el estado de polarización no se toma hasta que los fotones están en vuelo –demasiado tarde para enviar un mensaje al otro fotón diciéndole la dirección de medida–.

La decisión acerca de qué direcciones de polarización de los fotones iban a ser medidas no se tomaba hasta que los fotones estaban en pleno vuelo entre la fuente y los detectores situados en A y B. Los resultados de dichas medidas mostraban claramente que las probabilidades conjuntas para los estados de polarización de los fotones detectados en A y en B coincidían con las predicciones de la mecánica cuántica, como la mayoría de las personas, incluyendo al propio Bell, habían creído, pero que suponen una violación de la hipótesis natural de que estos dos fotones son objetos separados e independientes. El experimento de Aspect estableció efectos de enmarañamiento cuántico sobre una distancia de unos doce metros. Me han informado que existen ahora algunos experimentos relativos a criptografía cuántica en los que efectos similares tienen lugar sobre distancias del orden de kilómetros.

Debería resaltar que, en estos efectos *no locales,* los sucesos ocurren en puntos separados en A y en B, pero están relacionados de forma misteriosa. La manera en que están relacionados -o *enmarañados*- es algo muy sutil, de modo que no hay manera de utilizar dicho enmarañamiento para enviar una señal de A a B -esto muy importante para la consistencia de la teoría cuántica con la relatividad-. De lo contrario, habría sido posible utilizar el enmarañamiento cuántico para enviar mensajes más rápidos que la luz. El enmarañamiento cuántico es un fenómeno muy extraño, se sitúa a mitad de camino entre objetos que están separados y que están en comunicación mutua -es un fenómeno puramente mecano-cuántico y no tiene ningún análogo en la física clásica-.

Un segundo ejemplo de un misterio z se refiere a las *medidas nulas,* y está bien ilustrado por el *problema de la comprobación de bombas de Elitzur-Vaidman.* Imaginen que ustedes pertenecen a un grupo terrorista y han acumulado una gran colección de bombas. Cada bomba contiene en su morro un detonador ultrasensible, tanto que un solo fotón de luz visible reflejado en un pequeño espejo situado en la punta de su morro le proporciona el impulso suficiente para hacer que la bomba estalle violentamente. Hay, sin embargo, una proporción bastante grande de bombas inservibles dentro del conjunto total de bombas. Son inservibles en una forma muy particular. El problema consiste en que el delicado interruptor al que está unido el espejo se quedó atascado durante la fabricación y, por ello, cuando un fotón incide en el espejo de una bomba inservible, el interruptor no se mueve y la bomba no explota (figura 2.6.a). El punto clave es que el espejo del morro de la bomba inservible actúa ahora tan solo como un espejo ordinario fijo, y no como un espejo móvil que forma parte del mecanismo de detonación. Por tanto, aquí está el problema: encontrar una bomba con garantía de funcionamiento,

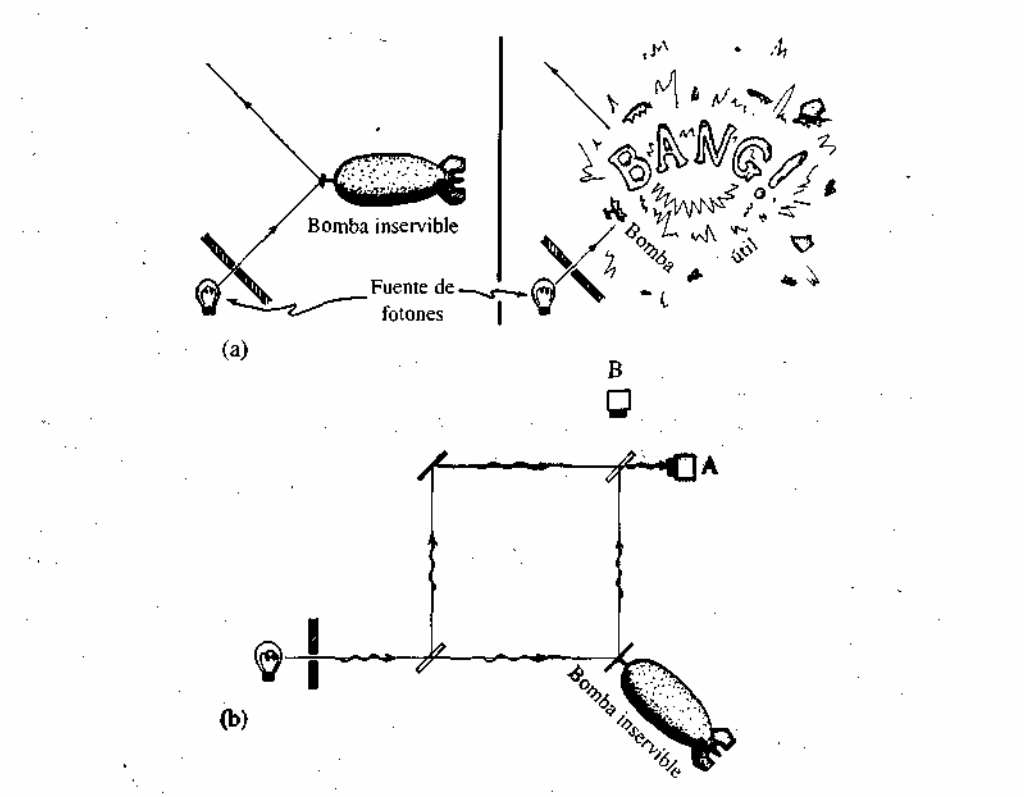


Figura 2.6. (*a*) El problema de la comprobación de bombas de Elitzur-Vaidman. El detonador ultrasensible de la bomba responderá al impulso de un solo fotón de luz visible –suponiendo que la bomba no es una bomba inservible debido a que su detonador está atascado–. El problema consiste en encontrar una bomba útil con garantía, dado un suministro grande de bombas dudosas. (*b*) El montaje para encontrar bombas útiles en presencia de bombas inservibles. En el caso de una bomba útil, el espejo inferior de la derecha actúa como aparato de medida. Cuando mide que un fotón ha seguido el otro camino, esto permite que el detector en A reciba el fotón –lo que no puede suceder en el caso de una bomba inservible–.

dada una gran colección que incluye cierto número de bombas inservibles. En la física clásica no hay manera de hacerlo. La única manera de comprobar si es una bomba útil sería agitar el detonador y entonces la bomba explotaría.

Resulta extraordinario que la mecánica cuántica nos capacite para comprobar si algo *podría* haber sucedido pero no sucedió. Comprueba lo que los filósofos llaman *hechos contrafácticos.* Resulta notable que la mecánica cuántica permita que ¡efectos reales resulten de hechos contrafácticos!

Permítanme mostrarles cómo pueden resolver el problema. La figura 2.6.b muestra la versión original de la solución dada por Elitzur y Vaidman en 1993. Supongan que tenemos una bomba inservible. Tiene un espejo que está atascado -es simplemente un espejo fijo— y, por ello, cuando un fotón rebota en él no provoca una agitación importante del espejo y, en consecuencia, no hay explosión. Realizamos el montaje que se muestra en la figura 2.6.b. Se emite un fotón que incide en primer lugar en un espejo semitransparente. Este es un espejo que transmite la mitad de la luz incidente y refleja la otra mitad. Ustedes podrían pensar que esto significa que la mitad de los fotones que inciden en el espejo son transmitidos a través de él y la otra mitad rebotan. Sin embargo, lo que sucede en el nivel cuántico de los fotones individuales no es esto en absoluto. Lo que de verdad ocurre es que cada fotón, emitido individualmente desde la fuente, estaría en un estado de superposición cuántica de las dos rutas alternativas para el fotón: la transmitida y la reflejada. El espejo de la bomba está en el camino del fotón transmitido y forma con él un ángulo de 45°. La parte del haz de fotones que se refleja en el espejo semitransparente encuentra otro espejo totalmente reflectante, también formando un ángulo de 45°, y ambos haces se reúnen luego en un espejo semitransparente final, como se muestra en la figura 2.6.b. Hay detectores en dos lugares, A y B.

Consideremos lo que le sucede a cada fotón individual, emitido por la fuente, cuando la bomba es inservible. Cuando encuentra el primer espejo semitransparente, el estado del fotón se divide en dos estados separados, uno de ellos correspondiente al fotón que atraviesa el espejo semitransparente y se dirige hacia la bomba inservible, y el otro correspondiente al fotón que es reflejado hacia el espejo fijo. (Esta superposición de caminos alternativos para el fotón es exactamente la misma que la que se produce en el experimento de la doble rendija ilustrado en la figura 2.2. Es también esencialmente el mismo fenómeno que sucede cuando sumamos espines.) Suponemos que las longitudes de los caminos entre el primero y el segundo espejo semitransparente son exactamente iguales. Para ver cuál es el estado del fotón cuando llega a los detectores tenemos que comparar los dos caminos que puede tomar el fotón para llegar a cualquiera de ellos, al coexistir ambos caminos en una superposición cuántica. Encontramos que los caminos se cancelan en B, mientras que se suman en A. Así pues, solo puede haber una señal para activar el detector A y nunca el detector B. Sucede exactamente igual que en la figura de interferencia mostrada en la figura 2.2: existen algunas posiciones en las que la intensidad es siempre nula porque los dos fragmentos del estado cuántico se cancelan en dicho Punto. Así pues, por reflexión en una bomba inservible, siempre es activado el detector A y nunca el detector B.

Supongamos ahora que tenemos una bomba efectiva. En ella, el espejo de su morro ya no es un espejo fijo sino que su potencialidad para moverse convierte a la bomba en un *aparato de medida.* La bomba mide una u otra de las dos alternativas para el fotón en el espejo: puede estar en un estado correspondiente a que un fotón haya llegado o a que no haya llegado. Supongamos que el fotón atraviesa el primer espejo semitransparente y el espejo situado en el morro de la bomba mide que realmente ha seguido su camino. Entonces, ¡boom!, la bomba estalla. La hemos perdido. Así que sacamos otra bomba y ensayamos de nuevo. Quizá esta vez la bomba mida que el fotón no llega: la bomba no explota, de modo que lo que se ha medido es que el fotón ha seguido el otro camino. (Esta es una medida nula.) Ahora bien, cuando el fotón alcanza el segundo espejo semitransparente, es transmitido y reflejado por igual y por eso es ahora posible que B sea activado. Así pues, con una bomba útil, un fotón es detectado de vez en cuando por B, lo que indica que la bomba midió que el fotón siguió el otro camino. El punto clave es que, cuando la bomba es una bomba útil, actúa como un aparato de medida, y esto interfiere con la cancelación exacta que es necesaria para impedir que el fotón sea detectado por B, incluso si el fotón no interacciona con la bomba -una *medida nula*-. Si el fotón no siguió ese camino, ¡entonces tuvo que haber seguido el otro camino! Si B detecta el fotón, entonces sabemos que la bomba actuó como un aparato de medida y, por tanto, era una bomba efectiva. Más aún, con una bomba efectiva el detector B mediría, de vez en cuando, la llegada del fotón y la bomba no explotaría. Esto *solo* puede suceder si es una bomba efectiva. Ustedes saben que es una bomba efectiva porque ha medido que el fotón ha seguido realmente el otro camino.

Es realmente extraordinario. En 1994, Zeilinger visitó Oxford y me dijo que había realizado el experimento de comprobación de bombas. En realidad, él y sus colegas no habían experimentado con bombas reales sino con algo, en principio, similar -debería recalcar que Zeilinger no es ciertamente un terrorista-. Me dijo entonces que él y sus colegas Kwiat, Weinfurter y Kasevich habían llegado a una solución mejorada en la que podían realizar el mismo tipo de experimento sin malgastar bombas en

absoluto. No entraré en cómo se consigue esto, puesto que supone un montaje mucho más sofisticado. En realidad, sí existe un desperdicio, pero es casi despreciable, y ustedes pueden encontrar una bomba efectiva garantizada.

Permítanme dejarles con estas ideas. Estos ejemplos ilustran algunos aspectos de la naturaleza extraordinaria de la mecánica cuántica y sus misterios z. Pienso que parte del problema es que algunas personas se quedan hipnotizadas por estas cosas, dicen, «Válgame Dios, qué extraña es la mecánica cuántica», y tienen razón. Tiene que ser bastante extraña para incluir todos estos misterios z como fenómenos reales. Pero entonces piensan que tienen que aceptar también los misterios x y, ¡creo que esto es erróneo!

Volvamos al gato de Schrödinger. La versión del experimento mental mostrado en la figura 2.7 no es exactamente la versión original de Schrödinger pero será más apropiada para nuestros propósitos. De nuevo tenemos una fuente de fotones y un espejo semitransparente que desdobla el estado cuántico de los fotones incidentes en una superposición de dos estados diferentes, uno reflejado y otro que atraviesa el espejo. Existe un detector de fotones en el camino del fotón transmitido que registra la llegada de un fotón y hace que se dispare una pistola que mata al gato. Podría considerarse el gato como el punto final de una medida; pasamos del nivel cuántico al mundo de los objetos ponderables cuando encontramos que el gato está muerto o vivo. Pero el problema es que si se toma el nivel cuántico como algo que es verdadero en todo el camino hasta el nivel de los gatos y así sucesivamente, entonces hay que creer que el estado real del gato es una superposición de estar muerto y vivo a la vez. El caso es que el fotón está en una superposición de estados que siguen un camino u otro, el detector está en una superposición de estados *activado y desactivado, y* el gato está en una superposición de estados *estar vivo y muerto.* Este problema es conocido desde hace tiempo. ¿Qué dicen sobre él diversas personas? Hay probablemente más actitudes diferentes con respecto a la mecánica cuántica que físicos cuánticos. Esto no es inconsistente porque algunos físicos cuánticos sostienen opiniones diferentes al mismo tiempo.

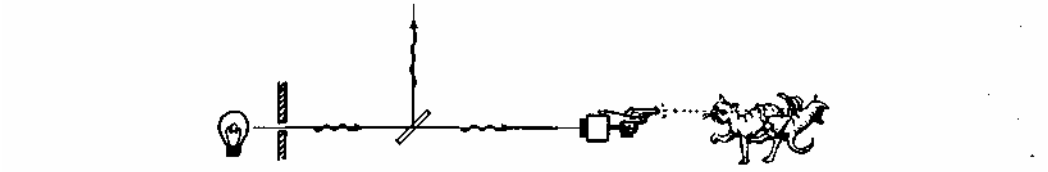


Figura 2.7. *El gato de Schrödinger.* El estado cuántico implica una superposición de un fotón reflejado y uno transmitido. La componente transmitida activa un dispositivo que mata al gato, y, por ello, según la evolución-U, el gato existe en una superposición de vida y muerte.

Quiero ilustrar una amplia clasificación de puntos de vista con un maravilloso comentario de sobremesa hecho por Bob Wald, dice así:

«Si realmente *crees* en la mecánica cuántica, entonces no puedes tomarla *en serio.*»

Me parece que este es un comentario muy profundo y verdadero sobre la mecánica cuántica y las actitudes de la gente hacia ella. He dividido a los físicos cuánticos en diversas categorías en la figura 2.8. En particular, los he clasificado en aquellos que *creen* y aquellos que son *serios.* ¿Qué entiendo por serios? Las personas serias consideran que el vector de estado $|\psi>$ describe el mundo real: el vector de estado *es* la realidad. Aquellos que *realmente* creen en la mecánica cuántica no creen que esta sea la actitud correcta hacia ella. He situado los nombres de varias personas en el diagrama. Hasta donde yo puedo ver, Niels Bohr y los seguidores del punto de vista de Copenhague son creyentes. Bohr creía ciertamente en la mecánica cuántica pero no tomaba en serio el vector de estado como una descripción del mundo. De alguna forma $|\psi>$ estaba por completo en la mente: era nuestra manera de describir el mundo, no era el mundo en sí mismo. Y esto lleva también a lo que John Bell llamaba PTPP, que significa «Para Todos

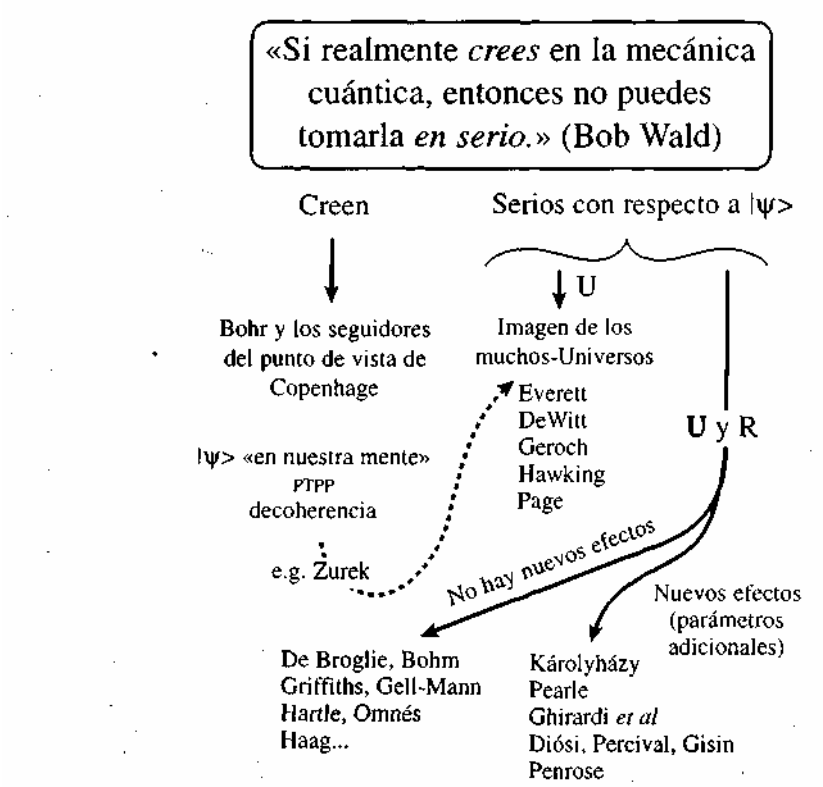


Figura 2.8.

los Propósitos Prácticos». A John Bell le gustaba ese término, pienso yo que porque tiene un sonido ligeramente despectivo[11]. Se basa en el *punto de vista de la decoherencia* sobre el que tendré algo que decir más adelante. Suele pasar que, cuando se interroga minuciosamente a alguno de los más ardientes defensores de PTPP, tales como Zurek, se repliegan al centro del diagrama de la figura 2.8. Ahora bien, ¿qué entiendo yo por *el centro del diagrama?*.

He dividido a la gente *seria* en diferentes categorías. Existen quienes creen que U es toda la historia -entonces tenemos que admitir que la evolución unitaria es toda la historia-. Esto conduce al punto de vista de los *muchos-Universos.* Desde este punto de vista el gato está realmente vivo y muerto a la vez, pero los dos gatos habitan, en cierto sentido, en Universos diferentes. Más adelante insistiré sobre ello. También he señalado algunos de aquellos que han defendido un punto de vista de este tipo general, al menos en alguna etapa de su pensamiento. ¡Los defensores de los muchos-Universos son los únicos que están en el centro de mi diagrama!

Las personas a las que considero *realmente serias* con respecto a |ψ>, y yo mismo me incluyo entre ellas, son aquellas que creen que tanto **U** como **R** son fenómenos reales. No solo tiene lugar la evolución unitaria, mientras el sistema es en cierto sentido pequeño, sino que hay también algo diferente que ocurre y que es esencialmente lo que he llamado **R** -*quizá*, no sea exactamente **R** sino algo parecido lo que está sucediendo ahí fuera-. Si uno cree eso, entonces parece que es posible adoptar uno de los dos puntos de vista. Se puede adoptar el punto de vista de que no hay nuevos efectos físicos que tener en cuenta -y yo he incluido aquí el punto de vista de Broglie/Bohm, además de los muy diferentes de Griffiths, Gell-Mann, Hartle y Omnés-. **R** tiene algún papel que jugar, además de la mecánica cuántica estándar **U**, pero uno no esperaría encontrar ningún efecto nuevo. Y están aquellos que mantienen el segundo punto de vista *realmente serio,* que yo suscribo personalmente, de que algo nuevo tendrá que entrar y cambiar la estructura de la mecánica cuántica. **R** contradice realmente a **U** -algo nuevo está interviniendo-. He incluido los nombres de algunos de aquellos que toman este punto de vista al pie de la letra.

Quiero decir algo más detallado sobre las matemáticas y, en concreto, considerar cómo tratan los diferentes puntos de vista al gato de Schrödinger. Volvamos a la imagen del gato de Schrödinger pero incluyamos ahora las ponderaciones complejas w y z (figura 2.9.a). El fotón se desdobla en los dos estados

[1] En inglés es FAPP (For All Practical Purposes). *(N. del T.)*

y, si uno es serio con respecto a la mecánica cuántica, debe creer que el vector de estado es real y entonces cree también que el gato debe estar realmente en algún tipo de superposición de estados de estar muerto y vivo a la vez. Es muy conveniente representar estos estados de estar muerto y vivo utilizando paréntesis de Dirac, como he mostrado en la figura *2.9*.b. ¡Podemos poner tantos gatos como símbolos dentro de los paréntesis de Dirac! El gato no acapara toda la historia porque también están la pistola y el fotón, y el aire circundante, de modo que hay también un entorno —cada componente del estado es realmente un producto de todos estos efectos juntos, pero seguimos teniendo una superposición (figura 2.9.b)-.

¿Cómo entiende esto el punto de vista de los *muchos-Universos!* Alguien viene y mira al gato y ustedes se preguntan: «¿Por qué no ve esa persona estas superposiciones de estados del gato?». Esto es porque un creyente en los muchos-Universos describiría la situación como se muestra en la figura 2.9.c. Hay un estado de un gato vivo, acompañado por la persona que ve y percibe un gato vivo; y hay otro estado del gato muerto, acompañado por una persona que observa un gato muerto. Estas dos alternativas están superpuestas: he colocado también dentro de los paréntesis de Dirac los estados mentales de la persona que observa el gato en cada uno de estos dos estados -la expresión de la persona refleja el estado mental del individuo-. Por eso, la visión del creyente de los muchos-Universos es que todo está bien: existen copias diferentes de la persona que Percibe el gato, pero dichas copias habitan en *Universos diferentes.* Ustedes podrían imaginar que son una de estas copias, pero existe otra copia de ustedes en otro Universo *Paralelo* que ve la otra posibilidad. Por supuesto, esta no es una descripción muy económica del Universo pero creo que eso no es lo peor de la descripción de los muchos-Universos. No es solo su carencia de economía lo que me preocupa. La principal dificultad es que, de hecho, no resuelve el problema. Por ejemplo, ¿por qué nuestra conciencia no nos

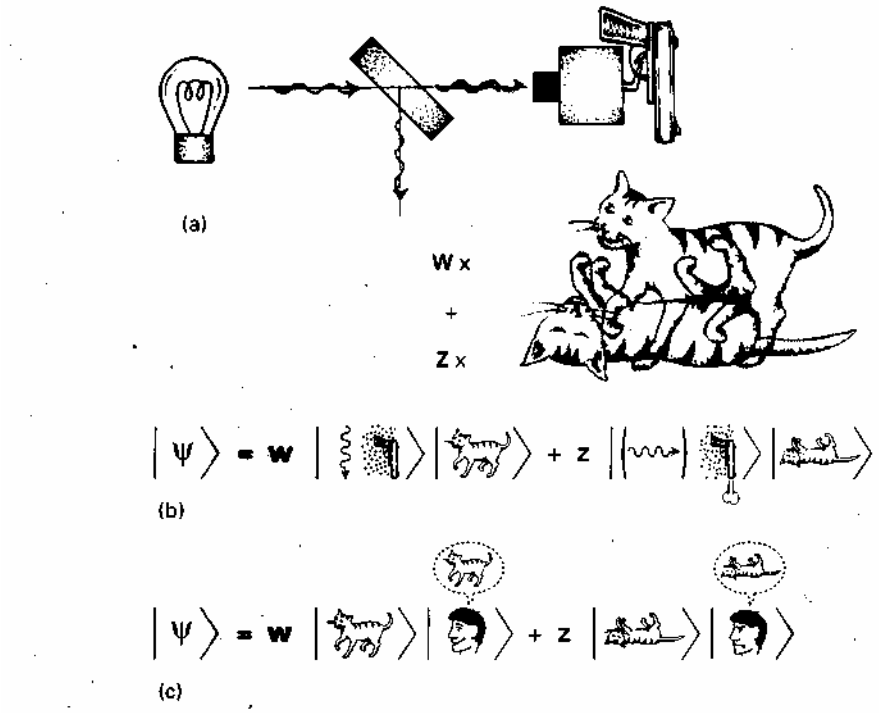


Figura 2.9.

permite percibir superposiciones macroscópicas? Tomemos el caso especial en que w y z son iguales. Entonces, podemos reescribir este estado tal como se muestra en la figura 2.10, es decir: gato vivo, más gato muerto, junto con una persona que percibe el gato vivo, más una persona que percibe el gato muerto, *más* gato vivo menos gato muerto, junto con una persona que percibe el gato vivo, menos una persona que percibe el gato muerto -es solo un poco de algebra-. Ahora ustedes pueden decir, «¡Oiga, usted no puede hacer eso; no son así los estados de percepción!». Pero, ¿por qué no? No sabemos lo que significa percibir. ¿Cómo sabemos que un estado de percepción no podría estar percibiendo un gato vivo y muerto al mismo tiempo? A menos que ustedes sepan qué es la percepción, y tengan una buena teoría de por qué están prohibidos tales estados de percepción mezclados -y eso sería ir mucho más allá del capítulo 3- me parece que esto no proporciona ninguna explicación. No explica por qué tiene lugar la percepción de uno u otro pero no la percepción de una superposición. Podría incluirse en una teoría pero tendríamos que tener también una teoría de la percepción. Hay otra objeción y es que si admitimos que los números w y z son

números generales, la teoría no nos dice por qué las probabilidades son las que proporciona la mecánica cuántica, a las que se llega por la regla de los módulos al cuadrado que describí antes. Estas probabilidades, después de todo, se pueden comprobar de forma muy precisa.

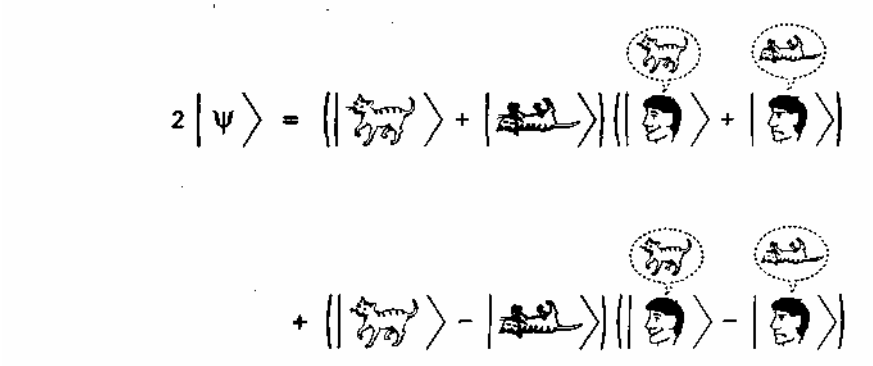
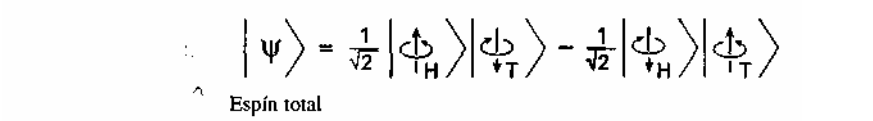


Figura 2.10.

Permítanme llegar un poco más lejos en la cuestión de la medida cuántica. Tendré que decir algo más sobre el *enmarañamiento cuántico.* En la figura 2.11 he dado una descripción del experimento EPR en versión de Bohm que, recordemos, es uno de los misterios z cuánticos. ¿Cómo describimos el estado de las partículas de espín 1/2 que salen en las dos direcciones? El espín total es cero y, por ello, si recibimos aquí una partícula con espín-arriba, sabemos que la partícula allí debe tener espín-abajo. En este caso, el estado cuántico para el sistema combinado sería un producto de *arriba-aqui* por *abajo-allí.* Pero, si encontramos que el espín es *abajo-aquí,* entonces debe ser *arriba-allí.* (Estas alternativas aparecerían si decidimos examinar el espín de la partícula *aquí* en la dirección *arriba/abajo.)* Para obtener el estado cuántico para el sistema entero debemos superponer estas alternativas. De hecho, necesitamos un signo menos para hacer que el espín total del par de partículas juntas sume cero cualquiera que sea la dirección que escojamos.



Espín total

Figura 2.11.

Supongamos ahora que estamos contemplando la realización de una medida del espin sobre la partícula que viene hacia mi detector *aquí,* y que la otra está llegando muy lejos, digamos, a la Luna -de modo que *allí* ¡está en la Luna!-. Imaginemos ahora que hay un colega en la Luna que mide su partícula en una dirección *arriba/abajo.* Tiene la misma probabilidad de encontrar su partícula con *espín-arriba* o *espín-abajo.* Si encuentra *espín-arriba,* entonces el estado de espín de mi partícula debe ser *abajo.*

Si encuentra *espín-abajo,* entonces el estado de espín de mi partícula es *arriba.* Así pues, considero que el vector de estado de la partícula que voy a medir es una mezcla de estados con la misma probabilidad de *espín-arriba* y *espín-abajo.*

En mecánica cuántica existe un procedimiento para tratar mezclas de probabilidad como esta. Uno utiliza una magnitud denominada *matriz densidad.* La matriz densidad que *yo aquí* utilizaría en la situación actual será la expresión indicada en la figura 2.12 El primer 1/2 en la expresión es la probabilidad de que *yo* encuentre que el espín *aquí* está hacia *arriba, y* el segundo 1/2 en la expresión es la probabilidad de que *yo* encuentre que el espín *aquí* está hacia *abajo.* Estas son simplemente probabilidades clásicas ordinarias, que expresan mi incertidumbre con respecto al estado de espín real de la partícula que voy a medir. Las probabilidades ordinarias son simplemente números reales ordinarios (entre 0 y 1), y la combinación indicada en la figura 2.12 no es una superposición cuántica, en la que los coeficientes serían números complejos, sino una combinación de probabilidades ponderadas. Nótese que las cantidades que están multiplicadas por los dos factores de probabilidad (de 1/2) son expresiones que incluyen un primer factor, en el que el paréntesis angulado señala a la derecha -denominado un vector *ket* (de Dirac)- y también un segundo factor en el que el paréntesis angulado apunta a la izquierda -un vector *bra-*. (El vector *bra* es lo que se conoce como el *complejo conjugado* del vector *ket).*

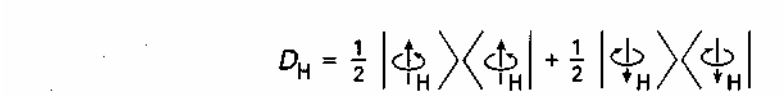


Figura 2.12.

No es este el lugar apropiado para intentar explicar en detalle la naturaleza de las matemáticas involucradas en la construcción de matrices densidad. Baste decir que la matriz densidad contiene toda la información necesaria para calcular las probabilidades de los resultados de medidas que pudieran realizarse en una parte del estado cuántico del sistema, suponiendo que no disponemos de ninguna información relativa a la parte restante de dicho estado. En nuestro ejemplo, el estado cuántico entero se refiere al *par* de partículas juntas (un estado enmarañado) y suponemos que yo no dispongo *aquí* de ninguna información relativa a medidas que pudieran ser realizadas *allí,* en la Luna. sobre la compañera de la partícula que yo voy a examinar *aquí.*

Cambiemos ahora ligeramente la situación y supongamos que mi colega en la Luna decide medir el espín de su partícula en una dirección izquierda/derecha en lugar de arriba/abajo. Por eso, eventualmente es más conveniente utilizar la descripción del estado dada en la figura 2.13. De hecho, es exactamente el mismo estado de antes, representado en la figura 2.11 (como revelará un poco de álgebra basada en la geometría de la figura 2.4) pero el estado se representa de forma diferente. Aún no sabemos cuál es el resultado que obtendrá mi colega en la Luna en su medida de espín *izquierda/derecha.* – pero sabemos que hay una probabilidad 1/2 de que él encuentre *espín-izquierda* -en cuyo caso yo debo encontrar *espín-derecha-* y una probabilidad 1/2 de que él encuen-*espín-derecha* —en cuyo caso yo debo encontrar *espín-izquierda*—. En consecuencia, matriz densidad D_h debe venir dada como en la figura 2.13, y debe resultar que esta es la misma matriz densidad que antes (como aparecía en la figura 2.12). Por supuesto, así es como debería ser. La elección de medida que mi colega adopte en la Luna no debería suponer ninguna diferencia con respecto a las probabilidades que yo obtenga para mis propias medidas. (Si pudieran suponer una diferencia, ello haría posible que mi colega me enviara señales desde la Luna a más velocidad que la luz, estando su mensaje codificado en su elección de la dirección de la medida de espín.)

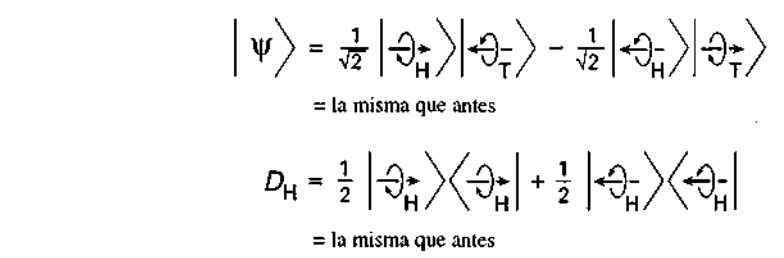


Figura 2.13.

Ustedes también pueden verificar el álgebra directamente para comprobar que las matrices densidad son realmente las mismas. Si les es familiar este tipo de álgebra, sabrán de lo que estoy hablando -si no, no se preocupen-. La matriz densidad es el mejor recurso en caso de que haya alguna parte del estado a la que no se puede acceder. La matriz densidad utiliza probabilidades en el sentido ordinario, aunque combinadas con la descripción mecano-cuántica en la que existen probabilidades mecano-cuánticas implícitamente involucradas. Si yo no tengo ningún conocimiento acerca de lo que está pasando *allí,* esta sería la mejor descripción del estado *aquí* que puedo tener.

Sin embargo, es difícil aceptar que la matriz densidad describa la *realidad.* El problema es que yo no puedo asegurar que no vaya a recibir, más tarde, un mensaje de la Luna diciéndome que mi colega midió realmente el estado y encontró que la respuesta era tal y cual. En tal caso, yo sé cuál debía ser *realmente* el estado de mi partícula. La matriz densidad no me decía *todo* sobre el estado de mi partícula. Por eso necesitaba conocer el estado real del par combinado. Así, la matriz densidad es una especie de descripción provisional, y ésa es la razón de por qué a veces se denomina PTPP (i.e., para todos los propósitos prácticos).

La matriz densidad no se utiliza normalmente para describir situaciones como esta sino más bien para describir situaciones como la que se muestra en la figura 2.14, donde, en lugar de tener un estado enmarañado dividido entre lo que es accesible a mí *aquí* y a mi colega *allí en* la Luna, el estado *aquí es* un gato muerto, o vivo, y el estado *allí* (quizá incluso en la misma habitación) proporciona el estado del entorno global que acompaña al gato. Por ello, como vector de estado enmarañado completo puedo tener

gato vivo junto con un cierto entorno, más gato muerto junto con otro entorno. Lo que dicen las personas PTPP es que uno nunca puede obtener información suficiente sobre el entorno y en consecuencia no utiliza el vector de estado: uno tiene que utilizar la matriz densidad (figura 2.15).

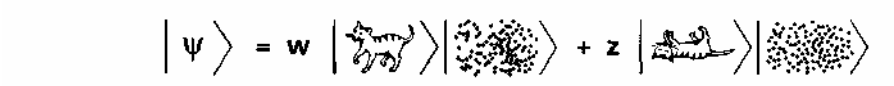


Figura 2.14.

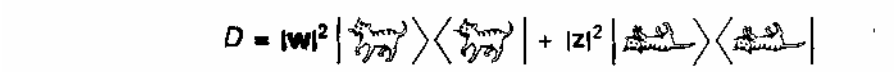
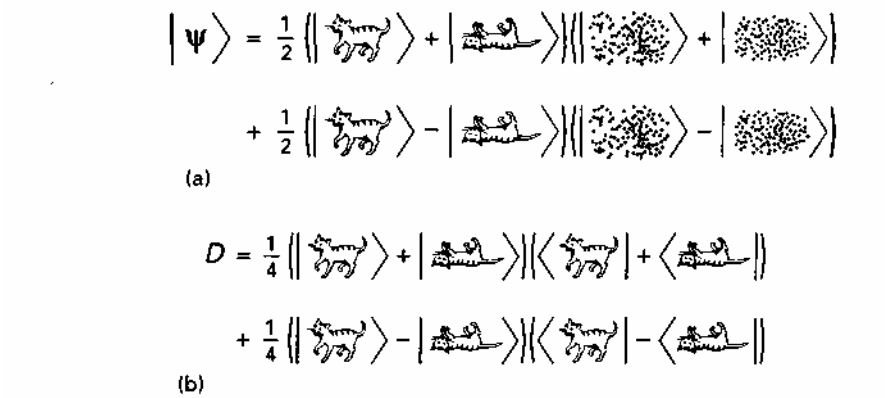


Figura 2.15.

La matriz densidad se comporta entonces como una mezcla probabilista y las personas PTPP afirman que, para todos los propósitos prácticos, el gato está muerto o vivo. Todo esto podría servir *para todos los propósitos prácticos* pero no proporciona una imagen de la realidad: no nos informa de qué podría suceder si alguna persona muy perspicaz llegara más tarde y nos dijera cómo extraer la información del entorno. De algún modo, es un punto de vista temporal: suficientemente bueno mientras nadie sea capaz de obtener dicha información. Sin embargo, podemos llevar a cabo el mismo análisis para el gato que el que llevamos a cabo para la partícula en el experimento EPR. Demostramos qué tan bueno es utilizar los estados de *espín-izquierda* y *espín-derecha* como utilizar *espín-arriba* y *espín-abajo.* Podemos obtener estos estados izquierda y derecha combinando los estados arriba y abajo de acuerdo con las reglas de la mecánica cuántica y obtener el mismo vector de estado enmarañado total para el par de partículas, como se representa en la figura 2.13.a, y la misma matriz densidad, como se representa en la figura 2.13.c.

En el caso del gato y su entorno (en la situación en que las dos amplitudes w y z son iguales), podemos utilizar las mismas matemáticas, donde ahora *gato vivo más gato muerto* desempeña el papel de *espín-derecha y* donde *gato vivo menos gato muerto* desempeña el papel de *espín-izquierda.* Obtenemos el mismo estado que antes (figura 2.14 con w = z) y la misma matriz densidad que antes (figura 2.15 con w = z). ¿Es un gato vivo más muerto o un gato vivo menos muerto tan bueno como un gato vivo o como un gato muerto? Bien, esto no es tan obvio, pero las matemáticas son sencillas. Seguiría siendo la misma matriz densidad para el gato que antes (figura 2.16). Así, conocer cuál es la matriz densidad no nos ayuda a determinar si el gato está realmente vivo o muerto. En otras palabras, la vida o la muerte del gato no está contenida en la matriz densidad: necesitamos algo más.



(a)

(b)

Figura 2.16.

Nada de esto explica por qué el gato está realmente vivo o muerto (y no en alguna combinación de ambos). No solo eso: ni siquiera explica por qué el gato se percibe como vivo o como muerto. Además, en el caso de amplitudes generales (w, z) no explica por qué las probabilidades relativas son $|w|^2$ y $|z|^2$. Mi opinión personal es que esto no es suficiente. Vuelvo al diagrama que muestra el conjunto de la física, ahora corregido, para mostrar lo que yo creo que la física tendrá que hacer en el futuro (figura 2.17). El procedimiento que he descrito mediante la letra **R** es una aproximación a algo que todavía no tenemos. Lo que no tenemos es un elemento que llamo **RO** y que significa *Reducción Objetiva.* Es algo objetivo -bien

una cosa *o* la otra sucede objetivamente-. Es una teoría que nos falta. **RO** es un bonito acrónimo porque incluye la disyuntiva *o*[22], *y* esto es, de hecho, lo que sucede, uno (R)**O** el otro.

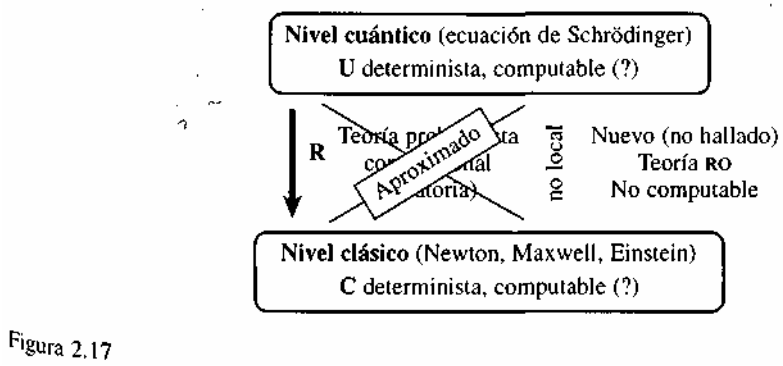


Figura 2.17

Pero, ¿cuándo tiene lugar este proceso? El punto de vista que estoy proponiendo es que algo falla en el principio de superposición cuando se aplica a *geometrías espacio-temporales.* Encontramos la idea de geometrías espacio-temporales en el capítulo I y represento dos de ellas en la figura 2.18.a. Además, he representado en la figura la superposición de estas dos geometrías espacio-temporales, exactamente como hicimos para la superposición de partículas y fotones. Cuando uno siente que está obligado a considerar superposiciones de espacio-tiempos diferentes, surgen montones de problemas debido a que los conos de luz de los dos espacio-tiempos pueden apuntar en direcciones diferentes. Este es uno de los grandes problemas a los que nos enfrentamos cuando tratamos, en serio, de cuantizar la relatividad general. Tratar de hacer física *dentro de* tan extraño tipo de espacio-tiempo superpuesto es algo en lo que, en mi opinión, ha fracasado todo el mundo hasta ahora.

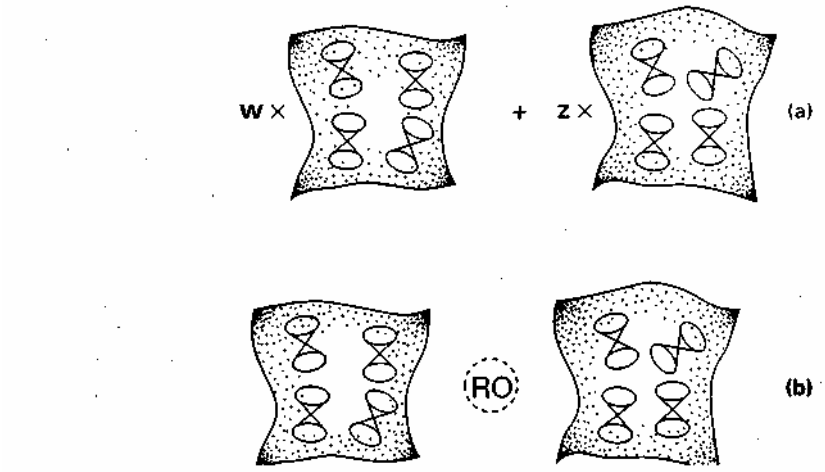


Figura 2.18.

Lo que estoy afirmando es que existen buenas razones para este fracaso general: simplemente porque no es esto lo que uno debería estar haciendo. De algún modo, esta superposición se convierte realmente en uno (**R**)**O** el otro y es en el nivel del espacio-tiempo donde esto sucede (figura 2.18.b). Ahora, ustedes podrían decir: «Todo esto está muy bien en principio, pero cuando usted trata de combinar la mecánica cuántica y la relatividad general llega a estos números ridículos, el tiempo y la longitud de Planck, que son muchos órdenes de magnitud menores que los tipos normales de longitudes y tiempos con los que trabajamos incluso en física de partículas. Esto no tiene nada que ver con cosas en la escala de gatos o personas. Entonces, ¿qué puede tener que ver la gravedad cuántica con ello?». Creo que tiene mucho que ver debido a la naturaleza fundamental de lo que está sucediendo.

¿Cuál es la relevancia de la longitud de Planck (10^{33} cm) para la reducción del estado cuántico? La figura 2.19 es una imagen muy esquemática de un espacio-tiempo que está

[2] "En el inglés original el acrónimo es OR (Objective Reduction) **que** coincide con la conjun-'°n disyuntiva «or». *(N. del T.)*

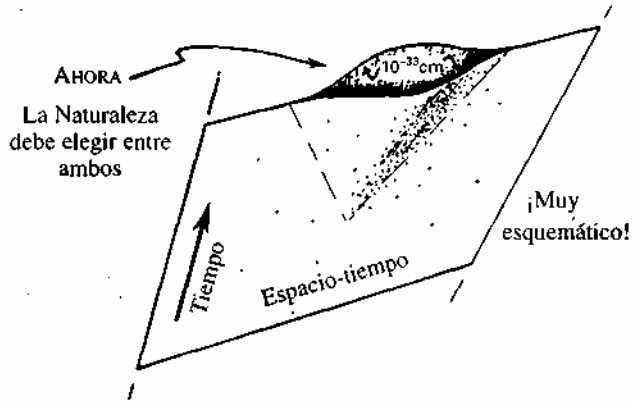


Figura 2.19. ¿Qué importancia tiene la escala de Planck de 10^{-23} cm para la reducción del estado cuántico? Idea aproximada: cuando entre los dos estados en superposición haya movimiento de masa suficiente para que los dos espacio-tiempos resultantes difieran en algo del orden de 10^{-23} cm.

tratando de bifurcarse. Hay una situación que conduce a una superposición de dos espacio-tiempos, uno de los cuales podría representar al gato muerto y el otro al gato vivo y, de algún modo, podría parecer que estos dos espacio-tiempos diferentes tienen que estar superpuestos. Tenemos que preguntar, «¿Cuándo llegan a ser suficientemente diferentes para que pudiéramos plantearnos un posible cambio de las reglas?» Hay que tratar de ver en qué momento la diferencia entre estas geometrías es, en un sentido apropiado, del orden de la longitud de Planck. Cuando las geometrías empiezan a diferir en dicha cantidad uno tiene que procuparse de qué hacer y es entonces cuando podrían cambiar las reglas. Debería hacer hincapié en que estamos tratando aquí con espacio-tiempos y no solo espacios. Para una *separación espacio-temporal de una escala de Planck*, una pequeña separación espacial corresponde a un tiempo mayor, y una mayor separación espacial, a un tiempo más corto. Lo que necesitamos es un criterio que nos permita estimar cuándo dos espacio-tiempos difieren significativamente, y esto conducirá a una *escala de tiempo* para la elección entre ambos por parte de la Naturaleza. Así pues, este punto de vista afirma que la Naturaleza escoge uno u otro de acuerdo con alguna regla que todavía no entendemos.

¿Cuánto tiempo necesita la Naturaleza para hacer esta elección? Podemos calcular dicha escala de tiempo en ciertas situaciones muy claras, cuando la aproximación newtoniana a la teoría de Einstein es suficiente y cuando existe una diferencia claramente definida entre los dos campos gravitatorios que están siendo sometidos a superposición cuántica (siendo aproximadamente iguales en magnitud las dos amplitudes complejas involucradas). La respuesta que estoy sugiriendo es la siguiente. Voy a reemplazar el gato por una masa -el gato ha tenido mucho trabajo que hacer y merece un descanso-. ¿Qué tamaño tiene la masa, cuánto tiene que moverse, y cuál es la escala de tiempo resultante para que ocurra el colapso del vector de estado (figura 2.20)? Voy a considerar la superposición de

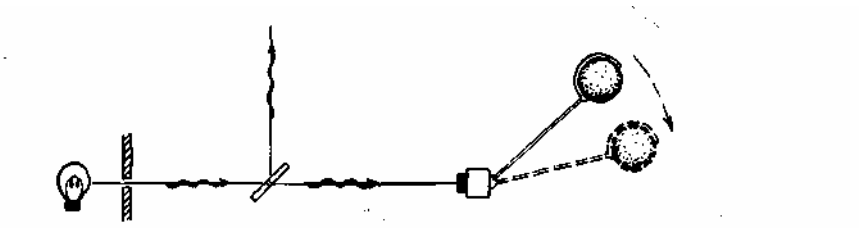


Figura 2.20. En lugar de tener un gato, la medida podría consistir en el simple movimiento de una masa esférica. ¿Qué tamaño o qué valor debe tener la masa?, ¿cuánto debe moverse?, ¿cuánto tiempo puede durar la superposición antes de que **R** tenga lugar?

un estado más el otro como un estado inestable -es un poco como una partícula, un núcleo de uranio o algo semejante, que se desintegra, y que podría desintegrarse en una cosa u otra, y hay una cierta escala de tiempo asociada con esa desintegración-. Que sea inestable es una hipótesis, pero esta inestabilidad va a ser una consecuencia de la física que aún no entendemos. Para calcular la escala de tiempo, consideremos la energía E que se necesitaría para desplazar una copia de la masa lejos del campo gravitatorio de otra copia. Entonces, tomamos *h*, la constante de Planck dividida por *2p*, *y* la dividimos por esta energía gravitatoria, y esta va a ser la escala de tiempo T para la desintegración en dicha situación.

$$T = h/E$$

Hay muchos esquemas que siguen este tipo general de razonamiento: todos los esquemas gravitatorios, aunque pueden diferir en detalle.

Hay otras razones generales para creer que un esquema gravitatorio de esta naturaleza podría ser algo a tener en cuenta. Una de estas es que todos los demás esquemas explícitos para la reducción del estado cuántico, que intentan resolver el problema de la medida cuántica introduciendo algunos fenómenos físicos nuevos, se enfrentan a dificultades con la conservación de la energía. Se encuentra que las reglas normales de la conservación de la energía tienden a ser violadas. Quizá sea así realmente, pero creo que, si adoptamos un esquema gravitatorio, hay una excelente oportunidad de que podamos ser capaces de evitar por completo este problema. Aunque yo no sé cómo hacer esto en detalle, permítanme exponer lo que tengo en mente.

En la relatividad general, masa y energía son conceptos más bien extraños. Ante todo, masa es igual a energía (dividida por la velocidad de la luz al cuadrado) y, por consiguiente, la energía potencial gravitatoria contribuye (negativamente) a la masa. En consecuencia, si ustedes tienen dos masas que están separadas, el sistema global posee una masa ligeramente mayor que si estuvieran juntas (figura 2.21). Aunque las densidades de masa-energía (medidas por el tensor energía-momento) son solo diferentes de cero dentro de las propias masas, y la cantidad en cada una de ellas no depende significativamente de la presencia de la otra masa, hay una diferencia entre las energías *totales* en cada uno de los dos casos ilustrados en la figura 2.21. La energía total es una magnitud no local. Hay, de hecho, algo esencialmente no local con respecto a la energía en la relatividad general, Este es ciertamente el caso del famoso ejemplo del pulsar binario, que mencioné en el capítulo 1: las ondas gravitatorias se llevan energía positiva y masa del sistema, pero esta energía reside de forma no local en todo el espacio. La energía gravitatoria es evasiva. Creo que si dispusiéramos de la forma correcta de combinar la relatividad general con la mecánica cuántica, dispondríamos de una buena oportunidad para evitar las dificultades con la energía que plagan las teorías del colapso del vector de estado. La cuestión es que, en el estado superpuesto, uno tiene que tener en cuenta la contribución gravitatoria a la energía en la superposición. Pero no se puede conceder un sentido local a la energía debida a la gravedad y, por ello, se produce una incertidumbre básica en la energía gravitatoria y dicha incertidumbre es del orden de la energía E descrita más arriba. Este es precisamente el tipo de cosas que uno se encuentra en las partículas inestables. Una partícula inestable tiene una incertidumbre en su masa-energía que está relacionada con su tiempo de vida por esta misma fórmula.

Déjenme concluir examinando las escalas de tiempo explícitas que aparecen en la aproximación que estoy proponiendo -volveré a esto en el capítulo 3-. ¿Cuáles son los tiempos de desintegración para sistemas reales en los que tienen lugar estas superposiciones espacio-temporales? Para un protón (provisionalmente considerado como una esfera rígida), la escala de tiempo es de algunos millones de años. Eso es bueno, ya que sabemos, por los experimentos de interferometría con partículas individuales, que no suceden estas cosas. Así pues, esto es consistente. Si uno toma una gotita de agua con un radio, digamos, de 10^{-5} cm, el tiempo de desintegración sería de algunas horas; si tuviera un radio de una micra, este tiempo sería de un veinteavo de segundo y, si fuera de una milésima de centímetro, necesitaría aproximadamente una millonésima de segundo. Estos números indican el rango de escalas sobre las que podría llegar a ser importante este tipo de física.

Hay, sin embargo, otro ingrediente esencial que tengo que traer aquí a colación. Quizá me estuve burlando ligeramente del punto de vista PTPP, pero un elemento de

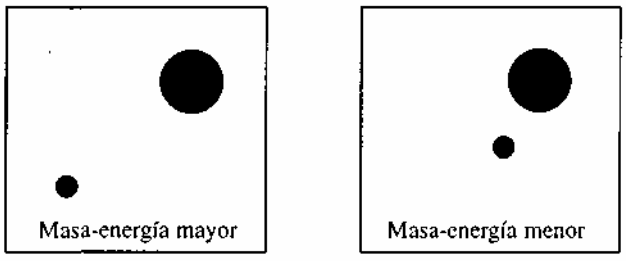


Figura 2.21. La masa-energía total de un sistema gravitante incluye contribuciones puramente gravitatorias que no son localizables.

dicha imagen tiene que ser tomado muy en serio: se trata del entorno. El entorno es vital en estas consideraciones y yo lo he ignorado hasta ahora en mi discusión. Por tanto, hay que hacer algo mucho más elaborado. Hay que considerar no solo la masa *aquí* superpuesta con la masa *allí* sino la masa y su entorno superpuestos con la otra masa y su entorno. Hay que tratar de ver con cuidado si el efecto principal está en la perturbación del entorno o en el movimiento de la masa. Si está en el entorno, el efecto va a ser aleatorio, y no tendremos nada diferente de los procedimientos estándar. Si el sistema puede aislarse lo suficiente para que el entorno no esté involucrado, entonces podríamos ver algo diferente de la mecánica cuántica estándar. Sería muy interesante saber si pueden sugerirse experimentos plausibles -y yo conozco varias posibilidades experimentales- que pudieran verificar si este tipo de esquema es verdadero en la naturaleza o si la mecánica cuántica convencional sobrevive una vez más y realmente hay que considerar que estas masas -o incluso los gatos- deben persistir en tales estados superpuestos.

Permítanme resumir en la figura 2.22 qué es lo que hemos estado tratando de hacer. En esta figura he situado las diferentes teorías en los vértices de un cubo distorsionado. Los tres ejes del cubo corresponden a tres de las constantes fundamentales de la física: la constante gravitatoria G (eje horizontal), la velocidad de la luz tomada en su forma inversa c^{-1} (eje diagonal), y la constante de Planck-Dirac, h (eje vertical hacia abajo). Cada una de estas constantes es minúscula en términos ordinarios y puede considerarse nula con una buena aproximación. Si consideramos las tres nulas, tenemos lo que llamo física galileana (arriba a la izquierda). La inclusión de una constante gravitatoria no nula nos lleva horizontalmente hacia la teoría gravitatoria newtoniana (cuya formulación espacio-temporal geométrica fue dada mucho más tarde por Cartan). Si, en su lugar, permitimos que c^{-1} sea distinto de cero, llegamos a la teoría especial de la relatividad de Poincaré-Einstein-Minkowski. El *cuadrado* superior de nuestro cubo distorsionado queda completo si permitimos que ambas constantes sean diferentes de cero, y así se obtiene la teoría general de la relatividad de Einstein. Sin embargo, esta generalización no es en absoluto sencilla -y he ilustrado este hecho en la figura 2.22 mediante las distorsiones en el cuadrado superior-. Permitiendo que h sea distinto de cero pero, por el momento, haciendo de nuevo $G = c^{-1} = 0$, obtenemos la mecánica cuántica estándar. Mediante una generalización no totalmente directa, c^{-1} puede ser incorporada y con ello se llega a la teoría cuántica de campos. Esto completa la cara izquierda del cubo, cuyas pequeñas distorsiones indican la falta de directividad.

Ustedes podrían pensar que todo lo que tenemos que hacer ahora es completar el cubo y así conoceríamos todo. Sin embargo, resulta que los principios de la física gravitatoria están en conflicto fundamental con los de la mecánica cuántica. Esto se manifiesta incluso en la gravedad newtoniana (donde mantenemos $c^{-1} = 0$) cuando utilizamos el marco geométrico apropiado (Cartan), en el que se utiliza el *principio de equivalencia de Einstein* (según el cual los campos gravitatorios uniformes son indistinguibles de las aceleraciones). Esto me fue señalado por Joy Christian, quien también me proporcionó la inspiración que hay tras mi figura 2.22. Por el momento, no hay ninguna unión apropiada entre la mecánica cuántica y la gravedad newtoniana -que tome en cuenta

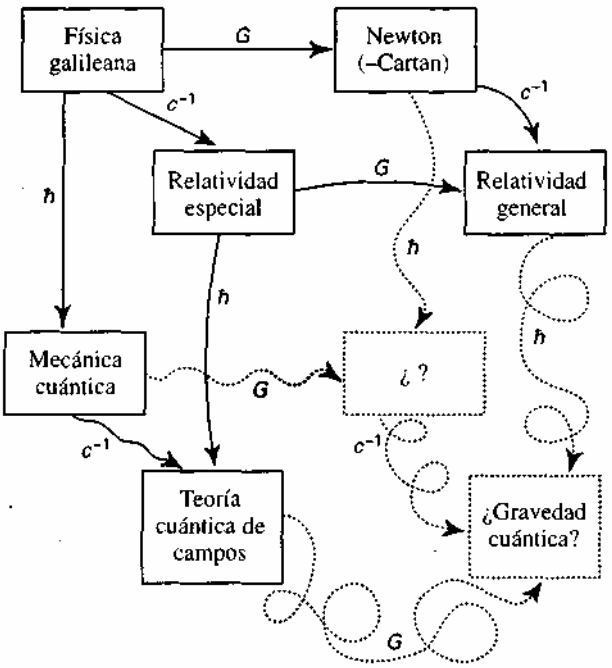


Figura 2.22.

enteramente el principio de equivalencia de Einstein, como hace la geometría de Cartan en la teoría clásica. En mi opinión, esta unión tendría que acomodar el fenómeno de la *reducción de estado cuántico* -aproximadamente según las líneas de las ideas **RO** subrayadas antes en este capítulo-. Tal unión estaría muy lejos de completarse en la cara trasera del cubo en la figura 2.22. La teoría global, que incorpora las tres constantes: h, G y c^{-1}, y con la que se completa el *cubo* entero, tendría que ser incluso más sutil y sofisticada matemáticamente. Evidentemente, este es un tema para el futuro.

3
La física y la mente

Los dos primeros capítulos se ocupaban del mundo físico y de las reglas matemáticas que utilizamos para describirlo, de lo notablemente precisas que son y lo extrañas que parecen a veces. En este tercer capítulo hablaré acerca del *mundo mental* y, en particular, sobre la relación que guarda con el mundo físico. Supongo que el obispo Berkeley habría pensado que, en cierto sentido, el mundo físico emerge de nuestro mundo mental, mientras que el punto de vista científico más general consiste en que la mentalidad es, de algún modo, una característica de algún tipo de estructura física.

Popper introdujo un tercer mundo llamado el *mundo de la cultura* (figura 3.1). Veía este mundo como un producto de la mentalidad y, por ello, utilizaba una jerarquía de números como se ilustra en la figura 3.2. En esta imagen, el mundo mental está relacionado en cierta manera -¿emerge de?- con el mundo físico y, de algún modo, la cultura surge de la mentalidad.

Ahora quiero considerar las cosas de manera un poco diferente. Más que entender la cultura como algo que surge de nuestra mentalidad, como hizo Popper, prefiero creer



Figura 3.1. *Mundo III* de Karl Popper.

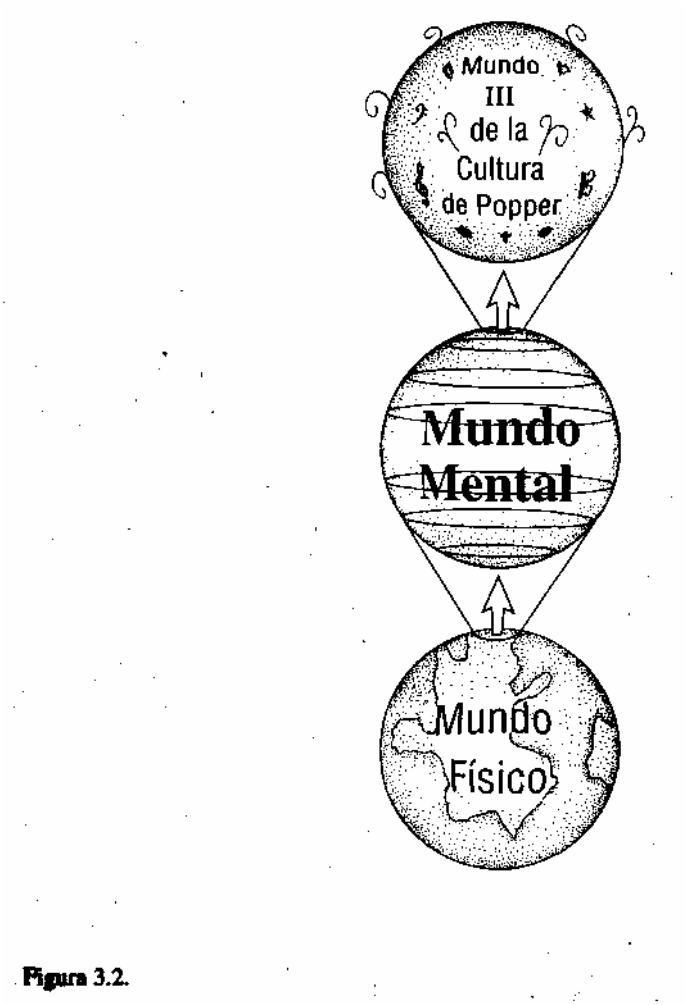


Figura 3.2.

que los mundos están conectados como se muestra en la figura 3.3. Más aún, mi «Mundo III» no es en realidad el mundo de la cultura sino el mundo platónico de las ideas -en particular la verdad matemática absoluta-. De esta forma, la disposición de la figura 1.3, que ilustra la dependencia profunda del mundo físico con respecto a leyes matemáticas exactas, está incorporada en nuestra ilustración.

Gran parte de este capítulo estará dedicada a la relación que se establece entre todos estos mundos diferentes. Me parece que hay un problema fundamental con la idea de que la mentalidad surge de lo físico: es algo que preocupa a los filósofos por muy buenas razones. Los conceptos que tratamos en física son materia, objetos masivos, partículas, espacio, tiempo, energía, cosas físicas y demás. ¿Cómo puede ser que nuestras sensaciones, nuestra

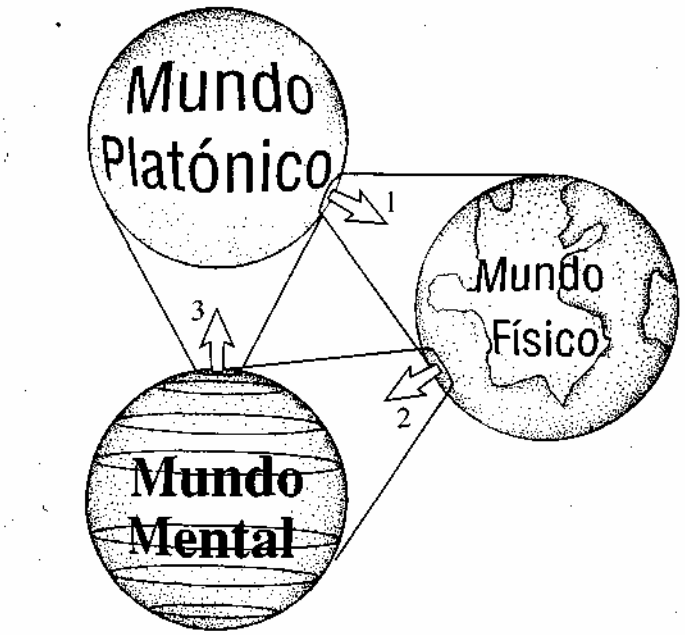


Figura 3.3. Tres Mundos y tres Misterios.

percepción de la rojez, o de la felicidad, tengan algo que ver con la física? Creo que esto es un misterio. Podemos considerar las flechas que conectan los diferentes mundos en la figura 3.3 como misterios. En los dos primeros capítulos discutí la relación que existe entre las matemáticas y la física (Misterio 1), y cité el comentario de Wigner con respecto a esta relación. La consideraba muy extraordinaria y yo también lo hago. ¿A qué se debe que el mundo físico parezca obedecer leyes matemáticas de una forma extremadamente precisa? Y no solo eso, sino que las matemáticas que parecen tener el control de nuestro mundo físico son excepcionalmente fructíferas y potentes simplemente *como* matemáticas. Creo que esta relación es un profundo misterio.

En este capítulo examinaré el Misterio 2: el misterio de la relación entre el mundo físico y el mundo de la mentalidad. Pero, en relación con esto, tendremos también que considerar el Misterio 3: ¿qué subyace a nuestra capacidad de acceder a la verdad matemática? Cuando me referí al mundo platónico en los dos primeros capítulos, estaba hablando principalmente acerca de las matemáticas y de los conceptos matemáticos que uno tiene que invocar para describir el mundo físico. Uno tiene la sensación de que las matemáticas que son necesarias para describir estos problemas están ahí fuera. Existe también, sin embargo, la sensación de que estas construcciones matemáticas son productos de nuestra mentalidad, es decir, que las matemáticas son un producto de la mente humana. Uno puede pensar las cosas de este modo, pero no es así realmente como el matemático considera la verdad matemática; y tampoco es esta mi forma de verla. Por ello, aunque hay una flecha que une el mundo mental y el mundo platónico, no pretendo indicar que esta, o de hecho cualquiera de estas flechas, significa que uno de estos mundos emerge simplemente de los otros. Aunque quizá en algún sentido están emergiendo, las flechas pretenden simplemente representar el hecho de que existe una relación entre los diferentes mundos.

Más importante es el hecho de que la figura 3.3 representa tres prejuicios propiamente míos. Uno de ellos es que el mundo físico entero puede, en principio, describirse en términos matemáticos. No estoy diciendo que todas las matemáticas puedan ser utilizadas para describir la física. Lo que afirmo es que, si uno escoge las partes correctas de las matemáticas, estas describen el mundo físico de forma muy precisa y, por ello, el mundo físico se comporta de acuerdo con las matemáticas. Así pues, existe una pequeña parte del mundo platónico que engloba nuestro mundo físico. Tampoco estoy diciendo que todos los elementos del mundo físico tengan su correspondencia en el mundo mental; más bien estoy sugiriendo que no existen objetos mentales flotando ahí fuera que no estén basados en la física. Este es mi segundo prejuicio. Existe un tercer prejuicio según el cual, en mi visión de las matemáticas, al menos en principio, cualquier objeto individual en el mundo platónico es accesible, en cierto sentido, a nuestra mentalidad. Algunas personas podrían preocuparse por este tercer prejuicio -de hecho, pueden sentirse inquietos por los tres prejuicios-. Tengo que decir que, solo después de haber dibujado este diagrama, me di cuenta de que reflejaba estos tres prejuicios personales míos. Volveré a él al final del capítulo.

Permítanme decir algo ahora sobre la *conciencia humana.* En particular, ¿es esta una cuestión sobre la que deberíamos pensar en términos de explicación científica? Mi punto de vista personal es decididamente que sí deberíamos hacerlo. En particular, tomo muy en serio la flecha que une el mundo físico con el mundo mental. En otras palabras, estamos ante el reto de comprender el mundo mental en términos del mundo físico.

He resumido algunas características del mundo físico y el mundo mental en la figura 3.4. En el lado derecho tenemos aspectos del *mundo físico:* este es percibido como si estuviera gobernado por leyes físicas y matemáticas precisas, como se discutió en los dos primeros capítulos. En el lado izquierdo situamos la conciencia, que pertenece al *mundo mental,* y palabras como *alma, espíritu, religión* y demás, aparecen con frecuencia. En la actualidad, se prefieren explicaciones científicas para las cosas. Más aún, se tiende a pensar que sería posible, en principio, introducir cualquier descripción científica en un ordenador. En consecuencia, si uno tiene un modelo matemático de algo, debería ser capaz, en principio, de programarlo en un ordenador. Esto es algo *contra lo que argumentaré enérgicamente* en este capítulo, a pesar de mi *sesgo fisicalista.*

Los términos utilizados para describir las leyes físicas en la figura 3.4 son *predecible, calculable* - estos tienen que ver con el hecho de si existe o no *deteriminismo* en nuestras leyes físicas y si podríamos o no utilizar un ordenador para simular la acción de estas leyes-. Por un lado, existe la opinión de que cualidades mentales como la emoción, la estética, la creatividad, la inspiración y el arte son ejemplos de cosas que serían difíciles de ver emergiendo de algún tipo de descripción computacional. En el extremo *científico* opuesto, algunas personas dirían: «nosotros somos simples ordenadores; quizá no sepamos cómo describir estas cosas aún, pero, de algún modo, si supiéramos el tipo correcto de cálculos que habría que realizar, seríamos capaces de describir todas las cualidades mentales enumeradas en la figura 3.4». A menudo se utiliza la palabra *emergencia* para describir este proceso. Estas cualidades *emergen,* según estas personas, como resultado del tipo correcto de actividad computacional.

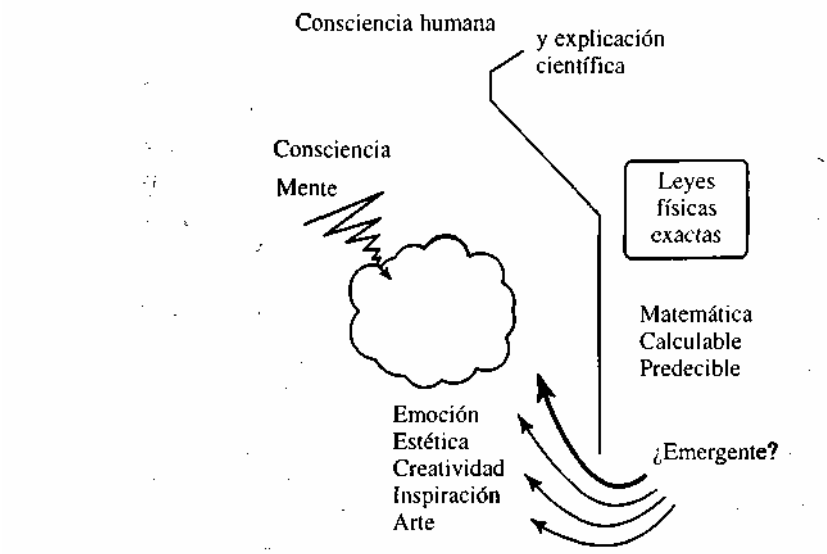


Figura 3.4.

¿Qué es la *conciencia* No sé cómo definirla. Pienso que no es este el momento de intentar definir la conciencia, puesto que no sabemos qué es. Creo que es un concepto físicamente accesible; pese a todo, definirlo sería probablemente definir la cosa equivocada. Sin embargo, voy a describirla hasta cierto grado. Me parece que hay al menos dos aspectos diferentes en la conciencia. Por un lado, existen manifestaciones *pasivas* de la conciencia, que implican *conocimiento.* Utilizo esta categoría para incluir elementos como la percepción del color, de la armonía, el uso de la memoria, y así sucesivamente. Por otro lado, están sus manifestaciones *activas,* que implican conceptos como el libre albedrío y la realización de acciones voluntarias y conscientes. La utilización de tales términos refleja aspectos diferentes de nuestra conciencia.

Aquí me concentraré principalmente en aquello que implica la conciencia de un modo esencial. Es diferente de los aspectos pasivo y activo de la conciencia, y quizá está en algún lugar entre ambos. Me refiero al uso del término *comprensión,* o quizá *intuición directa,* que a menudo es una palabra más adecuada. No voy a definir ninguno de estos términos; no sé lo que significan. Existen otros dos términos que tampoco entiendo: *conocimiento* e *inteligencia.* Entonces, ¿por qué estoy hablando sobre cosas que no

sé lo que significan realmente? Probablemente se debe a que soy un matemático y los matemáticos no pensamos tanto acerca de estos conceptos. No necesitamos definiciones precisas de los elementos que estamos manejando, con tal de que podamos decir algo sobre las *relaciones* entre ellos. El primer punto clave aquí es que opino que la inteligencia es algo que requiere comprensión. Utilizar el término inteligencia en un contexto en el que negamos que esté presente cualquier comprensión me parece poco razonable. De un modo análogo, comprender sin ningún conocimiento es también algo absurdo: comprender requiere algún tipo de conocimiento. Este es el segundo punto clave. De ello se deduce que la inteligencia requiere conocimiento. Aunque no estoy definiendo ninguno de estos términos, me parece razonable insistir sobre estas relaciones entre ellos.

Se podrían adoptar diversos puntos de vista acerca de la relación entre pensamiento consciente y computación. He resumido en la tabla 3.1 cuatro aproximaciones al conocimiento, que he designado A, B, C y D

Tabla 3.1.

A	Todo pensamiento es computación; en particular, las sensaciones de conocimiento conciente son producidas por la mera ejecución de cálculos apropiados.
B	La conciencia es una característica de la acción física del cerebro; y, mientras que cualquier acción física puede ser simulada computacionalmente, la simulación computacional no puede por sí misma producir conciencia.
C	La acción física apropiada del cerebro produce conciencia, pero esta acción física no puede siquiera ser propiamente simulada computacionalmente.
D	La conciencia no puede ser explicada en términos físicos, computacionales, o cualesquiera otros términos científicos.

El punto de vista que llamo A, a veces denominado *inteligencia artificial fuerte* (IA fuerte) *o funcionalismo* (computacional), afirma que todo pensamiento es simplemente la realización de ciertos cálculos y, en consecuencia, si uno realiza los cálculos apropiadas, el resultado será el conocimiento.

He designado como B al segundo punto de vista y, según este, en principio, sería posible simular la actividad de un cerebro cuando su propietario es conocedor de algo. La diferencia entre A y B es que, aunque dicha actividad puede ser simulada, esa mera simulación no podría en sí misma, según B, tener ningún sentimiento o ningún conocimiento -hay algo más, que quizá tiene que ver con la construcción física del objeto-. Por eso, un cerebro compuesto de neuronas y similares podría ser conocedor, mientras que una simulación de la actividad de dicho cerebro no sería conocedora. Este es, hasta donde puedo alcanzar, el punto de vista defendido por John Searle.

A continuación está mi propia manera de ver las cosas, que he llamado C. Estoy de acuerdo con B, hay algo en la actividad física del cerebro que provoca conocimiento -en otras palabras, hay algo en la física a lo que tenemos que dirigirnos-, pero esta actividad física es algo que ni siquiera puede ser simulado computacionalmente. No hay ninguna simulación que pudiera realizarse de dicha actividad. Esto requiere que exista algún factor en la actividad física del cerebro que esté más allá de la computación.

Finalmente, siempre existe el punto de vista D, según el cual es un error considerar estas cuestiones en términos de ciencia. Quizá el conocimiento no pueda ser explicado en términos científicos.

Yo soy un defensor del punto de vista C. Existen, sin embargo, diversas variedades de C. Está lo que podría denominarse C *débil y C fuerte. C débil* propone que, de algún modo, con la física conocida, bastaría solo con mirar con mucho cuidado para encontrar ciertos tipos de actividad que están más allá de la computación. Cuando digo «más allá de la computación», tengo que ser un poco más explícito, como lo seré en un instante. Según C *débil*, no hay nada que necesitemos buscar fuera de la física conocida para encontrar la actividad no computacional apropiada. C *fuerte*, por el contrario, requiere que exista algo fuera de la física conocida; nuestra comprensión física es inadecuada para la descripción del conocimiento. Es incompleta, y, como ustedes habrán deducido del capítulo 2, yo sí creo que nuestra imagen física es incompleta, como indiqué en la figura 2.17. Desde

el punto de vista de la *C fuerte* quizá la ciencia futura explicará la naturaleza de la conciencia, pero la ciencia actual no lo hace.

Incluí algunas palabras en la figura 2.17 que no comenté en ese momento, en particular, el término *computable.* En la imagen estándar, uno se encuentra con física básicamente computable en el nivel cuántico, y el nivel clásico es, probablemente, computable, aunque hay cuestiones técnicas sobre cómo pasar desde sistemas discretos computables a sistemas continuos. Es un punto importante pero permítanme que no me ocupe aquí de ello. De hecho, me parece que los defensores de la C *débil* tendrán que encontrar algo en estas incertidumbres, algo que no pueda explicarse en términos de una descripción computable.

Para pasar del nivel cuántico al nivel clásico en la imagen convencional, introducimos el procedimiento que he llamado R, y que es una operación completamente probabilista. Lo que tenemos entonces es computabilidad junto con aleatoriedad. Voy a argumentar que esto no es suficiente: necesitamos algo diferente, y esta nueva teoría, que construye un puente entre estos dos niveles, tiene que ser una teoría no computable. Diré algo más sobre lo que entiendo por este término en un momento.

Esta es mi versión de la *C fuerte:* buscamos la no computabilidad en la física que forma un puente entre los niveles cuántico y clásico, lo cual es pedir bastante. Estoy diciendo que no solo necesitamos una nueva física, sino que también necesitamos una nueva física que sea relevante para la actividad del cerebro.

En primer lugar, abordemos la cuestión de si es plausible o no que exista algo más allá de la computación en nuestra comprensión. Permítanme exponerles un bonito ejemplo de un sencillo problema de ajedrez. En la actualidad, los ordenadores juegan muy bien al ajedrez. Sin embargo, cuando el problema de ajedrez que se muestra en la figura 3.5 fue planteado al ordenador más potente disponible en su época, el ordenador Pensamiento Profundo, este hizo una jugada muy estúpida. En esta posición de ajedrez, las piezas blancas están superadas en número por las negras: existen dos torres negras extra y un alfil negro. Esto debería constituir una enorme ventaja, si no fuera por el hecho de que una barrera de peones protege a todas las piezas negras. Así, todo lo que las blancas tienen que hacer es pasearse detrás de su barrera de peones blancos, y no hay posibilidad de perder la partida. Sin embargo, cuando se le presentó esta posición a Pensamiento Profundo, instantáneamente capturó la torre negra, abrió la barrera de peones y quedó en una posición irremediablemente perdida. La razón de que hiciera eso es que había sido programado para calcular un cierto número de jugadas por anticipado y contar luego las piezas, o algo similar. En este ejemplo, ese no era el camino más adecuado. Por supuesto, si hubiera seguido calculando una jugada tras otra, y otra, algunas veces más, habría sido capaz de hacerlo. La cuestión es que el ajedrez es un juego computacional. En este caso, el jugador humano ve la barrera de peones y comprende que es impenetrable. El ordenador no tenía esa comprensión -simplemente computaba un movimiento tras otro-. Así pues, este ejemplo ilustra la diferencia que existe entre la mera computación y la cualidad de la comprensión.

He aquí otro ejemplo (figura 3.6). Es muy tentador capturar la torre negra con el alfil blanco, pero lo correcto es hacer como si el alfil blanco fuera un peón y utilizarlo para crear otra barrera de peones. Una vez que le hayamos enseñado a reconocer barreras de peones, el ordenador podrá ser capaz de resolver el primer problema pero fracasará en el segundo porque necesita un nivel de comprensión mayor. Ustedes podrían pensar, no obstante, que con suficiente cuidado sería posible programar todos los niveles de

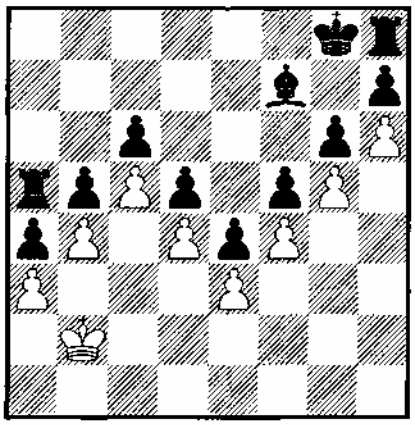


Figura 3.5. Blancas juegan y hacen tablas: fácil para los seres humanos, ¡pero Pensamiento Profundo capturó la torre! (Problema planteado por William Hartston, tomado de un artículo de Jane Seymore y David Norwood en *New Scientist*, n.º 1889, pág. 23, 1993).

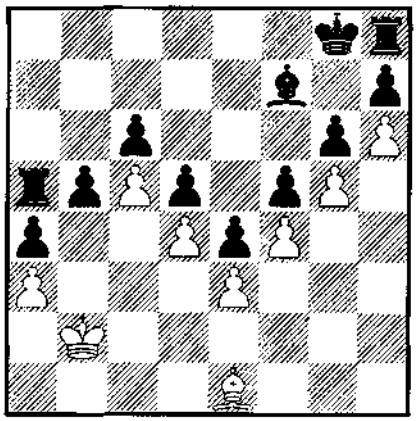


Figura 3.6. **Blancas juegan** y hacen tablas: de nuevo fácil para los seres humanos, **pero un** ordenador normal que **juegue al** ajedrez capturará la torre (de un test de Turing por **William Hartston** y **David** Norwood).

comprensión posibles. Bien, quizá pueda hacerse con el ajedrez. El problema está en que el ajedrez es un juego computacional y, por eso, en última instancia, sería posible calcular todas las posibilidades hasta el final con un ordenador suficientemente potente. Esto supera la capacidad de los ordenadores actuales pero, en principio, sería posible. De todas formas, uno tiene la sensación de que en la *comprensión* hay algo más que la computación directa. Ciertamente, el modo en que enfocamos estos problemas de ajedrez es muy diferente de como lo hace un ordenador.

¿Podemos construir un argumento más fuerte a favor de que hay realmente algo en nuestra comprensión que es diferente de la computación? Sí podemos. No quiero consumir mucho tiempo en este argumento, aunque es la piedra básica de toda la discusión. Pero tengo que dedicarle un poco de tiempo, aunque el argumento pueda hacerse algo técnico. Las docientas primeras páginas de *Las sombras de la mente* estaban dedicadas a tratar de mostrar que no hay cabos sueltos en el argumento que voy a ofrecerles.

Déjenme decir algo sobre *computaciones*. Las computaciones son lo que hace un ordenador. Los ordenadores reales tienen una capacidad de almacenamiento limitada pero voy a considerar un ordenador idealizado, una *máquina de Turing*, que difiere de un ordenador ordinario de propósito general solo en el hecho de que posee un espacio de almacenamiento ilimitado y puede seguir computando indefinidamente sin cometer errores y sin gastarse siquiera. Permítanme dar un ejemplo de computación. Una computación no tiene por qué involucrar simple aritmética sino que puede incluir también la realización de operaciones lógicas. Veamos un ejemplo:

• Encontrar un número que no sea la suma de tres números cuadrados.

Por un número entiendo un *número natural,* tal como 0, 1,2, 3, 4, 5..., y por *números cuadrados,* los números 0^2, 1^2, 2^2, 3^2, 4^2, 5^2... He aquí cómo debería hacerlo usted: es una manera muy estúpida, pero ilustra lo que podemos entender por una computación. Empezamos con 0 y comprobamos si es la suma de tres números cuadrados. Buscamos todos los cuadrados que son menores o iguales que 0 y solo encontramos 0^2. Por consiguiente, solo podemos ensayar: $0 = 0^2 + 0^2 + 0^2$; que resulta ser cierto y, por ello, 0 es la suma de tres cuadrados. A continuación ensayamos con 1. Escribimos todas las formas posibles de sumar todos los números cuyos cuadrados son menores que, o igual a uno y vemos si podemos sumar tres de ellos para obtener 1. Podemos hacerlo: $1 = 0^2 + 0^2 + 1^2$. Podemos seguir de esta forma más bien tediosa, como se indica en la tabla 3.2, hasta que llegamos al número 7, donde podemos comprobar que no existe ninguna forma de sumar tres cuadrados de O^2, 1^2 y 2^2 en cualquier combinación para que resulte el número 7 —todas las posibilidades se muestran en la tabla-. Así pues, 7 es la respuesta: es el número más pequeño que no es la suma de tres números cuadrados. Este ha sido un ejemplo de un cálculo.

Tabla 3.2

Ensayar 0	los cuadrados ≤ 0 son	0^2	$0 = 0^2 + 0^2 + 0^2$
Ensayar 1	los cuadrados ≤ 1 son	0^2, 1^2	$1 = 0^2 + 0^2 + 1^2$
Ensayar 2	los cuadrados ≤ 2 son	0^2, 1^2	$2 = 0^2+1^2+1^2$
Ensayar 3	los cuadrados ≤ 3 son	0^2,1^2	$3 = 1^2+1^2+1^2$
Ensayar 4	los cuadrados ≤ 4 son	0^2, 1^2, 2^2	$4 = 0^2 + 0^2 + 2^2$
Ensayar 5	los cuadrados ≤ 5 son	0^2, 1^2, 2^2	$5 = 0^2+l^2 + 2^2$
Ensayar 6	los cuadrados ≤ 6 son	0^2, 1^2, 2^2	$6 = 1^2 + 1^2 + 2^2$
Ensayar 7	los cuadrados ≤ 7 son	0^2, 1^2, 2^2	$7 \neq 0^2 + 0^2 + 0^2$
			$7 \neq 0^2 + 0^2 + 1^2$
			$7 \neq 0^2 + 0^2 + 2^2$
			$7 \neq 0^2 + 1^2 + 1^2$
			$7 \neq 0^2 + 1^2 + 2^2$
			$7 \neq 1^2 + 1^2 + 1^2$
			$7 \neq l^2 + 1^2 + 2^2$
			$7 \neq 1^2 + 2^2 + 2^2$
			$7 \neq 2^2 + 2^2 + 2^2$

En este ejemplo hemos tenido suerte porque el cálculo llegó a término, mientras que existen ciertos cálculos que no terminan nunca. Por ejemplo, supongamos que modifico ligeramente el problema:

• Encontrar un número que no sea la suma de cuatro números cuadrados.

Existe un famoso teorema debido a Lagrange, un matemático del siglo XVIII, que demuestra que todo número puede expresarse como la suma de cuatro cuadrados. Por tanto, si procedemos de un modo mecánico para encontrar dicho número, el ordenador, simplemente, seguirá funcionando sin parar y nunca encontrará una respuesta. Esto ilustra el hecho de que existen algunas computaciones que no terminan.

La demostración del teorema de Lagrange es bastante complicada, así que he aquí otro teorema más fácil que, ¡espero todos puedan apreciar!

- ***Encontrar un número impar que sea la suma de dos números pares.***

Ustedes podrían programar su ordenador para llevarlo a cabo y el ordenador seguiría funcionando incesantemente porque sabemos que, cuando sumamos dos números pares, siempre obtenemos un número par.

Hay aquí un ejemplo esencialmente más complicado:

- ***Encontrar un número par mayor que 2 que no sea la suma de dos números primos.***

¿Termina alguna vez esta computación? La creencia general es que no lo hace, pero esto es una mera conjetura, conocida como la *conjetura de Goldbach, y* es tan difícil que nadie sabe con certeza si es verdadera o no. Así pues, hay (probablemente) tres cálculos que no se detienen: uno fácil, uno difícil y un tercero que es tan difícil que nadie sabe todavía si realmente se detiene o no.

Planteémonos ahora la siguiente pregunta:

- ¿Están utilizando los matemáticos algún algoritmo computacional (llamémosle A) para convencerse de que ciertas computaciones no terminan?

Por ejemplo, ¿tenía Lagrange algún tipo de programa de ordenador en su cabeza, que finalmente le llevó a la conclusión de que todo número es la suma de cuatro cuadrados? Ustedes no necesitan siquiera ser Lagrange: ustedes simplemente tienen que ser alguien que pueda seguir el argumento de Lagrange. Nótese que yo no estoy interesado en la cuestión de la originalidad, sino solamente en la cuestión de la comprensión. Por eso es por lo que he expresado la pregunta de la forma anterior: *convencerse* significa *crear* comprensión.

El término técnico para un enunciado de la naturaleza de los que estamos considerando es el de Π_1-*sentencia*. Una Π_1-*sentencia* es una afirmación de que cierta computación especificada no termina. Para apreciar el argumento que sigue, solo necesitamos pensar en sentencias de esta naturaleza. Quiero convencerles de que no existe tal algoritmo A.

Para comprobar esto, necesito hacer una generalización. Tengo que hablar sobre computaciones que dependen de un número natural *n.* Veamos algunos ejemplos:

- ***Encontrar un número natural que no sea la suma de* n *números cuadrados.***

Hemos visto por el teorema de Lagrange que si *n* es cuatro, o más, la computación no tiene fin. Pero si n es menor o igual que tres, entonces sí se para. La siguiente computación es:

- ***Encontrar un número impar que sea la suma de n números pares.***

No importa cuál sea *n* -eso no va a ayudarle en absoluto-. El cálculo no se detiene para ningún valor de *n.* Para la extensión de la conjetura de Goldbach, tenemos:

- ***Encontrar un número par mayor que 2 que no sea la suma de tres o menos números primos.***

Si la conjetura de Goldbach es verdadera, entonces esta computación no se detendrá para ningún *n* (distinto de 0 y 1). En cierto sentido, cuanto mayor es *n*, más fácil es esto. De hecho, creo que existe un valor suficientemente grande de *n* para el que se sabe que la computación es *interminable.*

El punto importante es que estos tipos de computación dependen del número natural *n.* Esto es, de hecho, fundamental para el famoso argumento conocido como *argumento de Gödel.* Lo discutiré en una forma debida a Alan Turing, pero utilizaré su argumento de un modo ligeramente diferente. Si a ustedes no les gustan los argumentos matemáticos, pueden desconectar por un momento: lo importante es el resultado. En cualquier caso, el argumento no es muy complicado, ¡tan solo algo confuso!

Los cálculos que actúan sobre un número *n* son básicamente programas de ordenador. Ustedes pueden hacer una lista de programas de ordenador y asignar un número, digamos *p,* a cada uno de ellos. De este modo ustedes introducen en su ordenador de propósito general algún número *p* y el ordenador empieza a funcionar, realizando dicha computación pésima aplicada a cualquier número *n* que ustedes hayan seleccionado. El número *p* se escribe como subíndice en nuestra notación. Así pues, hago una lista de dichos programas de ordenador, o computaciones, que actúan sobre el número *n,* colocándolos uno detrás de otro.

$$C_0(n),\ C,(n),\ C_2(n),\ C_3(n)....C_p(n)...$$

Vamos a suponer que esta es una lista de todas las posibles computaciones *C(n) y* que podemos encontrar algún modo efectivo de ordenar estos programas de ordenador, de modo que el número *p* designe al *p-ésimo* programa en la ordenación. Entonces, $C_p(n)$ representa el *p-ésimo* programa aplicado al número natural *n.*

Supongamos ahora que disponemos de algún procedimiento computacional, o algorítmico (A) que puede actuar sobre un par de números *(p, n), y* tal que, cuando dicho procedimiento llega a un final, nos proporciona una demostración válida de que la computación *C(n)* no termina. El algoritmo A no siempre funcionará necesariamente, en el sentido de que puede haber algunas computaciones *C(n)* que sean interminables cuando A(*p,n)* tampoco termina. Pero quiero insistir en que A no comete errores y, por eso, si A(p,n) sí termina, $C_p(n)$ no lo hace. Tratemos de imaginar que los matemáticos humanos actúan de acuerdo con algún procedimiento computacional A cuando formulan (o siguen) cierta demostración matemática rigurosa de una proposición matemática (digamos, de una $\prod_1$-sentencia). Supongamos que también se les permite *conocer* qué es A y que ellos *creen* que es un procedimiento válido. Vamos a tratar de imaginar que A engloba *todos* los procedimientos a disposición de los matemáticos humanos para demostrar convincentemente qué computaciones no se detienen. El procedimiento A empieza considerando la letra *p* para seleccionar el programa de ordenador, y luego considera el número *n* para descubrir sobre qué número debe actuar. Entonces, si el procedimiento computacional A llega a término, ello implica que la computación *C(n)* no termina. Así pues:

si *A(p, n)* se para, entonces *C (n)* no se para. (1)

En esto consiste el trabajo de A: proporciona la forma de convencerse incuestionablemente de que ciertas computaciones no terminan.

Supongamos ahora que ponemos *p = n.* Esto puede parecer algo curioso**.** Es el famoso procedimiento conocido como *procedimiento diagonal de Cantor y* no hay nada erróneo en usarlo. Entonces llegamos a la conclusión de que:

;

si A(*n, n)* se para, entonces $C_n(n)$ no se para.

Pero ahora, A(*n, n)* depende solo de un número y, por ello, A(*n, n)* debe ser uno de los programas de ordenador *C(n),* puesto que la lista es exhaustiva para computaciones que actúan sobre una única variable *n.* Supongamos que el programa de ordenador que es idéntico a *A(n, n)* viene designado por *k.* Entonces,

$$A(n, n) = C_k(n).$$

Ahora hacemos *n - k y* encontramos que:

$$A(k, k) = C_k(k).$$

Entonces, consideramos el enunciado (1) y concluimos que: si $A(k, k)$ se para, entonces $C_k(k)$ no se para.

Pero $A(k, k)$ es lo mismo que $C_k(k)$. Por consiguiente, si $C_k(k)$ se para, entonces no se para. Eso significa que no se para: es pura lógica. Pero aquí está la trampa: esta computación particular no se para y, si creemos en A, entonces también debemos creer que $C_k(k)$ no se para. Pero A tampoco se para y, por tanto, no *sabe* que $C_k(k)$ no se para. Por consiguiente, el procedimiento computacional no puede, después de todo, englobar la totalidad del razonamiento matemático para decidir que ciertas computaciones no se paran -es decir, para establecer la verdad de $\prod_1$-sentencias-. Esto es lo esencial del argumento de Gödel-Turing en la forma en que lo necesito.

Ustedes puede cuestionarse la fuerza global de este argumento. Lo que afirma claramente es que la intuición matemática no puede ser codificada en forma de alguna computación de la que podamos saber que es correcta. Las personas discuten a veces sobre esto pero a mí me parece que es una implicación clara. Es interesante leer lo que Turing y Gödel decían acerca de este resultado. He aquí la afirmación de Turing:

> «En otras palabras, si se espera que una máquina sea infalible, no puede ser también inteligente. Existen varios teoremas que dicen casi exactamente eso. Pero estos teoremas no dicen casi nada sobre cuánta inteligencia puede ser mostrada si una máquina no tiene pretensiones de infalibilidad.»

Por eso, su idea era que los argumentos tipo Gödel-Turing pueden ser reconciliados con la idea de que los matemáticos son esencialmente ordenadores si los procedimientos algorítmicos de acuerdo con los que actúan, para discernir la verdad matemática, son básicamente inválidos. Podemos restringir la atención a enunciados aritméticos, por ejemplo, ($\prod_1$-sentencias, que constituyen un tipo muy restrictivo de enunciados. Creo que Turing pensaba que la mente humana utiliza algoritmos, pero que dichos algoritmos son simplemente erróneos -esto es, son realmente inválidos-. Yo encuentro que esta es una posición bastante poco plausible, especialmente porque uno no está interesado aquí en cómo podría obtener inspiración, sino simplemente en cómo podría seguir un argumento y comprenderlo. Me parece que la postura de Turing no es muy encomiable. Según mi esquema, Turing habría sido una persona del tipo A.

Veamos lo que pensaba Golee. En mi esquema, él era una persona del tipo D. Así, incluso si Turing y Gödel tenían la misma evidencia ante sí, ambos llegaban a conclusiones esencialmente opuestas. De todas formas, aunque Golee no creía que la intuición matemática pudiera reducirse a computación, no fue capaz de descartar rigurosamente esta posibilidad. He aquí lo que decía:

> «Por otra parte, sobre la base de lo que se ha demostrado hasta ahora, sigue siendo posible que pueda existir (e incluso sea empíricamente descubierta) una máquina capaz de demostrar teoremas que, de hecho, sea equivalente a la intuición matemática, pero que no puede probarse que sea así, ni puede probarse que produzca solo teoremas correctos de la teoría de números finitos.»

Su argumento era que existe una escapatoria para el uso directo del argumento de Gödel-Turing como refutación del computacionalismo (o funcionalismo), a saber: que los matemáticos podrían estar utilizando un procedimiento algorítmico que es válido pero del que no podemos saber con seguridad que es válido. Así pues, era la parte cognoscible la que Golee pensaba que constituía una escapatoria, y era la parte válida aquella por la que se decidió Turing.

Mi opinión es que ninguna de estas es un posible desarrollo del argumento. Lo que dice el teorema de Gödel-Turing es que si se encuentra que un procedimiento algorítmico (para establecer $\prod_1$-sentencias) es válido, entonces uno puede mostrar inmediatamente algo que se sale de ello. Pudiera ser que estemos utilizando un procedimiento algorítmico del que no puede saberse que es válido y podría haber algún tipo de dispositivo de aprendizaje que nos permita desarrollar esta capacidad. Estos temas, y muchos otros, son tratados *ad nauseam* en mi libro *Las sombras de la mente.* No quiero entrar aquí en estas digresiones, solo mencionaré dos puntos.

¿Cómo podría haber aparecido este supuesto algoritmo? En el caso de los seres humanos, presumiblemente tendría que haber sido resultado de la selección natural, o, en el caso de los robots, habría sido construido por IA (Inteligencia Artificial) deliberada. No entraré en detalle en estos argumentos sino que simplemente los ilustraré con dos dibujos de mi libro.

El primer dibujo tiene que ver con la selección natural (figura 3.7). Ustedes pueden ver que el matemático no está en una posición muy cómoda desde el punto de vista de la selección natural porque hay un tigre de dientes afilados dispuesto a saltar sobre él. Por el contrario, sus primos, en la otra parte del dibujo, están cazando mamuts, construyendo casas, recogiendo cosechas y otras labores. Estas actividades implican comprensión pero no son específicas de las matemáticas. Así pues, es posible que la cualidad de la comprensión fuera la causa por la que fuimos seleccionados, pero no es posible que lo fueran los algoritmos específicos para hacer matemáticas.



Figura 3.7. Para nuestros ancestros remotos, una capacidad específica **para hacer matemáticas sofisticadas difícilmente puede haber sido una ventaja selectiva, pero sí pudo haberlo sido perfectamente una** capacidad general para comprender.

El otro dibujo tiene que ver con la *construcción IA deliberada y* hay una pequeña historia en mi libro acerca de un experto en IA del futuro que mantiene una discusión con el robot (figura 3.8). El argumento completo dado en el libro es algo largo y complicado

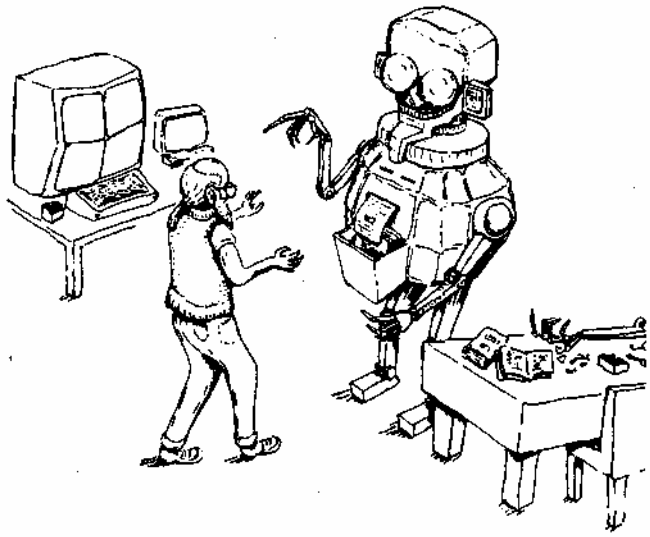


Figura 3.8. Ignacio Almirante se enfrenta al Joven Cibersistema Matemático. Las primeras 200 páginas de *Las sombras de la mente* están dedicadas a responder a las críticas a la utilización del argumento de Gödel-Turing. La esencia de estos nuevos argumentos está contenida en el diálogo entre la persona IA y su robot.

-no creo que sea necesario entrar aquí en ello-. Mi uso original del argumento de Gödel-Turing había sido atacado por todo tipo de personas desde todo tipo de ángulos diferentes y era necesario responder a todos esos ataques. Traté de resumir la mayoría de estos nuevos argumentos presentados en *Las sombras de la mente,* en la discusión que la persona IA mantiene con su robot.

Permítanme volver a la cuestión de lo que está pasando. El argumento de Gödel se refiere a enunciados particulares sobre números. Lo que Gödel nos dice es que ningún sistema de reglas computacionales puede caracterizar las propiedades de los *números naturales.* Pese al hecho de que no hay forma computacional de caracterizar los números naturales, cualquier niño sabe lo que son. Todo lo que hacemos es mostrar al niño números diferentes de objetos, como se ilustra en la figura 3.9, y, al cabo de cierto tiempo, puede abstraer la noción de número natural a partir de estos ejemplos concretos. No damos al niño un conjunto de reglas computacionales, lo que estamos haciendo es capacitar al niño para *comprender* qué son los números naturales. Yo diría que el niño es capaz de establecer algún tipo de *contacto* con el mundo platónico de las matemáticas. Algunas personas no son partidarias de esta manera de hablar sobre la intuición matemática pero, de todas formas, me parece que uno tiene que adoptar alguna posición de esta naturaleza acerca de lo que está sucediendo. De algún modo, los números naturales están ya *ahí,* existiendo en algún lugar del mundo platónico, y tenemos acceso a dicho mundo a través de nuestra capacidad para ser conocedores de cosas. Si fuéramos

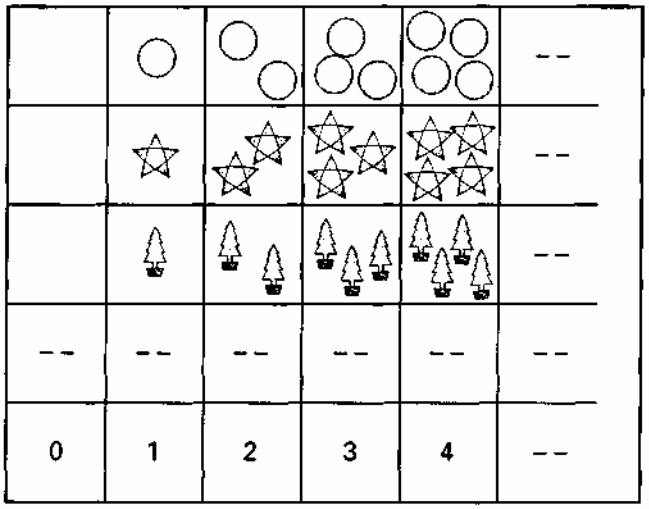


Figura 3.9. La noción platónica de un número natural puede ser abstraída por un niño a partir de tan solo unos pocos ejemplos simples.

simplemente ordenadores electrónicos, no tendríamos dicha capacidad. Las reglas no son los factores que nos capacitan para comprender la naturaleza de los números naturales, como muestra el teorema de Gödel. Comprender qué *son* los números naturales es un buen ejemplo del contacto con el mundo platónico.

Así pues, lo que estoy diciendo, con más generalidad, es que la comprensión matemática no es algo computacional, sino algo bastante diferente que depende de nuestra capacidad de ser conocedores de cosas. Algunas personas podrían decir: «Bien, todo lo que usted afirma haber demostrado es que la intuición matemática no es computacional. Eso no dice mucho sobre otras formas de conciencia». Pero me parece que esto ya es bastante. Es poco razonable trazar una línea entre la comprensión matemática y cualquier otro tipo de comprensión. Eso es lo que yo estaba tratando de ilustrar con mi primer dibujo (figura 3.7). La comprensión no es específica de las matemáticas. Los seres humanos desarrollan esta habilidad de comprensión general y *no* es una habilidad computacional porque la comprensión matemática no lo es. Tampoco trazo una línea divisoria entre la comprensión matemática y la conciencia humana en general. Por eso, aunque dije que no sé qué es la conciencia humana, me parece que la comprensión humana es un ejemplo de ella, o, al menos, es algo que la requiere. Tampoco voy a trazar una línea entre la conciencia humana y la conciencia animal. Aquí podría tener dificultades con diferentes conjuntos de personas. Me parece que los seres humanos son muy similares a muchos otros tipos de animales y, aunque podemos tener una comprensión de las cosas algo mejor que algunos de nuestros primos, en cualquier caso ellos también poseen algún tipo de comprensión, y por eso deben tener también conocimiento.

Por consiguiente, la no-computabilidad en *algún* aspecto de la conciencia y, específicamente, en la comprensión matemática, sugiere fuertemente que la no-computabilidad debería ser una característica de *toda* la conciencia. Esta es mi hipótesis.

Ahora bien, ¿qué entiendo por no-computabilidad? He hablado mucho sobre ello pero debería dar un ejemplo de algo que sea no-computacional para mostrar qué quiero decir. Lo que voy a describirles es un ejemplo de lo que a menudo se denomina un *modelo de Universo de juguete* -es el tipo de actividades que ocupan a los físicos cuando no pueden pensar en nada mejor que hacer- y, ¡no es una actividad tan fútil! Lo propio de un modelo de juguete es que no pretende ser un modelo real del Universo, puede reflejar algunas de sus características pero no pretende ser tomado en serio como un modelo del Universo real. Este modelo de juguete concreto no pretende, ciertamente, ser tomado en serio en este sentido. Se presenta solo para ilustrar un determinado aspecto.

En este modelo, existe un tiempo discreto que corre 0, 1, 2, 3, 4..., y el estado del Universo en cualquier instante viene dado por un *conjunto poliomino*. ¿Qué es un conjunto poliomino? Algunos ejemplos se ilustran en la figura 3.10. Un poliomino es una colección de cuadrados, todos ellos pegados a lo largo de aristas diversas para dar lugar a alguna forma plana. Estoy interesado en conjuntos de poliominos. Ahora, en este modelo de juguete, el estado del Universo en un instante viene dado por dos conjuntos finitos independientes de poliominos. En la figura 3.10 expongo una lista completa de todos los posibles conjuntos finitos de poliominos, enumerados S_0, S_1, S_2..., de alguna forma computacional. ¿Cuál es la evolución, o la dinámica, de este ridículo Universo? Empezamos en el instante cero, con los conjuntos de poliominos (S_0, S_0), y luego continuamos con otro par de conjuntos de poliominos de acuerdo con cierta regla precisa. Esta regla depende de si es posible o no utilizar un conjunto de poliominos dado para teselar el plano entero utilizando solamente los poliominos de dicho conjunto. La pregunta

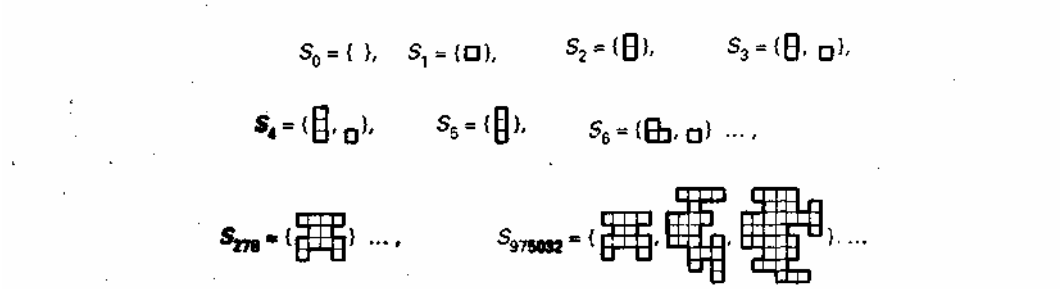


Figura 3.10. Un modelo de juguete no computable para el Universo. Los diferentes estados de este Universo de juguete determinista pero no computable están dados en términos de pares de conjuntos finitos de poliominos. Si el primer conjunto del par tesela el plano, la evolución temporal procede con un aumento del orden numérico del primer conjunto mientras que el segundo *marca el tiempo*. Cuando el primer conjunto no tesela el plano, los dos se permutan antes de que la evolución continúe. Sería algo así como: (S_0, S_0), (S_0, S_1), (S_1, S_1), (S_2, S_1), (S_3, S_1), (S_4, S_1)..., (S_{278}, S_{251}), (S_{251}, S_{278}), (S_{252}, S_{279})...

entonces es: ¿es posible cubrir el plano entero sin dejar huecos o sin que haya solapamientos utilizando solo los poliominos del conjunto dado? Supongamos ahora que el estado del Universo del modelo de juguete en cierto instante es el par de conjuntos de poliominos (S_q, S_r). La regla para la evolución de este modelo consiste en que, si es posible teselar el plano con los poliominos de S_q, entonces pasamos al siguiente S_{q+1}, lo que da el par (S_{q+1}, S_r) en el próximo instante de tiempo. Si no es posible hacerlo, entonces debemos además permutar el par para obtener (S_r, S_{q+1}). Es un Universo pequeño muy simple y estúpido -¿qué pasa con él?-. Lo que sucede es que, aunque su evolución es enteramente determinista -les he proporcionado una regla determinista absoluta y muy clara acerca de cómo va a evolucionar el Universo -es *no-computable*-. Se sigue de un teorema de Robert Berger que no hay ninguna operación de ordenador que pueda simular la evolución de este Universo porque no hay procedimiento decisorio computacional para decidir cuándo un conjunto de poliominos teselará el plano.

Esto ilustra el hecho de que computabilidad y determinismo son conceptos diferentes. Algunos ejemplos de teselados por poliominos se muestran en la figura 3.11. En los ejemplos (a) y (b), estas formas pueden teselar un plano completo, tal como se ilustra. En el ejemplo (c), las formas de la izquierda o de la derecha por sí solas no pueden teselar un plano —en ambos casos dejan huecos-. Pero, tomadas juntas, pueden teselar el plano entero, como se ilustra en (c). El ejemplo (d) teselará también el plano -solo puede teselar el plano en la forma mostrada y esto ilustra lo complejos que pueden llegar a ser estos teselados.

No obstante, las cosas pueden empeorar. Permítanme mostrarles el ejemplo de la figura 3.12; de hecho, el teorema de Robert Berger depende de la existencia de conjuntos de teselas como este. Las tres teselas mostradas en la parte superior de la figura cubrirán el plano entero, pero no hay forma de hacer esto

de tal modo que la figura se repita. Es siempre diferente a medida que sigamos avanzando y no es fácil ver lo que podemos hacer realmente con ello. Pero, en cualquier caso, puede hacerse y la existencia de teselados como estos entran en el argumento de Robert Berger a partir del cual se sigue que no hay programa de ordenador que pueda simular este universo de juguete.

¿Qué pasa con el Universo real? En el capítulo 2 he argumentado que en nuestra física está ausente algo fundamental. ¿Existe alguna razón derivada de la propia física para pensar que pudiera haber algo no-computable en dicha física ausente? Creo que hay algunas razones para creer esto: que la verdadera teoría de la gravitación cuántica podría ser no-computable. La idea no es en absoluto descabellada. Señalaré que la no-computabilidad es una característica de dos aproximaciones independientes a la gravedad cuántica. Lo que es característico de estas aproximaciones concretas es que involucran la superposición cuántica de espacio-tiempos tetradimensionales. Muchas otras aproximaciones involucran solo superposiciones de espacios tridimensionales.

La primera es el esquema de Geroch-Hartle para la gravedad cuántica, que resulta tener un elemento no-computable, ya que invoca un resultado, debido a Markov, que afirma que las cuatro variedades topológicas no son computacionalmente clasificables. No entraré en esta cuestión técnica, pero ello demuestra que esta característica de la no-

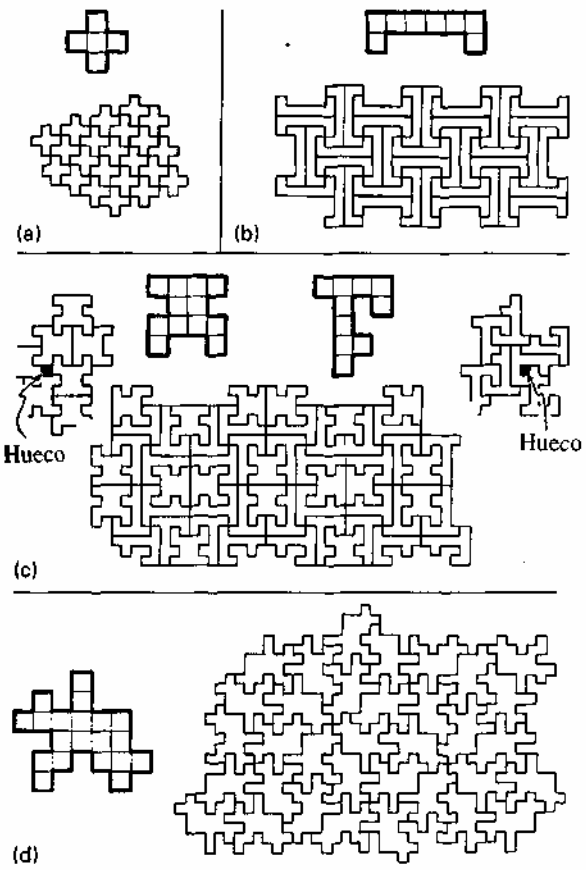


Figura 3.11. Diversos conjuntos de poliominos que teselarán el plano euclidiano infinito (si se permiten las teselas reflejadas). Sin embargo, ninguno de los poliominos del conjunto (c), por sí solo, teselará el plano.

computabilidad ha surgido ya de forma natural en los intentos de combinar la teoría general de la relatividad y la mecánica cuántica.

El segundo lugar donde ha surgido la no-computabilidad en una aproximación a la gravedad cuántica es en la obra de David Deutsch. Apareció en un borrador previo que distribuyó, pero luego, para mi pesar, cuando el artículo apareció en prensa, ¡el argumento no se encontraba por ninguna parte! Le

pregunté acerca de ello y me aseguró que lo suprimió, no porque estuviese equivocado, sino porque no era relevante para el resto del artículo. Su punto de vista es que, en estas divertidas superposiciones de espacio-

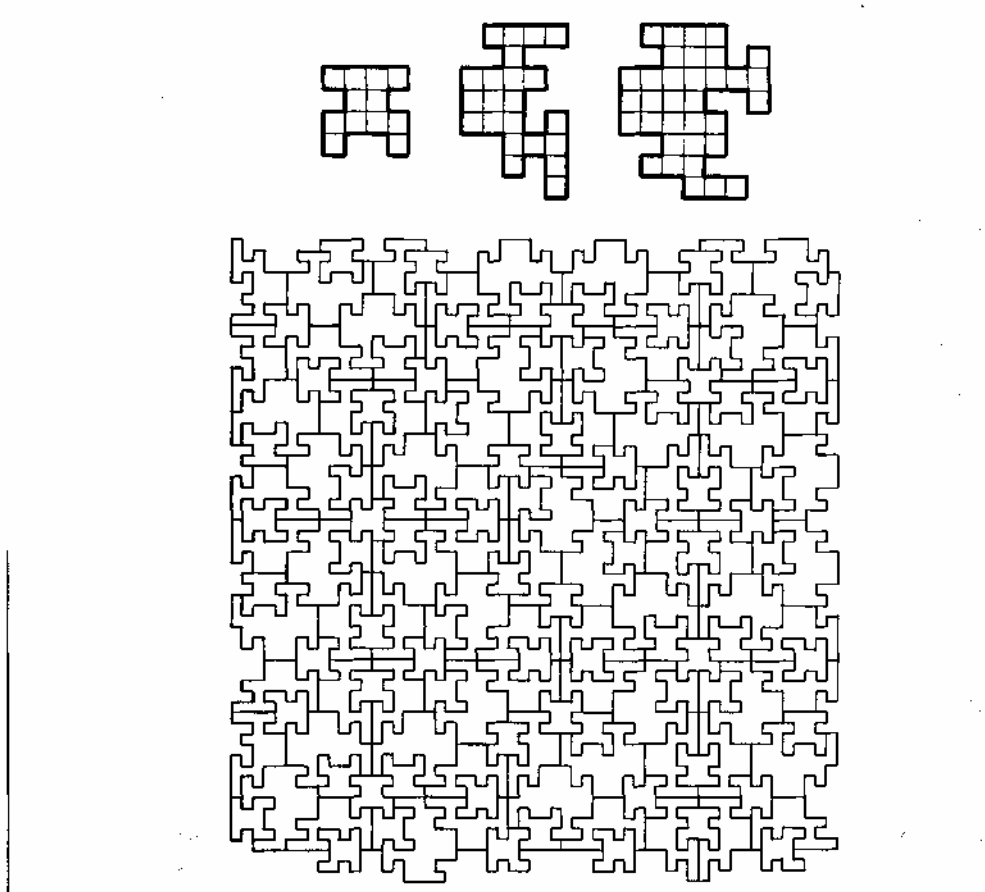


Figura 3.12. Este conjunto de tres polinomios teselará el plano solo de forma no periódica.

tiempos, uno tiene que considerar al menos la posibilidad de que algunos de estos Universos potenciales pudieran tener líneas cerradas de tipo tiempo (figura 3.13). En ellos, la causalidad se ha vuelto loca, futuro y pasado se mezclan, y las influencias causales se cierran en círculos. Ahora bien, aunque estas líneas cerradas solo tienen que jugar un papel como hechos contrafácticos, como en el problema de comprobación de bombas del capítulo 2, siguen ejerciendo influencia en lo que sucede realmente. Yo no diría que este es un argumento claro, pero es, al menos, un indicio de que fácilmente podría haber algo de naturaleza no-computacional en la teoría correcta, si es que la encontramos alguna vez.

Quiero plantear otra cuestión. Resalté que el determinismo y la computabilidad son conceptos diferentes. Ello tiene que ver ligeramente con la cuestión del *libre albedrío.*

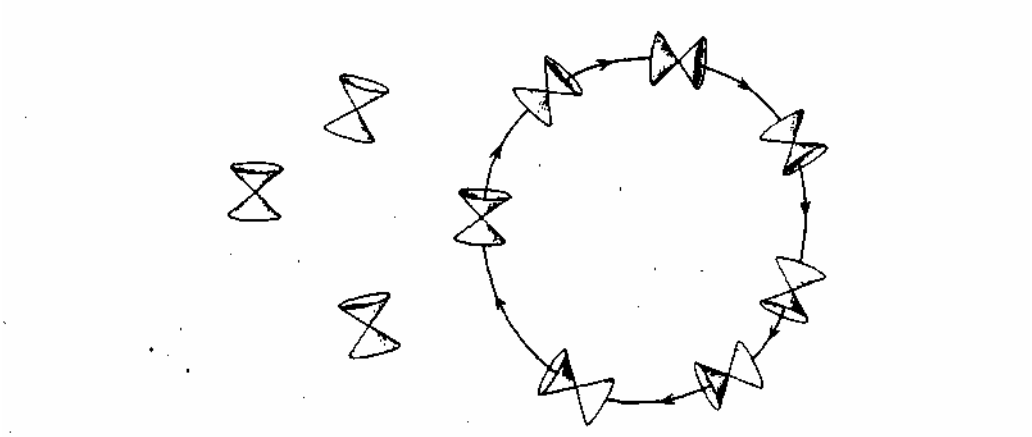


Figura 3.13. Con una inclinación suficientemente grande de los conos de luz, en el espacio-tiempo pueden darse líneas cerradas de tipo tiempo.

En discusiones filosóficas, el libre albedrío ha sido siempre considerado en términos de determinismo. En otras palabras: «¿está nuestro futuro determinado por nuestro pasado?», y otras cuestiones de esta naturaleza. Me parece que podrían plantearse otras muchas preguntas; por ejemplo: «¿está el futuro determinado computablemente por el pasado?» -esta es una cuestión diferente-.

Estas consideraciones plantean todo tipo de cuestiones diversas. Yo solo las plantearé —ciertamente no trataré de responderlas-. Siempre hay grandes discusiones acerca de hasta qué punto nuestras acciones están determinadas por *nuestra herencia y nuestro ambiente.* Y es extraño que no se mencione demasiado a menudo el papel de los *elementos aleatorios.* En cierto sentido, todas estas cosas están más allá de nuestro control. Ustedes podrían plantear la siguiente pregunta: «¿hay algo más, quizá una cosa denominada el yo, que es diferente de todas estas y que reside más allá de tales influencias?». Incluso asuntos legales tienen relevancia en el marco de una idea semejante. Por ejemplo, las cuestiones de derechos o responsabilidades parecen depender de las acciones de un *yo* independiente. Pero esto puede ser algo muy sutil. En primer lugar, existe la cuestión relativamente directa del *determinismo y no-determinismo.* El tipo normal de no-determinisrno . implica precisamente elementos aleatorios, pero eso no nos ayuda mucho. Estos elementos aleatorios siguen estando más allá de nuestro control. Podríamos tener en su lugar *no-computabilidad,* podríamos tener *tipos de no-computabilidad de orden superior.* En realidad es curioso que los argumentos del tipo de los de Gödel que he presentado pueden aplicarse de hecho en diferentes niveles. Pueden hacerlo en el nivel de lo que Turing llama *máquinas oráculo:* el argumento es realmente mucho más general que el que presenté más arriba. Por eso, uno tiene que considerar la cuestión de si podría o no existir algún tipo no-computabilidad de orden superior involucrado en la forma en que evoluciona Universo real. Quizá nuestras sensaciones de libre albedrío tengan algo que ver con esto.

He hablado del contacto con algún tipo de mundo platónico: ¿cuál es la naturaleza de este *contacto platónico*? Existen algunos tipos de palabras que parecerían involucrar elementos no-computables: juicio, sentido común, intuición, sensibilidad estética, compasión, moralidad... Creo que estas cualidades no son simples características de la Computación. Hasta ahora, he hablado del mundo platónico principalmente en términos de matemáticas, pero hay otros elementos que uno también podría incluir. Platón argumentaría ciertamente que no solo lo verdadero, sino también lo bueno y lo bello, son conceptos (platónicos) absolutos. Si existe algún tipo de contacto con los absolutos platónicos que nuestro conocimiento nos permite lograr, y que no puede explicarse en términos de comportamiento computacional, entonces creo que ésa es una cuestión importante.

¿Qué sucede con nuestros cerebros? La figura 3.14 muestra algo de un cerebro. Un constituyente principal del cerebro es su sistema de *neuronas.* Una parte importante de cada neurona es una fibra muy larga conocida como su *axón.* El axón se bifurca en ramas separadas en diversos lugares y cada una de estas ramas termina finalmente en una *sinapsis.* Estas sinapsis son las uniones donde las señales se transfieren desde cada neurona a (principalmente) otras neuronas por medio de sustancias químicas llamadas neurotransmisores. Algunas sinapsis son de naturaleza excitadora, con neurotransmisores que tienden a reforzar la activación de la neurona siguiente, y otras son inhibidoras, con tendencia a suprimir la activación de la siguiente neurona. Podemos referirnos a la fiabilidad de una sinapsis al pasar mensajes de una neurona a la siguiente como la *intensidad* de la sinapsis. Si todas las sinapsis tuvieran intensidades fijas, el cerebro sería muy parecido a un ordenador, pero se da el caso de que estas intensidades sinápticas

pueden cambiar y existen diversas teorías acerca de cómo cambian. Por ejemplo, el mecanismo de Hebb fue una de las primeras sugerencias para este proceso. La cuestión es, sin embargo, que todos los mecanismos que se han sugerido para inducir cambios son de naturaleza computacional, aunque con elementos probabilistas adicionales. Por eso, si tenemos algún tipo de regla probabilista-computacional que nos diga cómo cambian estas intensidades, entonces aún podremos simular la acción del sistema de neuronas y sinapsis mediante un ordenador (puesto que los elementos probabilistas también pueden ser fácilmente simulados computacionalmente) y obtener el tipo de sistema ilustrado en la figura 3.15.

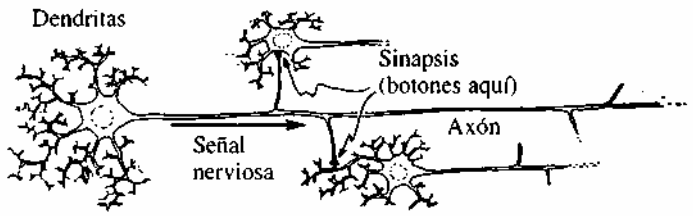


Figura 3.14. Un esbozo de una neurona, conectada a otras por medio de sinapsis.

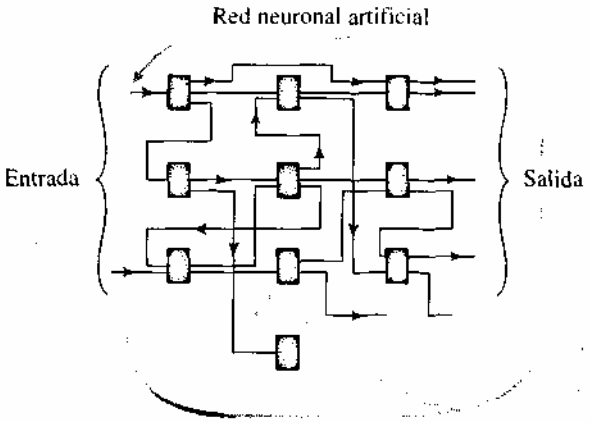


Figura 3.15.

Las unidades ilustradas en la figura 3.15, que podemos imaginar que son transistores, podrían desempeñar el papel de las neuronas en el cerebro. Por ejemplo, podemos considerar dispositivos electrónicos específicos conocidos como *redes neuronales artificiales.* En estas redes están incorporadas varias reglas concernientes a cómo cambian las intensidades de las sinapsis, normalmente para mejorar la calidad de alguna señal de salida. Pero las reglas son siempre de naturaleza computacional. Es fácil ver que esto debe ser así, por la muy buena razón de que estas cosas se simulan en ordenadores. Esa es la prueba: si uno es capaz de introducir el modelo en un ordenador, entonces es computable. Por ejemplo, Gerald Edelman tiene algunas sugerencias sobre cómo podría trabajar el cerebro que, según él, no son computacionales. ¿Cómo lo consigue? Edelman tiene un ordenador que simula todas estas sugerencias. Por eso, si hay un ordenador que se supone que lo simula, entonces es computacional.

Quiero abordar la cuestión siguiente: «¿qué están haciendo las neuronas individuales? ¿Están actuando simplemente como unidades computacionales?». Las neuronas son células y las células son estructuras muy elaboradas. De hecho, son tan elaboradas que, incluso si tuviéramos solo una de ellas, todavía podríamos seguir haciendo cosas muy complicadas. Por ejemplo, un paramecio, un ser unicelular, puede nadar hacia el alimento, alejarse del peligro, sortear obstáculos y, aparentemente, aprender por experiencia (figura 3.16). Todas estas son cualidades que uno pensaría que requieren un sistema nervioso pero, ciertamente, el paramecio no tiene sistema nervioso. Lo mejor que podría pasar es que, ¡el paramecio fuera una neurona en sí mismo! Evidentemente no hay neuronas en un paramecio: hay solo una única célula. Una afirmación semejante podría hacerse para una ameba. La pregunta es, ¿cómo lo hace?

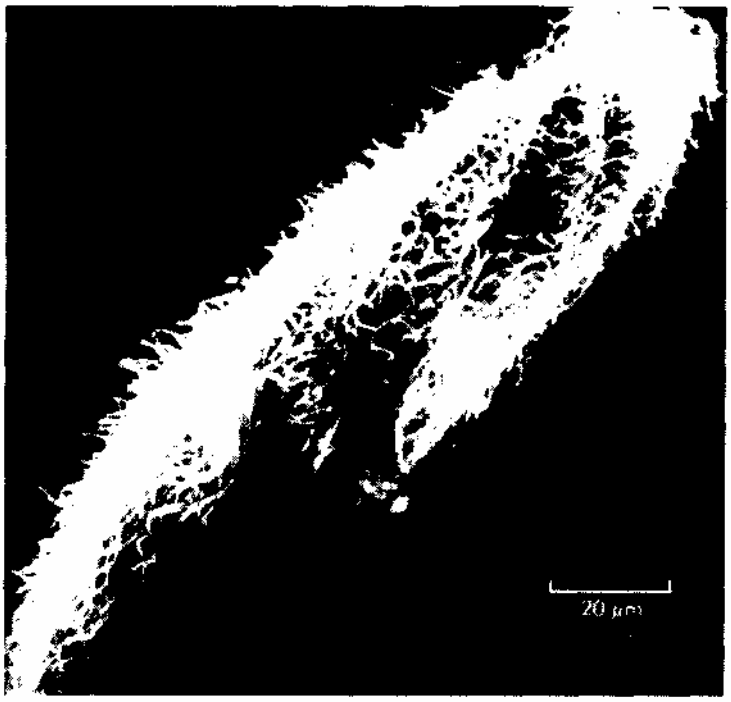


Figura 3.16. Un paramecio. Nótense los cilios similares a pelos que son utilizados para **nadar. Ellos** forman las extremidades externas del citoesqueleto del paramecio.

Una sugerencia es que el *citoesqueleto* -la estructura que, entre otras cosas, confiere su forma al paramecio- es lo que está controlando las acciones complejas de estos animales unicelulares. En un paramecio, los pequeños pelos, o cilios, que utiliza para nadar son los extremos del citoesqueleto, y están constituidos básicamente por pequeñas estructuras tubulares llamadas *microtúbulos.* El citoesqueleto está hecho de estos microtúbulos, además de actina y filamentos intermedios. Las amebas también se mueven, utilizando microtúbulos para impulsar sus pseudópodos.

Los microtúbulos son cosas extraordinarias. Los cilios, que los paramecios utilizan Para nadar, son básicamente haces de microtúbulos. Además, los microtúbulos están fuertemente involucrados en la mitosis, es decir, en la división celular. Esto es cierto Para los microtúbulos de las células ordinarias pero no, aparentemente, para las neuronas: las neuronas no se dividen y esto puede ser una diferencia importante. El centro de control del citoesqueleto es una estructura conocida como el *centrosoma,* cuya parte más prominente, el *centriolo,* consiste en dos haces de microtúbulos que forman una "T" separada. En una fase crítica, cuando el centrosoma se divide, cada uno de los dos cilindros del centriolo genera otro, de modo que se forman dos centriolos que luego se separan, y cada uno de ellos parece arrastrar un haz de microtúbulos con él. Estas fibras de microtúbulos conectan de algún modo las dos partes del centrosoma dividido con las hebras de ADN separadas en el núcleo de la célula y, entonces, dichas hebras de ADN se separan. Este proceso inicia la división celular.

Esto no es lo que sucede en las neuronas, porque las neuronas no se dividen, de modo que los microtúbulos deben de estar haciendo alguna otra cosa. ¿Qué están haciendo en las neuronas? Bien, probablemente montones de cosas, incluyendo el transporte de moléculas neurotransmisoras dentro de la célula, pero una actividad en la que sí parecen estar involucrados es en la determinación de las intensidades de las sinapsis. En la figura 3.17 se muestra una ampliación de una neurona y una sinapsis, donde también se indican las posiciones aproximadas de los microtúbulos, así como de las fibras de actina. Una forma posible de que los microtúbulos afectaran a la intensidad de la sinapsis es que tuvieran influencia en la naturaleza de una espina dendrítica (figura 3.17). Tales espinas aparecen en muchas sinapsis y, aparentemente, pueden crecer o contraerse o sufrir algún otro cambio en su naturaleza. Tales cambios pueden ser inducidos por alteraciones en la actina que hay en su interior -la actina es un constituyente esencial del mecanismo de la contracción muscular-. Los microtúbulos vecinos podrían influir fuertemente en esta actina que, a su vez, podría influir en la forma o las propiedades dieléctricas de la conexión sináptica. Existen al menos otras dos formas diferentes en las que los microtúbulos podrían estar implicados e influir en las intensidades de las sinapsis. Ciertamente están implicados en el transporte de

sustancias químicas neurotransmisoras que transportan la señal de una neurona a la siguiente. Son los microtúbulos los que las llevan a lo largo de los axones y dendritas y, de este modo, su actividad influiría en la concentración de dichas sustancias químicas en el extremo del axón y las dendritas. Esto, a su vez, podría influir en la intensidad de la sinapsis. Otra influencia del microtúbulo podría manifestarse en el crecimiento y la degeneración neuronal, al alterar la propia red de conexiones neuronales.

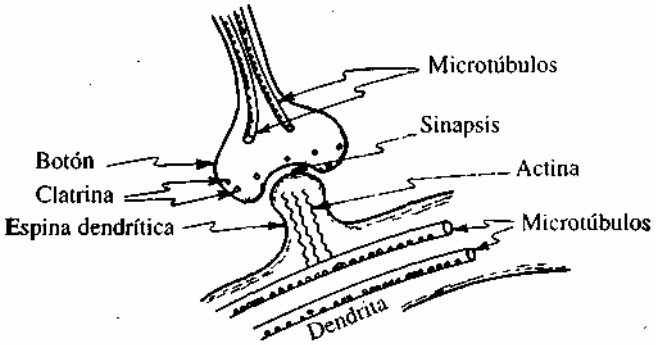


Figura 3.17. Las clatrinas (y las terminaciones de microtúbulos) se encuentran en los botones sinápticos del axón y parecen estar involucradas en la afectación de la intensidad de las sinapsis. Esto podría ocurrir por medio de los filamentos de actina en las espinas dendríticas.

¿Qué son los microtúbulos? En la figura 3.18 se muestra un esbozo de uno de ellos. Son pequeños tubos hechos de proteínas llamadas *tubulinas.* Son interesantes en varios aspectos. Las proteínas de tubulina parecen tener (al menos) dos estados, o conformaciones, diferentes, y pueden cambiar de una conformación a la otra. Aparentemente, pueden ser enviados mensajes a lo largo de los tubos. De hecho, Stuart Hameroff y sus colegas tienen ideas interesantes sobre el modo de enviar señales a lo largo de los tubos. Según Hameroff, los microtúbulos pueden comportarse como *autómatas celulares y* a lo largo de ellos pueden ser enviadas señales complicadas. Consideremos las dos conformaciones diferentes de cada tubulina como algo que representa los «0» y los «1» de un ordenador digital. Así, un único microtúbulo podría por sí solo comportarse como un ordenador, y tenemos que tener esto en cuenta si estamos considerando lo que hacen las neuronas. Cada neurona no solo se comporta como un interruptor sino que, más bien, implica muchos, muchos microtúbulos, y cada microtúbulo podría estar haciendo cosas muy complicadas.

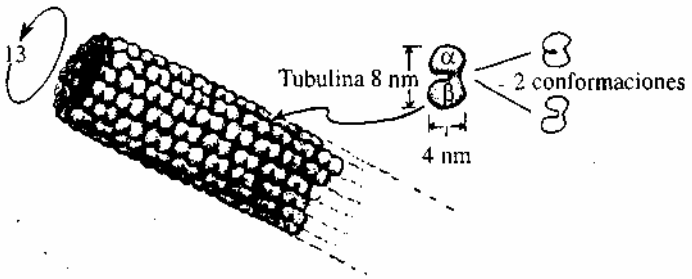


Figura 3.18. Un microtúbulo es un tubo hueco, que consiste normalmente en trece columnas de dímeros de tubulina. Cada molécula de tubulina parece ser susceptible de (al menos) dos conformaciones.

Aquí es donde entran mis propias ideas. Pudiera ser que la mecánica cuántica sea importante para comprender estos procesos. Una de las cosas que más me interesa de los microtúbulos es que son *tubos.* Al ser tubos, existe una posibilidad razonable de que sean capaces de aislar lo que está sucediendo en su interior frente a la actividad aleatoria en el entorno. En el capítulo 2 afirmé que necesitamos alguna forma nueva de física RO y, si esta nueva física va a ser relevante, debe haber movimientos de masas en superposiciones cuánticas que estén bien aislados del entorno. Muy bien pudiera ocurrir que, dentro de los tubos, se produzca algún tipo de actividad cuántica coherente a gran escala, algo parecido a un superconductor. Un movimiento de masa significativo estaría involucrado solo cuando su actividad empezara a acoplarse a las conformaciones de tubulina (tipo Hameroff) y entonces el comportamiento de *autómata celular* estaría en sí mismo sometido a superposición cuántica. El tipo de situaciones que podrían tener lugar se ilustra en la figura 3.19.

Como parte de esta imagen tendría que haber algún tipo de oscilación cuántica coherente que ocurriera dentro de los tubos y que necesitara extenderse sobre áreas muy

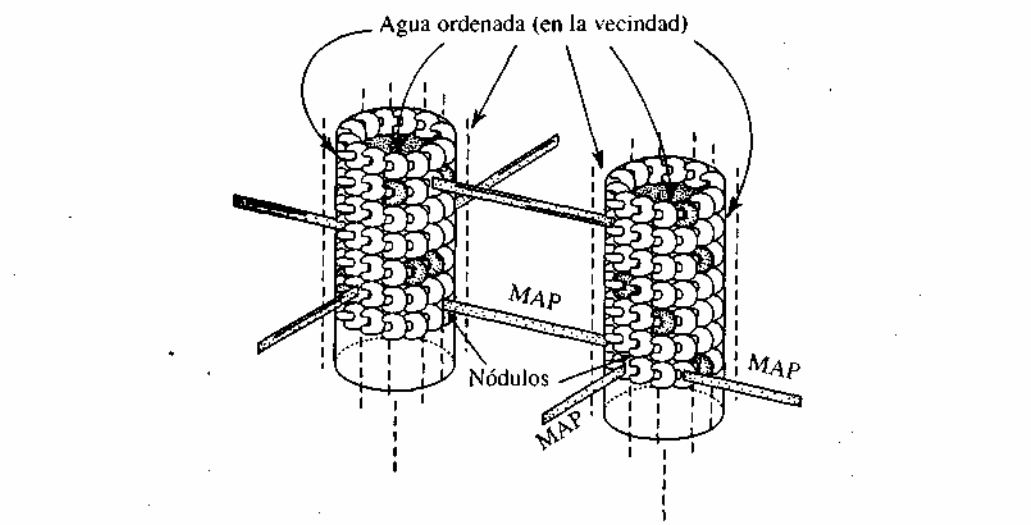


Figura 3.19. Sistemas de microtúbulos dentro de (colecciones de) neuronas podrían mantener actividad cuántica coherente a gran escala, y las ocurrencias RO individuales constituirían sucesos conscientes. Se necesita un aislamiento eficaz para esta actividad, posiblemente por agua ordenada que rodea los microtúbulos. Un sistema de interconexiones microtúbulos-proteínas asociadas (MAP) podría *sintonizar* esta actividad, fijando a los microtúbulos en *nódulos*.

amplias del cerebro. Hace bastantes años, Herbert Frolich presentó algunas sugerencias generales de este tipo, dando alguna plausibilidad al hecho de que elementos de esta naturaleza pudieran formar parte de los sistemas biológicos. Los microtúbulos parecen ser un buen candidato para las estructuras dentro de las cuales podría tener lugar esta actividad coherente cuántica a *gran escala*. Cuando utilizo el término gran escala, deben recordar ustedes que en el capítulo 2 describí el enigma EPR y los efectos de la no-localidad cuántica que muestran que efectos que están ampliamente separados no pueden ser considerados independientes entre sí. Efectos no-locales como este ocurren en mecánica cuántica y no pueden entenderse en términos de hechos separados unos de otros: está teniendo lugar algún tipo de actividad global.

Creo que la conciencia es algo global. Por consiguiente, cualquier proceso físico responsable de la conciencia tendría que presentar un carácter esencialmente global. La coherencia cuántica es idónea a este respecto. Para que sea posible semejante coherencia cuántica a gran escala, necesitamos un alto grado de aislamiento, tal como el que podrían suministrar las paredes de los microtúbulos. Sin embargo, también necesitamos algo más cuando empiezan a involucrarse las conformaciones de tubulina. Este necesario aislamiento adicional del entorno podría ser proporcionado por agua ordenada en el exterior inmediato de los microtúbulos. El agua ordenada (que se sabe que existe en células vivas) sería también probablemente un ingrediente importante de cualquier oscilación coherente cuántica que tuviera lugar dentro de los tubos. Aunque es bastante pedir, quizá no sea totalmente irrazonable que todo esto pudiera darse.

Las oscilaciones cuánticas dentro de los tubos tendrían que estar acopladas de alguna forma con la acción de los microtúbulos, a saber, la actividad de autómata celular de la que habla Hameroff; pero ahora su idea tiene que combinarse con la mecánica cuántica. Así pues, debemos tener, no solo actividad computacional en el sentido ordinario, sino también computación cuántica que implica superposiciones de diferentes acciones de este tipo. Si esa fuera toda la historia, aún seguiríamos en el nivel cuántico. En un cierto momento, el estado cuántico podría enmarañarse con el entorno. Entonces saltaríamos al nivel clásico de una forma aparentemente aleatoria, de acuerdo con el procedimiento R usual de la mecánica cuántica. Esto no es bueno si queremos que intervenga una no-computabilidad genuina. Para ello tienen que manifestarse los aspectos no-computables de RO, y eso requiere un aislamiento excelente. Por ello, afirmo que necesitamos algo en el cerebro que tenga un aislamiento suficiente para que la nueva física RO tenga una oportunidad de jugar un papel importante. Lo que necesitaríamos es que estas computaciones microtubulares superpuestas, una vez que se ponen en marcha, estuvieran suficientemente aisladas para que esta nueva física entrara realmente en juego.

Por eso, la imagen que yo tengo es que, durante algunos instantes, estas computaciones cuánticas continúan y permanecen aisladas del resto del material el tiempo suficiente -quizá algo del orden de casi un

segundo- para que criterios del tipo de los que estoy hablando dominen sobre los procedimientos cuánticos estándar, intervengan los ingredientes no computacionales y obtengamos algo esencialmente diferente de la teoría cuántica estándar.

Por supuesto, hay mucha especulación en muchas de estas ideas. Pese a todo, ofrecen una posibilidad genuina de una imagen mucho más específica y cuantitativa de la relación entre la conciencia y los procesos biológicos que la que han suministrado otros enfoques. Podemos al menos empezar a hacer un cálculo de cuántas neuronas tendrían que estar involucradas para que esta acción RO pudiera llegar a ser relevante. Lo que se necesita es alguna estimación de T, la escala de tiempo de la que hablé hacia el final del capítulo 2. En otras palabras, suponiendo que los sucesos conscientes están relacionados con tales ocurrencias RO, ¿cuál estimamos que es T? ¿Cuánto tiempo requiere la conciencia? Hay dos tipos de experimentos, ambos asociados con Libet y sus asociados, que resultan relevantes para estas ideas. Uno de ellos trata con el libre albedrío, o conciencia activa; el otro trata con la sensación, o la conciencia pasiva.

Primero, consideremos el libre albedrío. En los experimentos de Libet y Kornhuber se le pide a un sujeto que presione un botón en un instante enteramente dejado a su volición. Se colocan electrodos en la cabeza del sujeto para detectar la actividad eléctrica del cerebro. Se hacen muchas pruebas repetidas y los resultados se promedian (figura 3.20.a). El resultado global es que existe algún indicio claro de tal actividad eléctrica aproximadamente un segundo antes del instante en que el sujeto cree que toma la decisión real. Por ello, la libre voluntad parece implicar algún tipo de retardo temporal, del orden de un segundo.

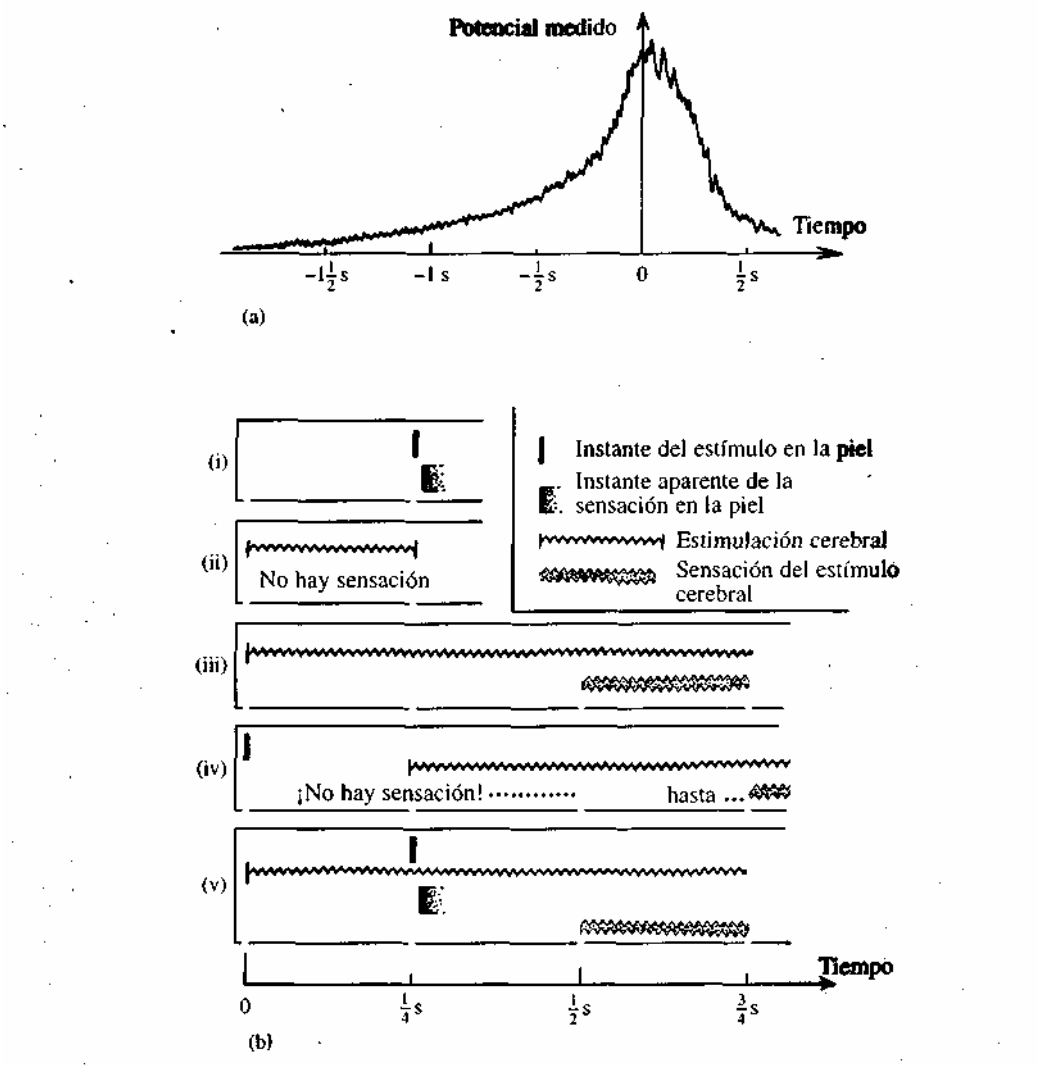


Figura 3.20. (a) Experimento de Kornhuber, posteriormente repetido y refinado por Liber y sus colegas. La decisión de flexionar el dedo parece tomarse en el instante 0, pero la señal precursora (promediada sobre muchos ensayos) sugiere un *preconocimiento* de la intención de flexionar. (b) Experimento de Libet. (i) El estímulo de la piel parece ser percibido aproximadamente en el momento real del estímulo. (ii) Un estímulo cortical de menos de medio segundo no es percibido. (iii) Un estímulo cortical de más de un segundo es percibido a partir de medio segundo en adelante. (iv) Tal estímulo subcortical puede *enmascarar retroactivamente* un estímulo previo en la piel, lo que indica que la consciencia del estímulo *aún no había tenido lugar* realmente en el momento del estímulo subcortical. (v) Si un estímulo en la piel se aplica inmediatamente *después* de semejante estímulo cortical, entonces la consciencia dérmica es *remitida hacia atrás* pero la consciencia cortical no lo es.

Más notables son los experimentos pasivos, que son más difíciles de realizar. Parecen sugerir que se necesita aproximadamente medio segundo de actividad en el cerebro antes de que una persona llegue a ser pasivamente conciente de algo (figura 3.20.b). En estos experimentos existen formas de bloquear la experiencia consciente de un estímulo en la piel, ¡hasta, aproximadamente, medio segundo *después* de que este estímulo haya ocurrido realmente! En aquellos casos en los que el procedimiento de bloqueo no se efectúa, el sujeto cree que la experiencia del estímulo en la piel ha ocurrido en el instante real de dicho estímulo. Pese a todo, podría haber sido bloqueado incluso medio segundo después del instante real del estímulo. Estos son experimentos muy intrigantes, en especial cuando se consideran conjuntamente. Sugieren que la voluntad consciente parece necesitar aproximadamente un segundo, y que la sensación consciente necesita aproximadamente medio segundo. Si imaginamos que la conciencia es algo que hace algo, entonces se nos presenta casi una paradoja. Necesitamos medio segundo antes de hacernos conscientes de algún suceso. Entonces tratamos de invocar a nuestra conciencia para hacer algo con ella. Necesitamos otro segundo para que nuestra libre voluntad haga ese algo. En resumen, necesitamos

aproximadamente un total de un segundo y medio. Por eso, si algo requiere una respuesta conscientemente voluntaria, necesitaríamos aproximadamente un segundo y medio antes de que realmente pudiéramos hacer uso de ella. Personalmente encuentro eso bastante difícil de creer. Consideremos, por ejemplo, la conversación ordinaria. Creo que, aunque buena parte de la conversación podría ser automática e inconsciente, el hecho de que se necesite un segundo y medio para dar una respuesta *consciente* me parece muy extraño.

Yo considero esto como indicio de que muy bien podría haber algo en nuestra forma de interpretar tales experimentos que presuponga que la física que estamos utilizando es básicamente física clásica. Recordemos el problema de comprobación de bombas donde hablábamos de hechos contrafácticos y el hecho de que los sucesos contrafácticos pudieran tener influencia sobre cosas, incluso si no ocurrieron realmente. El tipo ordinario de lógica que utilizamos tiende a fallar si no somos cuidadosos. Tenemos que tener en cuenta cómo se comportan los sistemas cuánticos, y podría suceder que algo curioso esté ocurriendo en estas medidas de tiempo debido a la no-localidad cuántica y los hechos contrafácticos cuánticos. Es muy difícil entender la no-localidad cuántica dentro del marco de la teoría especial de la relatividad. Mi opinión personal es que, para comprender la no-localidad cuántica, necesitaremos una teoría radicalmente nueva. Esta nueva teoría no será simplemente una ligera modificación de la mecánica cuántica sino algo tan diferente de la mecánica cuántica estándar como la teoría general de la relatividad es diferente de la gravedad newtoniana. Tendría que ser algo con un marco conceptual completamente diferente. En esta imagen, la no-localidad cuántica estaría incorporada en la teoría.

En el capítulo 2 se mostró que la no-localidad es algo que, aunque muy enigmático, puede describirse matemáticamente. Permítanme mostrarles la imagen de un triángulo imposible en la figura 3.21. Ustedes podrían preguntarse: «¿Dónde está la imposibilidad?». ¿Pueden ustedes localizarla? Ustedes pueden tapar partes diversas de la imagen y, sea cual sea el fragmento del triángulo que tapen, la imagen se hace

repentinamente posible. Por eso ustedes no pueden decir que la imposibilidad está en algún lugar concreto en la imagen: la imposibilidad es una característica de la estructura completa. De todos modos, existen formas matemáticas precisas para hablar de tales cosas. Esto puede hacerse descomponiendo la imagen, volviéndola a encolar y extrayendo ciertas ideas matemáticas abstractas de la estructura global y detallada del encolado. La noción de *cohomología* es la noción apropiada en este caso. Esta noción nos proporciona un medio de calcular el grado de imposibilidad de esta figura. Este es el tipo de matemática no-local que muy bien podría estar implicada en nuestra nueva teoría.

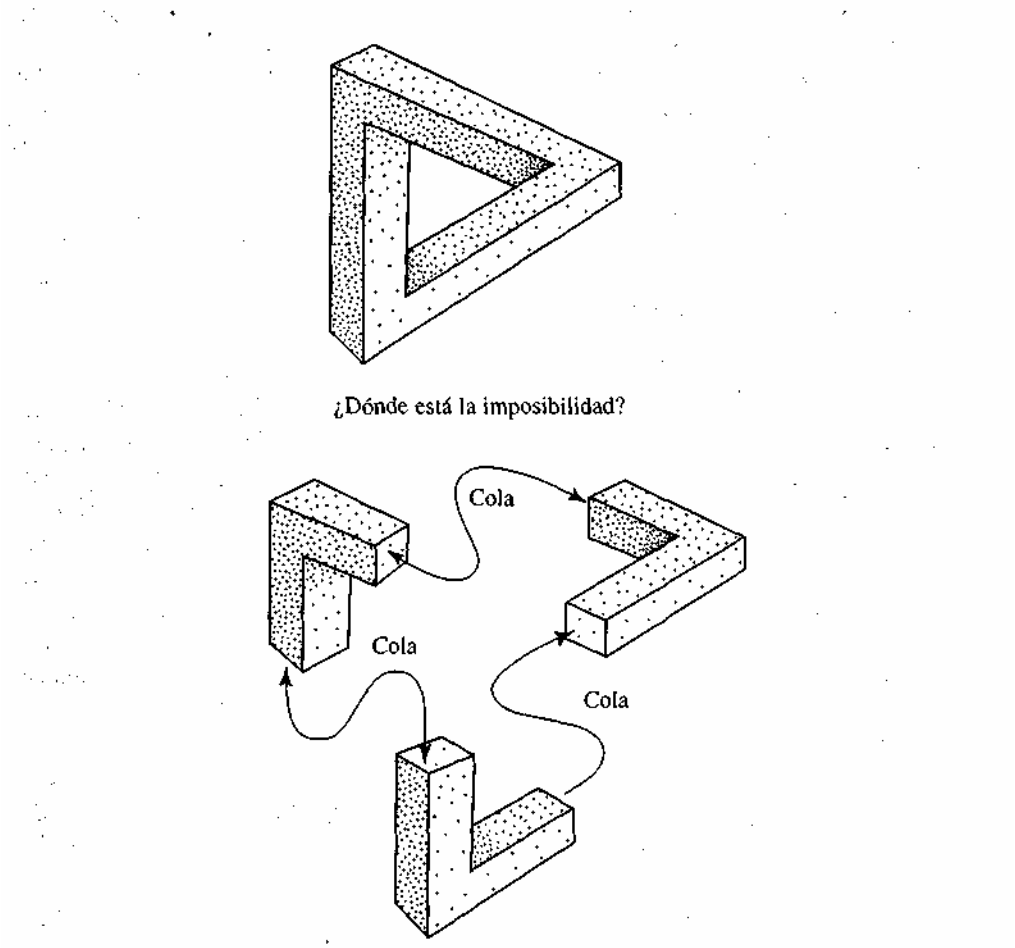


Figura 3.21. **Triángulo imposible.** La *imposibilidad* no puede localizarse; pese a todo, puede definirse en términos matemáticos precisos como una abstracción de las *reglas de encolado* subyacentes a su construcción.

¡No hay que suponer que sea un accidente el que la figura 3.21 se parezca a la figura 3.3! El dibujo de la figura 3.3 fue realizado deliberadamente para resaltar un elemento de paradoja. Hay algo característicamente misterioso en la forma en que estos tres mundos se interrelacionan -en donde cada uno de ellos parece casi *emerger* de una pequeña parte de su predecesor-. Pese a todo, como sucede con la figura 3.21, con comprensión adicional quizá seamos capaces de entender o incluso resolver parte de este misterio. Es importante reconocer los rompecabezas y los misterios cuando se presentan. Pero el solo hecho de que suceda algo muy enigmático no significa que nunca seremos capaces de comprenderlo.

Referencias

Albrecht-Buehler, G., (1981)
«Does the geometric design of centrioles imply their function?»,
Cell Motility, 1, págs. 237-245.

Albrecht-Buehler, G., (1991)
«Surface extensions of 3T3 cells towards distant infrared light sources»,
J. Cell Biol., 114, págs. 493-503.

Aspect, A., Grangier, P. y Roger, G. (1982),
«Experimental realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: a new violation of Bell's inequalities»,
Phys. Rev. Lett., 48, págs. 91-94.

Beckenstein, J. (1972),
«Black holes and the second law»,
Lett. Nuovo dm., 4, págs. 737-740.

Bell, J. S. (1991),
Lo decible y lo indecible en mecánica cuántica,
Alianza Editorial, Madrid .

Bell, J. S. (1990)
«Against measurement»,
Physics World, 3, págs. 33-40.

Berger, R. (1966)
«The undecidability of the domino problem»,
Memoirs Amer. Math. Soc., 66, pág. 72.

Bohm, D. y Hiley, B. (1994),
The Undivided Universe, Routledge, London.

Davenport, H. (1968),
The Higher Arithmetic,
Hutchinsons's University Library, London, 3.ª ed.

Deeke, L., Grotzinger, B. y Kornhuber, H. H. (1976),
«Voluntary finger movements in man: cerebral potentials and theory»,
Biol. Cybernetics, 23, pág. 99.

Deustch, D. (1985),
«Quantum theory, the Church-Turing principle and the universal quantumcomputer»,
Proc. Roy. Soc. (Lond.), A400, págs. 97-117.

DeWitt, B. S. y Graham, R. D., eds. (1973),
The Many-Worlds Interpretation of Quantum Mechanics,
Princeton University Press, Princeton.

Diósi, L. (1989),
«Models for universal reduction of macroscopic quantum fluctuations»,
Phys. Rev., A40, págs. 1165-1174.

Frohlich, H. (1968),
«Long-range coherence and energy storage in biological systems»,
Int. J. of Quantum. Chem., vol. II, págs. 641-649.

Gell-Mann, M. y Hartle, J. B. (1993),
«Classical equations for quantum systems»,
Phy. Rev., D 47, págs. 3345-3382.

Geroch, R. y Hartle, J. (1986),
«Computability and physical theories»,
Found. Phys., 16, pág. 533.

Godel, K. (1931),
«Über formal unentscheidbare Satze der Principia Mathematica und verwandter System 1»,
Monatshefte fur Mathematik und Physik, 38, págs. 173-198. [Hay traducción castellana en: Godel, K. (1987), *Obras complejas,* Alianza Editorial, Madrid.]

Golomb, S. W. (1966),
Polyominoes,
Scribner and Sons, Londres.

Haag, R. (1992),
Local Quantum Physics: Fields, Particles, Algebras,
Springer-Verlag, Berlín.

Hameroff, S. R. y Penrose, R. (1996),
«Orchestrated reduction of quantum coherence in brain microtubules, a model for conciousness», *Toward a Science of Consciousness: Contributions from the 1994 Tucson Conference,* eds.

Hameroff, S., Kaszniak, A. y Scott, A.,
M.T. Press, Cambridge, Mass.

Hameroff, S. R. y Penrose R. (1996),
«Conscious events as orchestrated space-time selections»,
J. Consciousness Studies, 3, págs. 36-53.

Hameroff, S. R. y Watt, R. C. (1982),
«Information processing in microtubules»,

J. Theor. Biol, 98, págs. 549-561.

Hawking, S. W. (1975),
«Particle creation by black holes»,
Comm. Math. Phys., 43, págs. 199-220.

Hughston, L. P., Jozsa, R. y Wooters, W. K. (1993),
«A complete classification of quantum ensembles having a given density matrix»,
Phys. Letters, A183, págs. 14-18.

Károlyházy, F. (1966),
«Gravitation and quantum mechanics of macroscopic bodies»,
Nuovo Cim., A42, pág. 390.

Károlyházy, F. (1974),
«Gravitation and quantum mechanics of macroscopic bodies»,
Magyar Fizikai PolyoirMat, 12, pág. 24.

Károlyházy, F., Frenkel, A. y Lukács, B. (1986),
«On the possible role of gravity on the reduction of the wave function»,
Quantum Concepts in Space and Time,
eds. R. Penrose y C. J. Isham, Oxford University Press, Oxford, págs. 109-128.

Kibble, T. W. B. (1981),
«Is a semi-classical theory of gravity viable?»,
Quantum Gravity 2: Second Oxford Symposium,
eds. C. J. Isham, R. Penrose y D.W. Sciama, Oxford University Press, Oxford, págs. 63-80.

Libet, B. (1992),
«The neural time-factor in perception, volition and free will»,
Revue de Métaphysique et de Morale, 2, págs. 255-272.

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, págs. 193-224.

Lockwood, M. (1989),
Mind, Brain and the Quantum,
Basil Blackwell, Oxford.

Lucas, J. R. (1961),
«Minds, Machines and Godel»,
Philosophy, 36, págs. 120-124; reimpresión en Alan Ross Anderson (1964), *Minds and Machines*, Prentice-Hall, New Jersey.

Majorana, E. (1932),

«Atomi orientati in campo magnético variabile»,
Nuovo Cimento, 9, págs. 43-50.

Moravec, H. (1988)
Mind Children: The Future of Robot and Human Intelligence,
Harvard University Press, Cambridge, Mass.

Omnés, R. (1992),
«Consistent interpretations of quantum mechanics»,
Rev. Mod. Phys., 64, págs. 339-382.

Pearle, P. (1989),
«Combining stochastic dynamical state-vector reduction with spontaneous localisation»,
Phys. Rev., A39, págs. 2277-2289.

Penrose, R. (1989)
The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics,
Oxford University Press, Oxford.
[Hay traducción castellana: *La nueva mente del emperador,* Mondadori, Madrid, 1991; reimpreso en Crítica, Barcelona, 1993.]

Penrose, R. (1989),
«Difficulties with inflationary cosmology»,
Proceedings of the 14th Texas Symposium on Relativistic Astrophysics, ed. E. Fenves,
Annals of N. Y. Acad. ScL, 571, pág. 249, N. Y. Acad. Science, New York.

Penrose, R. (1991),
«On the cohomology of impossible figures [La cohomologie des figures impossibles]»,
Structural Topology [Topologie structurale], 17, págs. 11-16.

Penrose, R. (1994),
Shadows of the Mind: An Approach to the Missing Science of Consciousness,
Oxford University Press, Oxford.
[Hay traducción castellana: *Las sombras de la mente,* Crítica, Barcelona, 1996.]

Penrose, R. (1996),
«On gravity's role in quantum state reduction»,
Gen. Reí. Grav., 28, pág. 581.

Percival, I. C. (1995),
«Quantum spacetime fluctuations and primary state diffusion»,
Proc. R. Soc. Land., A451, págs. 503-513.

Schrodinger, E. (1935),
«Die gegenwartige Situation in der Quantenmechanik»,
Naturwissenschaften, 23, págs. 807-812, 823-828, 844-849.

[Traducido por J. T. Trimmer (1980) en *Proc. Amer. Phil. Soc.*, 124, págs. 323-338].

Schrodinger, E. (1935),
«Probability relations between separated systems»,
Proc. Camb. Phil. Soc., 31, págs. 555-563.

Searle, J. R. (1980),
«Minds, Brains and Programs»,
The Behavioral and Brain Sciences, vol. III,
Cambridge University Press, Cambridge.

Seymore, J. y Norwood, D. (1993),
«A game for life»,
New Scientist, 39, págs. 23-26.

Squires, E. (1990),
«On an alleged proof of the quantum probability law»,
Phys. Lett., A145, págs. 67-68.

Turing, A. M. (1937),
«On computable numbers with an application to the Entscheidungsproblem»,
Proc. Land. Math. Soc., ser. 2, 42, págs. 230-265; una corrección (43), págs. 544-546.

Turing, A. M. (1939),
«Systems of logic based on ordinals»,
P. Lond. Math. Soc., 45, págs. 161-228.

Von Neumann, J. (1955),
Mathematical Foundations of Quantum Mechanics,
Princeton University Press, Princeton.
[Hay traducción castellana: *Fundamentos matemáticos de la mecánica cuántica*, Instituto Jorge Juan, csic, Madrid, 1955; reeditada en 1991.]

Wigner, E. P. (1960),
«The unreasonable effectiveness of mathematics in the physical sciences»,
Cornmun. Pure Appl. Math., 13, pags. 1-14.

Zurek, W. H. (1991),
«Decoherence and the transition from quantum to classical»,
Physics Today, 44, págs. 36-44.

4
Sobre mentalidad, mecánica cuántica y la actualización de potencialidades

ABNER SHIMONY

Introducción

Lo que más admiro en el trabajo de Roger Penrose es el espíritu de sus investigaciones: la combinación de dominio técnico, valentía y determinación para dirigirse al corazón de la cuestión. Penrose sigue la gran máxima de Hubert: «Mr *müssen wissen, wir werden wissen»*[3]. Con respecto al programa de su investigación, coincido con él en tres tesis básicas. Primera, la mentalidad puede tratarse científicamente. Segunda, las ideas de la mecánica cuántica son importantes para comprender el conjunto mente-cuerpo. Tercera, el problema mecanocuántico de la actualización de las potencialidades es un genuino problema físico que no puede resolverse sin modificar el formalismo cuántico. Soy escéptico, sin embargo, con respecto a muchos detalles de la elaboración que hace Roger de estas tres tesis y espero que mis críticas le estimulen para realizar mejoras.

4.1. La posición de la mentalidad en la Naturaleza

Aproximadamente una cuarta parte de los capítulos 1, 2 y 3, y casi la mitad de su libro *Las sombras de la mente* (en lo sucesivo lo abreviaré por *Sombras*) están dedicadas a establecer el carácter no-algorítmico de la capacidad matemática humana. La recensión que hizo Hilary Putnam de *Sombras* afirmaba que hay algunas lagunas en el argumento: Roger olvida la posibilidad de un programa para una máquina de Turing que simule capacidad matemática humana pero que no sea demostrablemente válido, y la posibilidad de que un programa semejante fuera tan complejo que, en la práctica, una mente humana no pudiera entenderlo. A mí no me convence la respuesta de Roger a Putnam[4], pero, por otra parte, no tengo un conocimiento suficiente sobre la teoría de la demostración para pronunciarme con fiabilidad. Creo, no obstante, que este punto es tangencial al interés central de Roger, y que es un alpinista que ha tratado de escalar la montaña equivocada. Su tesis central, a saber, que hay algo en los actos mentales que no puede ser alcanzado por ningún computador artificial, no depende de que sea establecido el carácter no algorítmico de las operaciones matemáticas humanas. De hecho, como un apéndice a su extenso argumento gódeliano, Roger presenta *(Sombras,* págs. 56-57) la tesis de «la habitación china» de John Searle según la cual una computación correcta, realizada por un autómata, no constituye un acto de comprensión. El núcleo del argumento es que un sujeto humano podría ser entrenado para comportarse como un autómata siguiendo en su conducta instrucciones presentadas acústicamente en chino, incluso si el sujeto no entiende el chino y sabe que este es el caso. Un sujeto que lleve a cabo correctamente una computación siguiendo estas instrucciones puede comparar directamente la experiencia normal de computar por comprensión y la experiencia anormal de computar como un autómata. La verdad matemática establecida por la computación en cuestión quizá sea enteramente trivial y, en cualquier caso, la diferencia entre computar mecánicamente y comprender es intuitivamente evidente.

Lo que Searle, con el apoyo de Roger, ha defendido con respecto a la comprensión matemática se aplica también a otros aspectos de la experiencia consciente: a las cualidades sensoriales, a las sensaciones de dolor y de placer, a los sentimientos de volición, a la intencionalidad (que es la experiencia referida a objetos, conceptos o proposiciones), etc. Dentro de la filosofía general del *fisicalismo* hay varias estrategias para dar cuenta de estos fenómenos[5]. En las teorías de *dos-aspectos,* dichas experiencias se consideran

[3] «Debemos saber, de modo que sabremos.» Esta máxima está grabada en la lápida de la tumba de Hubert. Ver Reid C. (1970), *Hilbert,* pág. 220, New York, Springer-Verlag.

[4] Penrose R. (1994), Carta a *The New York Times Book Review,* die. 18, pág. 39.

[5] Block, N. (1980), *Readings in Philosophy of Psychology,* Vol. I, caps. 2 y 3, Harvard University Press, Cambridge, **MA.**

[5] Whitehead, A. N. (1933), *Adventures of Ideas,* MacMilIan, Londres. Whitehead, A. N. (1929), *Process of Reality,* MacMillan, Londres.

como aspectos de estados cerebrales concretos; otras teorías identifican una experiencia mental con una clase de estados cerebrales, siendo la clase tan sutil que no puede darse una caracterización física explícita de ella, lo que impide la *reducción* explícita de un concepto mental a conceptos físicos; las teorías funcionalistas identifican las experiencias mentales con programas formales que pueden ser ejecutados, en principio, por muchos sistemas físicos diferentes incluso si, como cuestión de hecho contingente, son ejecutados por una red de neuronas. Un argumento fisicalista recurrente, resaltado particularmente por las teorías de dos aspectos, pero utilizado también por otras variedades de fisicalismo, es que una entidad caracterizada por un conjunto de propiedades puede ser idéntica a una entidad caracterizada por otro conjunto de propiedades completamente diferente. Las caracterizaciones pueden involucrar diferentes modalidades sensoriales, o una puede ser sensorial y la otra microfísica. El argumento procede entonces sugiriendo que la identidad de un estado mental y un estado cerebral (o una clase de estados cerebrales o un programa) es un ejemplo de esta lógica de la identidad general. Creo que existe un profundo error en este razonamiento. Cuando un objeto caracterizado por una modalidad sensorial es identificado con un objeto caracterizado por otra modalidad, hay una referencia tácita a dos cadenas causales, de las cuales, ambas tienen un término común en un único objeto y otro término común en el teatro de la conciencia del perceptor, pero con diferentes lazos causales intermedios en el entorno, y en los aparatos sensorial y cognitivo del perceptor. Cuando un estado cerebral y un estado de conciencia son identificados, según la versión de dos aspectos del fisicalismo, no hay dificultad en reconocer un objeto común como término: es, de hecho, el estado cerebral, puesto que el fisicalismo está comprometido con la primacía ontológica de la descripción física. Pero el otro término, el teatro de la conciencia del perceptor, está ausente. O quizá deberíamos decir que hay una equivocación general en la teoría de los dos aspectos, puesto que supone tácitamente un teatro común como lugar de combinación y comparación del aspecto físico y el mental; pero, por otra parte, si el fisicalismo es correcto, no existe un estado independiente para este teatro.

Un argumento afín en contra del fisicalismo descansa en un principio filosófico que yo llamo *el principio fenomenológico* (pero sería bienvenido un nombre mejor, si existe o pudiera ser sugerido en la literatura): esto es, cualquiera que sea la ontología que reconoce una filosofía coherente, dicha ontología debe bastar para explicar las apariencias. Una consecuencia de este principio es que el fisicalismo es incoherente. Una ontología fisicalista puede postular, y normalmente lo hace, una jerarquía ontológica, cuyo nivel fundamental consiste típicamente en partículas elementales o campos, y cuyos niveles superiores consisten en compuestos formados a partir de las entidades elementales. Estos compuestos pueden ser caracterizados de modos diferentes: las caracterizaciones de grano fino describen el microestado en detalle; las caracterizaciones de grano grueso suman, promedian o integran las descripciones de grano fino; las caracterizaciones relacionales dependen de conexiones causales entre los sistemas compuestos de interés y los instrumentos o los perceptores. ¿Dónde encajan las apariencias sensoriales en esta concepción de la naturaleza? No encajan en las caracterizaciones de grano fino, a menos que se introduzcan subrepticiamente propiedades mentales en la física fundamental, contrariamente al programa del fisicalismo. No encajan en la descripción de grano grueso sin algo similar a la teoría de los dos aspectos, cuya debilidad se ha señalado en el párrafo precedente; y no encajan en las caracterizaciones relaciónales a menos que el objeto esté causalmente conectado con un sujeto sensible. En resumen, las apariencias sensoriales no tienen cabida en una ontología fisicalista.

Estos dos argumentos contra el fisicalismo son simples pero robustos. Es difícil ver cómo podrían ser rebatidos y cómo podría considerarse la mente como ontológicamente derivada, si no fuera por varias consideraciones de peso. La primera es que no hay la más mínima evidencia de la existencia de mentalidad aparte de los sistemas nerviosos altamente desarrollados. Como dice Roger: «Si la *mente* es algo completamente externo al cuerpo físico, resulta difícil ver por qué tantos de sus atributos pueden asociarse muy estrechamente con propiedades de un cerebro físico» *(Sombras,* pág. 371). La segunda es el inmenso conjunto de evidencias de que las estructuras neuronales son productos de la evolución a partir de organismos primitivos privados de tales estructuras y, de hecho, si el programa de la evolución prebiótica es correcto, la genealogía puede extenderse más atrás hasta las moléculas inorgánicas y los átomos. La tercera consideración es que la física fundamental no atribuye propiedades mentales a estos constituyentes inorgánicos.

La *Filosofía del organismo* de A.N. Whitehead[6] (que tenía un antecedente en la monadología de Leibniz) tiene una ontología mentalista que toma en cuenta estas tres consideraciones precedentes, pero con matizaciones sutiles. Sus entidades últimas son *ocasiones reales*, que no son entidades persistentes sino cuantos espacio-temporales, dotados cada uno de ellos -normalmente con un nivel muy bajo- de características mentalistas como *experiencia, inmediatez subjetiva y apetito.* Los significados de estos conceptos se derivan de la mentalidad de alto nivel que conocemos por introspección, pero están enormemente extrapolados a partir de esta base familiar. Una partícula física elemental, que Whitehead concibe como una cadena temporal de ocasiones, puede ser caracterizada sin muchas pérdidas con los conceptos de la física ordinaria, porque su experiencia es oscura, monótona y repetitiva; pero de todas formas hay cierta pérdida: «La noción de energía física, que está en la base de la física, debe concebirse entonces como una abstracción a partir de la energía compleja, emocional y con propósito, inherente a la forma subjetiva de la síntesis final en la que cada ocasión se completa en sí misma»[6]. Solo la evolución de asociaciones de ocasiones altamente organizadas permite que la mentalidad primitiva llegue a ser intensa, coherente y completamente consciente: «los funcionamientos de la materia inorgánica permanecen intactos entre los funcionamientos de la materia viva. Parece que, en cuerpos que son obviamente vivos, se ha conseguido una coordinación que hace que destaquen algunas funciones inherentes en las ocasiones definitivas»[7].

El nombre de Whitehead no aparece en el índice de *Sombras* y su única aparición en *La nueva mente del emperador*[8] se refiere a los *Principia Mathematica* de Whitehead y Russell. No conozco las razones del olvido de Roger, pero puedo plantear algunas objeciones mías con las que él podría estar de acuerdo. Whitehead ofrece su ontología mentalista como un remedio para la *bifurcación de la naturaleza* en el mundo mecánico de la física y en la mente de la conciencia de alto nivel. El bajo nivel de protomentalidad que él atribuye a todas las ocasiones pretende cubrir este enorme hueco. Pero, ¿no hay una bifurcación comparable entre la protomentalidad de las partículas elementales y la experiencia de alto nivel de los seres humanos? Y, ¿hay alguna evidencia directa para la protomentalidad de bajo nivel? ¿La hubiera postulado alguien excepto para establecer

continuidad entre el Universo primitivo y el Universo actual habitado por organismos conscientes? Y, si no hay otra razón que esta, ¿no sería una equivocación el morfema «mental» en la palabra «protomental»? ¿No se convierte la filosofía entera del organismo en un truco semántico para tomar un problema y rebautizarlo como una solución? Además, ¿no constituye el concepto de ocasiones reales como entidades concretas definitivas del Universo un tipo de atomismo, más rico que el de Demócrito y Gassendi, pero, en cualquier caso, inconsistente con el carácter holístico de la mente que revela nuestra experiencia de alto nivel?

En la sección siguiente sugiero que estas objeciones pueden responderse en cierta medida elaborando un *whiteheadianismo* modernizado, que utiliza algunos conceptos extraídos de la mecánica cuántica[9].

4.2. La importancia de las ideas de la teoría cuántica para el problema mente-cuerpo

El concepto más radical de la teoría cuántica es que un estado completo de un sistema -es decir, uno que especifica el sistema al máximo- no se agota en un catálogo de propiedades reales del sistema sino que debe incluir potencialidades. La idea de potencialidad está implícita en el principio de superposición. Si se especifican una propiedad A de un sistema cuántico y un vector estado Φ (que suponemos por conveniencia de norma unidad), entonces Φ puede expresarse en la forma $\Sigma_i c_i u_i$,donde cada u_i es un vector de estado de

---

[6] Whitehead A. *N., Adventures of Ideas,* cap. 11, secc. 17.

[7] [7] Whitehead A. N., *Adventures of Ideas,* cap. 13, secc. 6.

[8] [8] Penrose, R. (1989), *The Emperor's New Mind,* Oxford University Press, Oxford. [Existe traducción castellana: (1991) *La nueva mente del emperador,* Mondadori, Madrid; reeditada en (1993) Crítica, Barcelona.]

[9] [9] Shimony, A. (1965), «Quantum physics and the philosophy of Whitehead», *Philosophy in America,* ed. Black, M, George Allen & Unwin, Londres; reimpresión: Shimony, A. (1993), *Search for a Naturalistic World View,* Vol. II, págs. 291-309, Cambridge University Press, Cambridge; Shimon, M. (1988), *A Whiteheadian approach to Bell's correlations,* Foundations of Physics (18), pág. 1035.

norma unidad que representa un estado en el que A tiene un valor definido a_i, *y* donde cada c_i es un número complejo tal que la suma de $|c_i|^2$ es igual a la unidad. Entonces Φ es una superposición de los u_i con pesos apropiados y, a menos que la suma contenga solo un único término, el valor de A en el estado representado por Φ no está definido. Si el estado cuántico se interpreta de forma realista, como una representación del sistema tal como es, más que como un compendio de nuestro conocimiento acerca de él, y si la descripción cuántica es completa, no susceptible de ser completada por *variables ocultas,* entonces esta indefinición es objetiva. Además, si el sistema interacciona con su entorno de forma tal que A se hace definido, por ejemplo, por medio de una medida, entonces el resultado es una cuestión de azar objetivo, y las probabilidades $|c_i|^2$ de los diversos resultados posibles son probabilidades objetivas. Estas características de indefinición objetiva, azar objetivo y probabilidad objetiva, se resumen caracterizando el estado cuántico como una red de potencialidades.

El segundo concepto radical de la teoría cuántica es el de enmarañamiento *(entanglement).* Si u_i son vectores de estado de norma unidad que representan estados del sistema I, en los que cierta propiedad A tiene valores distintos en estos estados, y v_i. son vectores de estado del sistema II, en los que una propiedad B tiene valores distintos, entonces existe un vector de estado $x = \Sigma_i c_i u_i v_i$ (donde los $|c_i|^2$ suman la unidad) del sistema compuesto I + II con características peculiares. Ni I ni II por separado están en un estado cuántico puro. En particular, I no es una superposición de los u_i *y* II no es una superposición de los v_i, pues tales superposiciones omiten el modo en que están correlacionados los u_i y los v_i. El estado x es un tipo de estado holístico, llamado *enmarañado.* La teoría cuántica tiene así un modo de composición que no encuentra análogo en la física clásica. Si ocurre un proceso por el que A se hace actual, por ejemplo, tomando el valor a_i, entonces B será también automáticamente actualizado y tendrá el valor b_i. El enmarañamiento entraña así que las potencialidades de I y II son actualizadas simultáneamente.

El *whiteheadianismo* modernizado al que yo me refería crípticamente al final de la sección 4.1 incorpora los conceptos de potencialidad y enmarañamiento de forma esencial. La potencialidad es el instrumento con el que puede salvarse la embarazosa bifurcación en la oscura protomentalidad y la conciencia de alto nivel. Incluso un organismo complejo con un cerebro altamente desarrollado puede hacerse inconsciente. La transición entre conciencia e inconciencia no tiene por qué ser interpretada como un cambio de estado ontológico, sino como un cambio de estado, y las propiedades pueden pasar de la definición a la indefinición, y a la inversa. En el caso de un sistema simple, como un electrón, uno no puede imaginar nada más que una transición desde una completa indefinición de experiencia a un mínimo atisbo. Pero, en esta unión entra en juego el segundo concepto, el enmarañamiento. En efecto, para un sistema de muchos cuerpos en estados enmarañados hay un espacio mucho más rico de propiedades observables que para una única partícula, y los espectros de los observables colectivos son mucho más amplios que los de las partículas componentes. El enmarañamiento de sistemas elementales, cada uno de ellos con un estrecho abanico de atributos mentales, puede generar, concebiblemente, un abanico más amplio que cubre todo el camino que va desde la inconciencia hasta la conciencia de alto nivel.

¿Cómo se compara este *whiteheadianismo* modernizado con la aplicación que hace Roger de las ideas cuánticas al problema mente-cuerpo? En el capítulo 7 de *Sombras,* y en los capítulos *2 y 3,* Roger hace un uso esencial de las dos grandes ideas de potencialidad y enmarañamiento. La potencialidad es invocada en su conjetura de que las *computaciones cuánticas* son ejecutadas por un sistema de neuronas, donde cada rama de una superposición realiza un cálculo independiente de los realizados en otras ramas *(Sombras,* págs. 375-376). El enmarañamiento (al que Roger se refiere normalmente como «coherencia») es invocado en varias fases para dar cuenta de la realización de estos cálculos: se supone que los microtúbulos en las paredes de las células juegan un papel organizador en el funcionamiento de las neuronas, y, para este propósito, se postula un estado enmarañado de un microtúbulo *(Sombras,* págs. 388-389); se supone luego que los microtúbulos de una misma neurona están en un estado enmarañado; y, finalmente, hay un supuesto estado enmarañado de un gran número de neuronas. Es necesario el enmarañamiento a gran escala porque: «la unidad de una sola mente puede aparecer, en una descripción semejante, solo si existe alguna forma de coherencia cuántica que abarca, al menos, una parte apreciable del cerebro entero» *(Sombras,* pág. 394). Roger sostiene que su propuesta es plausible a la vista de los fenómenos de superconductividad y superfluidez, especialmente de superconductividad a alta temperatura, y de los cálculos de Fróhlich de que el enmarañamiento a gran escala es posible en sistemas biológicos a temperatura corporal *(Sombras,* págs. 388-389). Una idea cuántica adicional en el tratamiento de la mente por parte de Penrose está sacada, no de la teoría cuántica vigente, sino de la teoría cuántica que él prevé en el futuro y que será discutida en la sección 4.3. Esta idea es la reducción objetiva (abreviado RO) de una superposición, en la cual se selecciona un valor real de un observable A, entre un intervalo inicialmente

grande de valores posibles. Que tal actualización es indispensable para una teoría de la mente es algo que parece exigido por los fenómenos indudables de los pensamientos y las sensaciones definidas en nuestra experiencia consciente. Es necesario incluso si existe algo tal como la computación cuántica, puesto que al final del procesamiento paralelo en las diferentes ramas de una superposición debe leerse un *resultado (Sombras, pág.* 376). Finalmente, Roger conjetura que RO proporcionará los aspectos no computables de la actividad mental.

Desde un punto de vista *whiteheadiano* modernizado, lo que está ausente -inadvertida o deliberadamente- en la teoría de Roger de la mente es la idea de mentalidad como algo ontológicamente fundamental en el Universo. La explicación de Roger suena sospechosamente parecida a una versión cuántica del fisicalismo. En las versiones del fisicalismo mencionadas en la sección 4.1, las propiedades mentales eran tratadas como propiedades estructurales de estados cerebrales o como programas para realizar cálculos en conjuntos neuronales. Roger suministra nuevos ingredientes para el programa de explicar físicamente la mentalidad -a saber, coherencia cuántica a gran escala y una supuesta modificación de la dinámica cuántica para explicar la reducción de las superposiciones-. Pero esta sofisticación no debilita el sencillo, aunque robusto, argumento contra el fisicalismo ofrecido en la sección 4.2. Las apariencias de nuestra vida mental no tienen lugar en una ontología fisicalista, y un fisicalismo gobernado por reglas cuánticas sigue siendo fisicalismo. La filosofía del organismo de Whitehead, por el contrario, es radicalmente no fisicalista, puesto que atribuye propiedades mentalísticas a las entidades más primitivas del Universo, enriqueciendo conjeturalmente de este modo su descripción física. La versión modernizada de *whiteheadianismo* que propuse provisionalmente no utiliza la teoría cuántica como un sucedáneo del estado ontológico fundamental de la mentalidad, sino como un instrumento intelectual para explicar la inmensa gama de manifestaciones de mentalidad en el mundo, desde la total depresión de la mentalidad intrínseca hasta la ampliación de alto orden de la misma.

El contraste puede expresarse de otra forma. La teoría cuántica es una herramienta, que despliega conceptos tales como *estado, observable, superposición, probabilidad de transición y enmarañamiento.* Los físicos han aplicado con éxito esta herramienta a dos ontologías muy diferentes: la ontología de partículas, en la mecánica cuántica no relativista estándar de electrones, átomos, moléculas y cristales; y la ontología de campos, en electrodinámica cuántica, cromodinámica cuántica y teoría cuántica de campos en general. Es concebible que la teoría cuántica pueda aplicarse a ontologías completamente diferentes, tales como una ontología de mentes, una ontología dualista o una ontología de entidades dotadas de protomentalidad. Las aplicaciones fisicalistas usuales de la teoría cuántica han sido maravillosamente fértiles en la explicación de los fenómenos observables de sistemas compuestos, incluyendo sistemas macroscópicos, en términos microscópicos. Me parece que Roger está tratando de hacer algo similar al explicar los fenómenos mentales en una ontología fisicalista mediante un empleo delicado de conceptos cuánticos. El *whiteheadianismo* modernizado, por el contrario, aplica el marco de la teoría cuántica a una ontología que es *ab initio* mentalista. Por supuesto, el *whiteheadianismo* modernizado es embrionario, impresionista, y carece de predicciones y confirmaciones experimentales que le pudieran dar sus credenciales como *teoría prometedora.* Pero tiene la gran virtud de reconocer la inderivabilidad de la mentalidad, que está ausente en todas las variedades de fisicalismo. Puede ser que yo haya leído u oído mal a Roger, y que, de hecho, él sea más *whiteheadiano* de lo que yo haya advertido. Si esto es así o no, que hiciera explícita la cuestión clarificaría mucho su postura.

Si una versión modernizada de Whitehead, o cualquier teoría cuántica de la mente, va a alcanzar madurez y solidaridad científica, tendrá que prestar mucha atención a los fenómenos psicológicos. Existen algunos fenómenos que tienen un cierto *sabor cuántico:* por ejemplo, el paso de la visión periférica a la visión focal; el paso de la conciencia a la inconciencia; la penetración de la mente a través del cuerpo; la intencionalidad; las anomalías en la localización temporal de los sucesos mentales; y las refundiciones y ambigüedades del simbolismo freudiano. Varios libros importantes sobre la relación entre la mente y la teoría cuántica, en particular los de Lockwood[10] y Stapp[11], han examinado fenómenos mentales que tienen un sabor cuántico. El propio Roger discute algunos de estos fenómenos, por ejemplo los experimentos de Kornhuber y Libet sobre los aspectos de temporización activa y pasiva de la conciencia *(Sombras,* págs. 407-408).

[10] Lockwood, M. (1989), *Mind, Brain and the Quantum,* Blackwell, Oxford.

[11] Stapp, H. P. (1993), *Mind, Matter and Quantum Mechanics,* Springer-Verlag, Berlin.

Una seria aplicación de la teoría cuántica a la mente debe considerar también las estructuras matemáticas del espacio de estados y el conjunto de observables. Estas no son suministradas por el marco cuántico. En el caso de la mecánica cuántica no relativista estándar y de la teoría cuántica de campos, estas estructuras están determinadas de varias formas: por consideraciones de representación de grupos espacio-temporales, por heurística basada en mecánica clásica y en teoría clásica de campos y, por supuesto, por el experimento. Uno de los grandes artículos de Schródinger sobre mecánica ondulatoria de 1926 presenta una analogía maravillosamente fructífera: la óptica geométrica es a la óptica ondulatoria lo que la mecánica de partículas es a una hipotética mecánica ondulatoria. ¿No podría ser heurísticamente valioso considerar una nueva analogía: la física clásica es a la física cuántica lo que la psicología clásica es a una hipotética psicología cuántica? Por supuesto, una de las dificultades para explotar esta analogía es que la estructura de la *psicología clásica* es mucho peor conocida y está quizá menos definida intrínsecamente que la estructura de la mecánica clásica.

He aquí una sugerencia adicional. Posiblemente, los conceptos cuánticos puedan ser aplicados a la psicología, pero no con tanta estructura geométrica como en física cuántica. Incluso si hay algo semejante a un espacio de estados mentales, ¿podemos suponer que este espacio tendrá la estructura de un espacio de Hilbert proyectivo? En particular, ¿habrá un producto interno definido entre dos estados mentales que determine la probabilidad de transición entre uno y otro? ¿No podría darse el caso de que exista una estructura más débil en la Naturaleza, aunque sea una estructura de tipo cuántico? Hay artículos muy interesantes de Mielnik[12] que sugieren que un concepto cuántico mínimo es la expresabilidad de un estado *mezcla* como combinación convexa de estados puros en más de una forma, mientras que en la mecánica estadística clásica un estado mezcla puede expresarse de forma unívoca en términos de estados puros. Una especulación más es que la fenomenología de los colores puede construirse como un ejemplo de la idea de Mielnik -por ejemplo, muchas formas diferentes de componer blanco perceptual a partir de una mezcla de luz coloreada-.

4.3. El problema de la actualización de las potencialidades

En el capítulo 2, Roger clasificó el problema de la actualización de las potencialidades (también llamado problema de la reducción del paquete de ondas y problema de la medida) como un misterio-x, un misterio que no puede resolverse sin un cambio radical de la propia teoría y no un misterio que puede ser exorcizado por hábito. Estoy completamente de acuerdo. Si la teoría cuántica describe objetivamente un sistema físico, entonces hay observables del sistema que están objetivamente indefinidos en un estado especificado pero que se hacen definidos cuando se realiza una medida. Pero la dinámica lineal de la teoría cuántica impide la actualización por medio de la medida. Una consecuencia de la linealidad es que el estado final del sistema compuesto por el aparato de medida más el objeto es una superposición de términos en los que el observable «puntero» del aparato tiene valores diferentes. Yo comparto el escepticismo de Roger sobre todos los intentos de interpretar este misterio, por ejemplo, mediante las interpretaciones de los muchos universos, decoherencia, variables ocultas, etc. En una u otra fase en un proceso de medida, la evolución unitaria del estado cuántico deja de ser válida y ocurre una actualización. Pero, ¿en qué fase? Existen muchas posibilidades.

La fase puede ser física, y ocurrir cuando un sistema macroscópico se enmaraña con un objeto microscópico, o cuando la métrica espacio-temporal se enmaraña con un sistema material; o puede ser mental, y desarrollarse en la psique del observador. Roger opta por la hipótesis de que la actualización es un proceso físico, debido a la inestabilidad de una superposición de dos o más estados de la métrica espacio-temporal; cuanto mayor es la diferencia de energía entre los estados superpuestos, menor es el tiempo de vida de la superposición *(Sombras,* págs. 359-367). Sin embargo, la conjunción de esta conjetura con la determinación de Roger de explicar experiencias reales en la conciencia impone algunas serias restricciones. Él necesita la superposición de estados cerebrales, como se indicó antes, para dar cuenta de la globalidad de la mente, pero monstruosidades tales como la superposición de ver un destello rojo y ver un destello verde deben, o bien no ocurrir en absoluto, o bien ser tan transitorias que difícilmente inciden en la conciencia. Roger argumenta -provisionalmente y en esbozo- que las diferencias de energía en los estados cerebrales correspondientes a tales percepciones distintas son suficientemente grandes para dar un tiempo de vida corto de la superposición. Sin embargo, él admite en varios pasajes *(Sombras,* págs. 419, 420, 440, 363-364) que está tratando de realizar un difícil paseo por la cuerda floja, pues debe mantener suficiente coherencia para explicar la globalidad de la mente y suficiente ruptura de coherencia para explicar los

[12] Mielnik, B. (1974), «Generalized quantum mechanics», *Communications in Mathematical Physics* (37), pág. 221.

sucesos conscientes definidos. El cómo podría ser robusto en la operación diaria un cerebro/mente que actúa según las líneas esbozadas por Roger es realmente algo muy misterioso.

Los recursos de la familia de modificaciones de la dinámica cuántica con objeto de explicar objetivamente la actualización de potencialidades no han sido todavía completamente explorados, ni por Roger ni por la comunidad de investigadores. Mencionaré brevemente dos caminos que encuentro atractivos. El modelo de reducción espontánea, de Ghirardi-Rimini-Weber y otros, es mencionado y convincentemente criticado por Penrose *(Sombras*, pág. 365), pero puede haber variantes de esta dinámica que escapen a sus críticas. Un segundo camino, que él no menciona, es la posibilidad de una *regla de superselección* en la naturaleza que impida la superposición de distintos isómeros o conformaciones de macromoléculas. La motivación para esta conjetura es la consideración de que las macromoléculas actúan típicamente a modo de interruptores en la célula, activando o desactivando procesos de acuerdo con la conformación molecular. Si dos conformaciones distintas estuvieran superpuestas, tendríamos un análogo celular al gato de Schródinger -un proceso en un limbo entre la ocurrencia y la no ocurrencia-. Si la naturaleza obedeciera a una regla de superselección que prohibe tales superposiciones se evitarían las dificultades, pero la razón sería misteriosa: ¿por qué la naturaleza prohibe superposiciones de estados de conformación de moléculas complejas cuando los permite para las sencillas?, ¿dónde está la línea divisoria? Sin embargo, una superselección semejante podría explicar todas las actualizaciones de potencialidades de las que tenemos buena evidencia, y puede tener la preciosa propiedad de ser comprobable por espectroscopia molecular[13].

Finalmente, vale la pena apuntar que, desde un punto de vista *whiteheadiano,* la hipótesis de que la actualización de potencialidades se logra por la psique del perceptor no es tan ridicula, antropocéntrica, mística y acientífica como normalmente se considera. Según Whitehead, algo parecido a la mentalidad permea toda la naturaleza, pero la mentalidad de alto nivel es contingente en la evolución de complejos de ocasiones especialmente hospitalarios. La capacidad para que un sistema actualice potencialidades, modificando con ello la mecánica lineal de la mecánica cuántica, puede estar omnipresente en la naturaleza, pero solamente no es despreciable en sistemas con mentalidad de alto nivel. Yo matizaría esta expresión de tolerancia, sin embargo, diciendo que la atribución a la psique del poder de reducir superposiciones debería tomarse en serio solo si sus implicaciones para un amplio abanico de fenómenos psicológicos son cuidadosamente desarrolladas, pues solo entonces habría una posibilidad de someter las hipótesis a la prueba experimental controlada.

[13] Quack M. (1989), «Structure and dynamics of chiral molecules», *Angew. Chem. Int.*, Ed. Engl-(28), pág. 571.

5
¿Por qué Física?

NANCY CARTWRIGHT

Nosotros discutimos el libro de Roger Penrose *Las sombras de la mente* en una serie de seminarios conjuntos LSE/King's College de Londres con el título: «Filosofía: Ciencia o Teología». Quiero empezar planteando la misma pregunta que me formuló uno de los participantes en el seminario: «¿Qué razones tiene Roger para pensar que las respuestas a preguntas sobre la mente y la conciencia deben encontrarse en la física antes que en la biología?». Hasta donde yo podía ver, existen tres tipos de razones que Penrose sugiere.

Primer argumento

Si lo hacemos de ese modo, podemos establecer un programa muy prometedor. Este es el tipo de razón potencialmente más poderosa que puede darse para un proyecto como el de Roger. Como positivista que soy, opuesta al mismo tiempo a la metafísica y a la argumentación trascendental, yo estaría dispuesta a discutir que este es el único tipo de argumento al que deberíamos dar peso. Por supuesto, la fuerza con que este tipo de argumento apoye un proyecto dependerá de lo prometedor -y lo detallado- que sea el programa. Una cosa que está clara es que la propuesta de Penrose -postular primero coherencia cuántica macroscópica en los microtúbulos del citoesqueleto y luego buscar las características no computacionales especiales de la conciencia en un nuevo tipo de interacción cuántico-clásica- no es un programa detallado. Sus promesas no residen ciertamente en el hecho de que sea un próximo paso natural en una agenda de investigación progresiva y bien verificada. Si uno la encuentra prometedora, debe ser debido a la audacia e imaginación de las ideas, a la convicción de que alguna nueva interacción de este tipo es necesaria en cualquier caso para poner en orden la mecánica cuántica, y al fuerte compromiso previo de que, si va a haber una explicación científica para la conciencia, debe, en definitiva, ser una explicación *física.* Creo que esto último debe jugar seguramente un papel clave si vamos a juzgar prometedor el programa de Roger. Pero, obviamente, en la medida en que juega un papel, el hecho de que juzguemos el programa prometedor no puede proporcionarnos una razón para pensar que sea la física, y no alguna otra ciencia, la que haga la tarea.

Segundo argumento

El segundo tipo de argumento para pensar que la física por sí sola proporcionará la explicación última es el hecho indudable de que ciertas áreas de la física -especialmente el electromagnetismo- contribuyen a nuestra comprensión del cerebro y el sistema nervioso.

La descripción estándar actual de la transmisión de mensajes utiliza conceptos de la teoría de circuitos eléctricos. Parte de la propia historia de Roger se basa en ideas muy recientes tomadas del electromagnetismo: se supone que los diferentes estados de polarización eléctrica en un dímero de tubulina constituyen la base de las diferencias en la configuración geométrica que hacen que los dímeros se curven formando ángulos diferentes con respecto al microtubo. Pero este tipo de argumento no funciona. El hecho de que la física cuente parte de la historia es una pobre razón para concluir que debe contar toda la historia.

A veces se invoca a la química en esta etapa para argumentar lo contrario. Nadie negaría ahora que una parte de la historia será contada por la química. Pero se supone que las partes relevantes de la química son, en sí mismas, simple física. Así es exactamente como el propio Roger habla de ello: «Las fuerzas químicas que controlan las interacciones de átomos y moléculas tienen en verdad un origen mecano-cuántico, y es fundamentalmente la acción química la que gobierna el comportamiento de las sustancias *neurotransmisoras* que transfieren señales de una neurona a otra -a través de minúsculas separaciones que se denominan *espacios sinópticos*—. Análogamente, los potenciales de acción que controlan físicamente la transmisión de la señal nerviosa tienen un origen ciertamente mecano-cuántico» *(Sombras,* pág. 368). La química entra en juego en defensa de la física para responder a mis preocupaciones, provocadas por el gigantesco salto inferencial desde «la física cuenta parte de la historia» a «la física cuenta toda la historia». Pero ahora este mismo salto inferencial ha reaparecido una vez más en un nivel inferior. Ciertamente no tenemos nada semejante a una reducción real de las partes relevantes de la química-física a la física -ya sea

cuántica o clásica[14]-. La mecánica cuántica es importante para explicar aspectos de los fenómenos químicos pero los conceptos cuánticos son utilizados siempre junto con conceptos *sui generis* -es decir, no reducidos- de otros campos. Ellos no explican los fenómenos por sí solos.

Tercer argumento

La tercera razón para pensar que la física explicará la mente es metafísica. Podemos ver la cadena argumental de Roger. Nos gustaría suponer que la función de la mente *no* es *misteriosa;* eso significa que puede explicarse en *términos científicos;* lo que equivale a que puede explicarse en términos *físicos.* En mi seminario, la pregunta: *¿por qué no biología?,* fue planteada por el bien conocido experto en estadística James Durbin, y creo que la cuestión encierra gran interés. Como experto en estadística, Durbin vive en un mundo abigarrado y estudia pautas de características que proceden de todo tipo de campos, tanto científicos como prácticos. Por el contrario, el mundo de Roger es el mundo del *sistema unificado,* con la física como base para la unificación. Pienso que la razón para este tipo de fisicismo es la idea de que no tenemos ninguna otra metafísica satisfactoria. Sin el sistema nos quedamos con algún tipo de dualismo inaceptable o, para utilizar el mismo término que Roger, misterioso. Este es el tema que quiero discutir[15], pues creo que la opinión de que no hay alternativa razonable es una idea que ha calado en muchos físicos. Existe la sensación de que cualquiera que tome la física seriamente como algo que describe realmente el mundo tendrá que creer en su hegemonía.

¿Por qué? Aparentemente, existe un número muy, muy elevado, de propiedades diferentes que actúan simultáneamente en el mundo. Algunas son estudiadas por una disciplina científica específica; otras lo son por otra disciplina diferente; unas cuantas están en la intersección de ciencias distintas; y, la mayoría, no son el objeto de estudio de ninguna ciencia. ¿Qué es lo que legitima la visión de que, detrás de las apariencias, todas estas propiedades son realmente las mismas? Yo creo que dos cosas: una es una confianza excesiva en la sistematicidad de sus interacciones y la otra, una excesiva estimación de lo que la física ha conseguido.

Debería señalar, sin embargo, que esta limitación en la visión metafísica que ve como posible solo un tipo de monismo fisicista está también ampliamente extendida en filosofía, incluso entre aquellos que se resisten a la reducción de las ciencias particulares a la física. Consideremos la filosofía de la biología, en la que el reduccionismo quedó pasado de moda hace bastante tiempo; ahora se toma de nuevo en serio un tipo de emergentismo en el que las propiedades y las leyes aparecen recién nacidas con niveles de complejidad y organización crecientes. Pese a todo, la mayor parte de las personas no puede ir más allá de un tipo de monismo; se sienten obligadas a insistir en la *sobrevenida.* En términos generales, decir que las propiedades de la biología descansan en las de la física es afirmar que, si tuviéramos dos situaciones que fueran idénticas con respecto a sus propiedades físicas, entonces deberían ser idénticas con respecto a sus propiedades biológicas. Esto no significa, según ellos, que las leyes biológicas se reduzcan a leyes físicas, puesto que las propiedades biológicas no necesitan ser definibles en términos físicos. Pero sí significa que las propiedades biológicas no son propiedades independientes y separadas, pues están fijadas por las propiedades de la física. Una vez que se ha establecido la descripción física, la descripción biológica no puede ser sino la que es. Las propiedades biológicas no se sitúan en una categoría completamente independiente: son ciudadanas de segunda clase.

Tomar en serio que las propiedades biológicas son propiedades separadas, causalmente efectivas por sí mismas, no es burlarse de la evidencia empírica. Doy por hecho lo que vemos en la ciencia: a veces la física ayuda a explicar lo que sucede en sistemas biológicos. Pero aquí sucede lo mismo que dije de la química: rara vez sin la ayuda también de descripciones biológicas *sui generis y* no reducidas. Podemos recordar una frase que he utilizado en otro lugar de un modo algo diferente: *ni dentro de la biología, ni . fuera de la biología*[16]. *Lo* que vemos se describe de forma más natural como una interacción entre

[14] Ver Hendry R. F, «Approximations in quantum chemistry», *Idealisation in Contemporary Physics*, Niall Shanks (ed.), Poznán Studies in the Philosophy of the Sciences and Humanities, Rodopi, Amsterdam (próximo 1997).

[15] Para detalles sobre argumentos en contra del sistema único, ver Dupre, J.(1993), *The Disorder of Things: Metaphysical Foundations of Disunity of Science,* Harvard University Press, Cambridge MA; Neurath, O. (1987), *Unified Science,* Vienna Circle Monograph Series, traducción: Kael, H. (D. Reidel: Dordrecht).

[16] Durante la discusión, Abner Shimony hizo los siguientes comentarios en relación con este punto:

características biológicas y físicas que se afectan mutuamente. También tenemos identificaciones muy contextúales entre una descripción biológica y una física, así como una buena cantidad de cooperación causal -propiedades biológicas y físicas que actúan conjuntamente para producir efectos que ninguna de ellas puede causar por sí sola-. Pasar de esto a *todo debe ser física* es precisamente el salto inferencial gigante por el que me he venido preocupando. Lo que vemos puede ser consistente con que todo sea física, pero ciertamente no apunta hacia esa conclusión en particular sino que parece más bien apartarse de ella[17].

En buena medida, la razón para creer que todo debe ser física está, creo yo, en una visión acerca del cierre. Se supone que los conceptos y las leyes de una buena teoría física constituyen un sistema cerrado en sí mismo: es todo lo que uno necesita para ser capaz de hacer predicciones sobre estos mismos conceptos. Pienso que esta es una errónea —o, al menos, excesivamente optimista- visión del éxito de la física. Casi al mismo tiempo en que la idea de *sobrevenida* se hizo importante en filosofía, también lo hizo la idea de una ciencia particular. Esencialmente todas las ciencias excepto la física son ciencias particulares. Eso significa que, en el mejor de los casos, sus leyes son válidas solo *ceteris paribus:* válidas en tanto no intervenga nada que esté fuera del dominio de la teoría en cuestión.

Pero ¿qué es lo que genera la confianza en que las leyes de la física son más que leyes *ceteris paribus?* Nuestros sorprendentes éxitos de laboratorio no muestran tal cosa; ni lo hace el éxito newtoniano con el sistema planetario, que tanto impresionó a Kant. Y tampoco lo hacen las grandes transferencias tecnológicas de la física: los tubos de vacío o los transitores o los magnetómetros SQUID. Pues estos dispositivos son construidos para asegurar que no hay interferencia. No prueban si las leyes siguen siendo buenas cuando factores externos al dominio de la teoría juegan un papel. Existe por supuesto la fe general en que, en el caso de la física, nada podría interferir salvo factores adicionales que en sí mismos pueden ser descritos en el lenguaje de la física y que están sujetos a sus leyes. Pero este precisamente es el punto en cuestión.

Quiero terminar con un comentario acerca del realismo. He estado apuntando hacia un tipo de visión pluralista de todas las ciencias, que permanecen codo a codo y aproximadamente en pie de igualdad con varios tipos diferentes de interacciones que se producen entre los factores estudiados en sus distintos dominios. Esta es una imagen que, a menudo, va acompañada de una visión en la que la ciencia es una construcción humana que no refleja la naturaleza. Pero esta no es una conexión necesaria. Kant mantenía la postura exactamente opuesta: es precisamente porque nosotros construimos la ciencia por lo que el sistema unificado es no solo posible sino necesario. De todas las maneras, en nuestros días esta imagen pluralista suele estar asociada con el construccionismo social. Por eso es importante resaltar que *pluralismo* no implica *antirrealismo.* Decir que las leyes de la física son verdaderas *ceteris paribus* no es negar que sean verdaderas; simplemente, no son completamente soberanas. No es el realismo en la física el que está en juego bajo el pluralismo, sino más bien el imperialismo. Por eso no quiero que nos dirijamos a una discusión del realismo científico. Más bien quiero que Penrose discuta su compromiso metafísico de que

---

«Nancy Cartwright defiende que hay que discutir la mente en el contexto de la biología antes que en el de la física. Yo aplaudo la parte positiva de su petición. Por supuesto, hay mucho que aprender sobre la mente desde la biología evolutiva, la anatomía, la neurofisiología, la biología del desarrollo, etc. Pero no estoy de acuerdo en que la investigación de la relación entre la mente y la física sea estéril. Deberían estudiarse las conexiones entre disciplinas con la máxima profundidad posible. No sabemos *a priori* dónde llevarán estas investigaciones, y los resultados han sido muy diferentes en dominios diferentes. Así, el Teorema de Bell y los experimentos que inspiró han mostrado que las correlaciones que se manifiestan en sistemas enmarañados espacialmente separados no pueden ser explicadas por ninguna teoría que atribuya estados definidos al sistema individual -un gran triunfo para el holismo-, La demostración de Onsager de que el modelo de Ising bidimensional experimenta transiciones de fase muestra que el orden a gran alcance puede manifestarse en un sistema infinito en el que los componentes solo interaccionan con sus vecinos más próximos -un triunfo para el punto de vista analítico y para la reducibilidad de la macrofísica a la microfísica-. Ambos tipos de descubrimientos -holístico o analítico- revelan algo importante acerca del mundo. La investigación de las relaciones entre disciplinas no infringe la validez de las leyes fenomenológicas dentro de las disciplinas. Tales investigaciones pueden proporcionar una heurística para leyes fenomenológicas refinadas, y también pueden ofrecer una comprensión más profunda de tales leyes. Cuando Pasteur sugirió que la quiralidad de las moléculas es responsable de la rotación del plano de polarización de la luz que atraviesa las disoluciones, él fundó la estéreo-química.»

[17] Para una discusión adicional sobre este punto, ver Cartwright N. (1993), *Synihese* (94), pág. 291. Para una discusión más elaborada del punto de vista general sugerido aquí, ver Cartwright N. (1994), *Proceedings of the Aristotelian Society y* (1995) *Physik, Philosophic uncí die Einheit der Wissenschaft, Philosophia Naturalis,* ed. L. Kreuger y B. Falkenburg, Spektrum, Heidelberg.

debe ser la física la que haga el trabajo. Pues eso debe ser presupuesto si la discusión se plantea ya acerca de si será este tipo de física o aquél. La cuestión no es si las leyes de la física son verdaderas y sostienen de alguna forma la actuación de la mente, sino si ellas son toda la verdad o si deben llevar el grueso de la carga explicativa.

6
Las objeciones de un reduccionista descarado

STEPHEN HAWKING

Para empezar, debería decir que soy un reduccionista descarado. Creo que las leyes de la biología pueden reducirse a las de la química -ya hemos visto cómo sucede esto con el descubrimiento de la estructura del ADN-; y pienso, además, que las leyes de la química pueden reducirse a las de la física. Confío en que la mayoría de los químicos estarán de acuerdo con esto.

Roger Penrose y yo trabajamos juntos en la estructura a gran escala del espacio y del tiempo, incluyendo singularidades y agujeros negros. Coincidimos bastante en la teoría clásica de la relatividad general pero los desacuerdos empezaron a surgir cuando entramos en la gravedad cuántica. Ahora tenemos enfoques muy diferentes con respecto al mundo, físico y mental. Básicamente, él es un platónico que cree que existe un único mundo de ideas que describe una única realidad física. Yo, por el contrario, soy un positivista que cree que las teorías físicas son simplemente modelos matemáticos que nosotros construimos, y que es absurdo preguntarse si se corresponden con la realidad; sólo hay que cuestionarse si predicen o no observaciones.

Esta diferencia de enfoque ha llevado a Roger a hacer tres afirmaciones en los capítulos 1, 2 y 3, con las que estoy en fuerte desacuerdo. La primera es que la gravedad cuántica provoca lo que él llama RO, reducción objetiva de la función de onda. La segunda es que este proceso tiene un papel importante en el funcionamiento del cerebro por su efecto en flujos coherentes a través de las moléculas. Y la tercera es que, debido al teorema de Godel, se necesita algo parecido a la RO para explicar la autoconciencia. Empezaré con la gravedad cuántica, que es lo que conozco mejor. Su reducción objetiva de la función de onda es una forma de decoherencia. Esta decoherencia puede venir por interacciones con el entorno o por fluctuaciones en la topología del espacio-tiempo. Pero Roger no parece necesitar ninguno de estos mecanismos. En su lugar, afirma que ocurre debido a la ligera curvatura del espacio-tiempo producida por la masa de un objeto pequeño. Pero, según las ideas aceptadas, esta curvatura no es obstáculo para una evolución hamiltoniana sin decoherencia o reducción objetiva. Quizá las ideas aceptadas sean erróneas pero Roger no ha presentado una teoría detallada que nos permita calcular cuándo ocurriría la reducción objetiva.

Parece que la motivación de Roger al proponer la reducción objetiva ha sido rescatar al pobre gato de Schródinger de su estado medio-vivo, medio-muerto. Ciertamente, en estos días de defensa de los derechos de los animales nadie se atrevería a sugerir un procedimiento semejante, ni siquiera como experimento intelectual. Sin embargo, Roger llegó a afirmar que la reducción objetiva era un efecto tan débil que no podría distinguirse experimentalmente de la decoherencia causada por interacción con el entorno. Si este es el caso, entonces la decoherencia ambiental puede explicar el gato de Schródinger. No existe la necesidad de invocar la gravedad cuántica. A menos que la reducción objetiva sea un efecto suficientemente fuerte para ser medido experimentalmente, no puede hacer lo que Roger quiere que haga.

La segunda afirmación de Roger era que la reducción objetiva tenía una influencia significativa en el cerebro, quizá mediante su efecto en flujos coherentes a través de los micro-túbulos. Yo no soy un experto en el funcionamiento del cerebro, pero esto me parece muy poco probable, incluso si yo creyera en la reducción objetiva, cosa que no hago. No puedo pensar que el cerebro contenga sistemas que están suficientemente aislados para que la reducción objetiva pueda ser distinguida de la decoherencia ambiental. Si estuviesen tan bien aislados no podrían interaccionar con la rapidez suficiente para tomar parte en los procesos mentales.

La tercera afirmación de Roger es que la reducción objetiva es, de algún modo, necesaria porque el teorema de Gódel implica que una mente consciente no es computable. En otras palabras, Roger cree que la conciencia es algo propio de los seres vivos y que no podría ser simulada en un ordenador. No deja claro de qué forma podría la reducción objetiva explicar la conciencia. Más bien, su argumento parece consistir en que la conciencia es un misterio y que la gravedad cuántica es otro misterio, de modo que deben estar relacionados.

Personalmente me siento incómodo cuando las personas, en especial los físicos teóricos, hablan sobre la conciencia. La conciencia no es una cualidad que uno pueda medir desde fuera. Si un hombrecillo verde apareciera mañana en el descansillo de nuestra puerta, no tendríamos forma de decir si él era

consciente y autoconsciente, o era simplemente un robot. Yo prefiero hablar de inteligencia, que es una cualidad que puede medirse desde fuera, y no veo ninguna razón por la que la inteligencia no pueda ser simulada en un ordenador. Nosotros, ciertamente, no podemos simular inteligencia humana por el momento, como demostró Roger con su problema de ajedrez. Pero Roger admitió también que no existía línea divisoria entre inteligencia humana e inteligencia animal. Por eso será suficiente considerar la inteligencia de una lombriz. No creo que haya ninguna duda de que uno puede simular el cerebro de una lombriz en un ordenador. El argumento de Godel es irrelevante porque las lombrices no se preocupan por las Π_1-sentencias.

La evolución desde los cerebros de lombrices hasta los cerebros humanos tuvo lugar presumiblemente por selección natural darwiniana. La cualidad seleccionada era la capacidad para escapar de los enemigos y para reproducirse, y no la capacidad para hacer matemáticas. Por eso, una vez más, el teorema de Godel no es relevante. Se trata simplemente de que la inteligencia necesaria para la supervivencia puede utilizarse también para construir demostraciones matemáticas. Pero es una ocupación muy dura y complicada. Ciertamente no tenemos un procedimiento cognosciblemente válido.

Yo les he dicho por qué estoy en desacuerdo con las tres afirmaciones de Penrose acerca de que haya reducción objetiva de la función de onda, que esta tenga un papel en el funcionamiento del cerebro y que sea necesaria para explicar la conciencia. Ahora debería dejar que Roger responda.

7
Roger Penrose responde

Agradezco los comentarios de Abner, Nancy y Stephen, y deseo hacer algunas puntualizaciones como respuesta. En lo que sigue, contestaré por separado a cada uno de ellos.

Respuesta a Abner Shimony

En primer lugar, permítanme decir que aprecio mucho los comentarios de Abner, que creo son extraordinariamente útiles. Sin embargo, él sugiere que al concentrarme en la cuestión de la computabilidad quizá esté intentando escalar la montaña equivocada. Si, con esto, está señalando que hay muchas manifestaciones importantes de la mentalidad distintas de la no-computabilidad, entonces estoy completamente de acuerdo con él. También estoy de acuerdo en que el argumento de la habitación china de Searle proporciona un alegato convincente contra la posición de la *IA-fuerte* que sostiene que la sola computación puede evocar mentalidad consciente. El argumento original de Searle estaba interesado en la cualidad mental de *comprensión*, como lo está mi propia discusión *godeliana*, pero la habitación china puede utilizarse también (quizá con fuerza aún mayor) contra otras cualidades mentales, tales como la sensación de un sonido musical o la percepción del color rojo. La razón por la que no he utilizado esta línea argumental en mi propia discusión, sin embargo, es que se trata de un argumento de carácter puramente negativo y no nos proporciona ninguna clave real con respecto a lo que verdaderamente está sucediendo con la conciencia, ni nos indica ninguna dirección a seguir si queremos avanzar hacia el descubrimiento de una base científica para la mentalidad.

La línea de razonamiento de Searle está interesada solamente en la distinción A/B, según la terminología que adopté en el capítulo 3 (cf. también, *Sombras*, págs. 26-31). Es decir, desea mostrar que los aspectos *internos* de la conciencia no quedan englobados en la computación. Esto no es suficiente para mí, porque yo necesito mostrar que las manifestaciones *externas* de la conciencia tampoco son alcanzables por computación. Mi estrategia no es tratar en esta etapa los problemas internos, mucho más difíciles, sino tratar de hacer algo mucho más modesto para empezar, intentando comprender qué tipo de física podría dar lugar al tipo de comportamiento externo que puede ser exhibido por un ser consciente -de modo que es la distinción A/C o B/C la que me interesa en esta etapa-. Mi alegato es que aquí sí es posible algún progreso. Por supuesto, aún no estoy intentando acometer un asalto mayor a la *verdadera* cumbre, pero opino que, si podemos Primero escalar con éxito una de sus estribaciones importantes, entonces estaremos en mejores condiciones para descubrir el camino hacia la cima real desde nuestra nueva posición de ventaja.

Abner se refiere a mi(s) carta(s) en respuesta a la recensión que hizo Hilary Putnam de *Sombras*, cuando comenta que él no ha quedado convencido con lo que yo tenía que decir. De hecho, yo no hice ningún intento real de responder a Putnam en detalle porque yo no creía que la sección de cartas en una revista fuera el lugar apropiado para entrar en una discusión detallada. Simplemente quería señalar que, en mi opinión, las críticas de Putnam eran una parodia. Resultaban particularmente irritantes porque no daba ninguna impresión de haber leído siquiera aquellas partes del libro que estaban dirigidas a los mismos puntos que él planteaba. Habrá una respuesta mucho más detallada en la revista (electrónica) *Psyche*, donde se abordan varias recensiones de *Sombras*, que espero responda a los puntos que le preocupan a Abner[18]. De hecho, creo que el argumento *godeliano* es, de raíz, muy potente, incluso si algunas personas se muestran muy reacias a admitirlo. ¡No voy a ceder en lo que creo que es un argumento básicamente correcto simplemente por las dificultades que les plantea a ciertas personas! Pienso que nos suministra una clave importante con respecto al tipo de física que posiblemente podría subyacer al fenómeno de la conciencia, incluso si este por sí solo no nos va a proporcionar la respuesta.

Creo que estoy básicamente de acuerdo con los puntos positivos que Abner cita. Está intrigado por la ausencia de mención a la obra filosófica de A. N. Whitehead en *Emperador* o en *Sombras.* La razón principal de esto es mi desconocimiento. No quiero decir que ignore la posición general de Whitehead, que

[18] Ya aparecida: enero de 1996; http//psyche.es.monash.edu.au/psiche-index-v2.l.html; y ahora existe una versión impresa, publicada por *un press* (1996).

mantiene una forma de *panpsiquismo.* Quiero decir que yo no había leído en detalle la obra filosófica de Whitehead y por eso hubiera sido reacio a hablar acerca de ella o acerca de su proximidad o cualquier otra relación con mis propias ideas. Creo que mi posición general no está muy lejana a lo que Abner está proponiendo, aunque yo no estaba preparado para hacer ningún enunciado definido en esta línea, debido en parte a una carencia de convicción clara acerca de lo que yo creo realmente.

Encuentro el *whiteheadianismo modernizado* de Abner particularmente sorprendente, con una sugestiva plausibilidad. Me doy cuenta ahora de que el tipo de pensamiento que debe de haber estado en mi subconsciente está muy cerca de lo que Abner expresa tan elocuentemente. Más aún, él tiene razón en que los *enmarañamientos* a gran escala son necesarios para que la unidad de una única mente aparezca como cierta forma de estado cuántico colectivo. Aunque yo no había afirmado explícitamente, ni en el *Emperador* ni en *Sombras,* la necesidad de que la mentalidad sea *ontológicamente fundamental en el Universo,* pienso que realmente es necesario algo de esta naturaleza. Sin duda hay algún tipo de protomentalidad asociado con cada instancia de RO, según mi propio punto de vista, pero tendría que ser extraordinariamente *minúscula* en un sentido apropiado. Sin ningún enmarañamiento ampliamente extendido con alguna estructura altamente organizada y soberbiamente adaptada a cierto tipo de *capacidad de procesamiento de información* -como ocurre en los cerebros- es presumible que la genuina mentalidad no se manifestaría de forma significativa. Creo que esto se debe solo a que mis propias ideas están tan mal formuladas a este respecto que yo no aventuré enunciados más claros sobre mi posición en estas materias. Estoy ciertamente agradecido a Abner por sus comentarios clarificadores.

También estoy de acuerdo en que quizá puedan obtenerse algunas intuiciones significativas de la exploración de posibles analogías y descubrimientos experimentales en la psicología. Si los efectos cuánticos son fundamentales para nuestros procesos mentales conscientes, entonces deberíamos empezar a ver algo de las implicaciones de este hecho en aspectos de nuestro pensamiento. Por otra parte, uno debe ser extraordinariamente cuidadoso, en este tipo de discusión, para no saltar a conclusiones y caer en falsas analogías. El área entera es un hervidero lleno de trampas potenciales; de eso estoy seguro. Quizá fuera posible, no obstante, realizar experimentos razonablemente claros y sería interesante explorar tales posibilidades. Por supuesto, podrían muy bien realizarse otros tipos de pruebas experimentales que fueran más concretas para la hipótesis de los microtúbulos.

Abner menciona la mecánica cuántica no hilbertiana de Mielnik. Siempre me ha parecido interesante este tipo de generalización del marco de la teoría cuántica, y creo que es algo que debería estudiarse más. Sin embargo, no estoy completamente convencido de que este sea precisamente el tipo de generalización que se necesita. Dos aspectos de esta idea concreta me hacen sentirme incómodo. Uno de estos es que, como sucede con otras aproximaciones a la mecánica cuántica (generalizada), se concentra de hecho en *la matriz densidad,* en lugar de hacerlo en el estado cuántico, como modo de describir la realidad. En la mecánica cuántica ordinaria, el espacio de matrices densidad constituye un conjunto convexo, y los *estados puros,* que se describirían unívocamente mediante un vector de estado, ocurren en la frontera de este conjunto. Esta imagen surge de un espacio de Hubert ordinario, siendo un subconjunto del producto tensorial del espacio de Hubert y su complejo conjugado (es decir, su dual). En la generalización de Mielnik se retiene esta imagen general de *matriz densidad,* pero no hay ningún espacio de Hubert lineal subyacente a partir del que se construya el conjunto convexo. A mí me gusta la idea de generalizar a partir de la noción de un espacio de Hubert lineal, pero me siento incómodo con la pérdida de los aspectos holomórficos (analíticos en el plano complejo) de la teoría cuántica, cuya pérdida parece ser una característica de esta aproximación. No se retiene un análogo del vector de estado, hasta donde puedo ver, sino solo de un vector de estado salvo fase. Esto hace que las superposiciones complejas de la teoría cuántica resulten particularmente oscuras dentro del formalismo. Por supuesto, podría argumentarse que son estas superposiciones las que causan todas las dificultades a escala macroscópica, y quizá uno debería deshacerse de ellas. De todas formas, son absolutamente fundamentales en el nivel cuántico, y creo que en esta forma particular de generalizar las cosas estamos perdiendo la parte positiva más importante de la teoría cuántica.

Mi otra fuente de insatisfacción tiene que ver con el hecho de que los aspectos no lineales de nuestra mecánica cuántica generalizada deberían ser dispuestos para tratar el proceso de medida, siendo un elemento de la *asimetría temporal* implicada aquí (ver *Emperador,* capítulo 7). Yo no veo que este aspecto de las cosas tenga un papel en el esquema de Mielnik tal como está.

Finalmente, me gustaría expresar mi apoyo a la búsqueda de mejores esquemas teóricos en los que se modifiquen las reglas básicas de la mecánica cuántica, y también al diseño de experimentos que puedan ser capaces de distinguir tales esquemas de la teoría cuántica convencional. Por el momento, yo no he

tropezado con ninguna sugerencia para un experimento actualmente factible que sea capaz de poner a prueba el tipo concreto de esquema que propuse en el capítulo 2. Estamos aún a algunos órdenes de magnitud por debajo, pero quizá alguien dé con una idea mejor para una prueba.

Respuesta a Nancy Cartwright

Me siento animado (y halagado) al oír que *Sombras* ha sido discutido seriamente en la serie KSE/King's College a la que Nancy se refiere. Sin embargo, ella se muestra escéptica acerca de que las cuestiones relativas a la mente deban responderse en términos de física más que de biología. En primer lugar, debería dejar claro que ciertamente no estoy diciendo que la biología no sea importante en nuestros intentos de abordar esta cuestión. De hecho, creo que es probable que los avances realmente significativos, en un futuro cercano, vengan desde el lado biológico más que desde el lado físico; pero principalmente porque lo que nosotros necesitamos de la física, en mi opinión, es una revolución fundamental y, ¡quién sabe cuándo llegará esta!

Supongo, no obstante, que este tipo de concesión no es lo que ella pretende -como algo que sirviera para que yo considere a la biología capaz de proporcionar *el ingrediente fundamental* de la comprensión de la mentalidad en términos científicos. De hecho, según mi propio punto de vista sería posible tener una entidad consciente que no fuera biológica en absoluto, en el sentido en que utilizamos el término *biología* actualmente; pero no sería posible que una entidad fuera consciente si no incorporase el tipo particular de *proceso físico* que yo mantengo que es esencial.

Una vez dicho esto, no tengo claro en absoluto cuál es la posición de Nancy con respecto al tipo de línea que hay que trazar entre la biología y la física. Tengo la impresión de que ella está siendo más bien pragmática sobre estas cuestiones al decir que es correcto considerar la conciencia como un problema físico si eso nos ayuda a avanzar. Por eso ella pregunta: ¿puedo señalar realmente un programa de investigación concreto donde los físicos, mejor que los biólogos, puedan ayudarnos a avanzar de un modo fundamental? Creo que mis propuestas sí llevan a un programa mucho más concreto que el que ella parece estar sugiriendo. Afirmo que debemos buscar estructuras en el cerebro con algunas propiedades físicas muy claras. Deberían ser tales que permitan que existan estados cuánticos extendidos espacialmente y bien protegidos, que persistan un tiempo al menos del orden de un segundo, y tal que los enmarañamientos involucrados en dichos estados le proporcionen una dispersión sobre grandes áreas del cerebro que implican probablemente muchos miles de neuronas simultáneamente. Para soportar un estado semejante, necesitamos estructuras biológicas con una construcción interna muy precisa, probablemente con una estructura de tipo cristalino, y capaz de ejercer una influencia importante sobre las intensidades sinápticas. Yo no veo que la transmisión nerviosa ordinaria pueda ser suficiente por sí misma porque no hay posibilidad real de obtener el aislamiento necesario. Las mallas vesiculares presinápticas, como han sugerido Beck y Eccles, podrían estar jugando un papel pero, a mi modo de ver, los micro-túbulos del citoesqueleto parecen mostrar más cualidades relevantes. Puede ser que haya muchas otras estructuras (tales como las clatrinas) en este tipo de escala que son necesarias para la imagen global. Nancy está sugiriendo que mi imagen no es muy detallada; pero me parece que lo es mucho más que casi todas las demás que yo he visto, y es potencialmente desarrollable posteriormente de una forma muy concreta, con muchas oportunidades de comprobación experimental. Estoy de acuerdo en que se necesita mucho más antes de que podamos acercarnos a una imagen *completa* -pero creo que debemos avanzar con cautela, y no espero pruebas definitivas durante bastante tiempo-. Se necesitará mucho más trabajo.

El punto más serio de Nancy parece tener más que ver con el papel que para ella tiene la física en nuestra imagen del mundo global. Pienso que quizá para ella el prestigio de la física está sobrevalorado. Quizá lo está -o, al menos, es muy posible que la imagen del mundo que los físicos actuales tienden a presentar esté muy sobrevalorada con respecto a su proximidad a la compleción, ¡o incluso con respecto a su corrección!-.

Viendo (correctamente, en mi opinión) que la teoría física actual es un mosaico de teorías, Nancy sugiere que quizá siga siendo así para siempre. Quizá el objetivo último del físico de una imagen completamente unificada sea un sueño inalcanzable. Ella adopta el punto de vista de que es metafísica, y no ciencia, abordar siquiera tal cuestión. Yo mismo no estoy seguro de qué actitud tomar sobre esto, pero no creo que necesitemos ir tan lejos al considerar lo que se necesita aquí. La unificación ha sido una tendencia global manifiesta en la física, y veo muchas razones para esperar que esta tendencia continúe. Requeriría una atrevida expresión de escepticismo afirmar lo contrario. Tomemos lo que considero que es la pieza

principal del *mosaico* en la teoría física moderna, a saber, la forma en que están empalmados los niveles de descripción clásico y cuántico: una forma muy poco convincente, en mi opinión. Uno podría seguir la línea de que simplemente debemos aprender a vivir con dos teorías básicamente incompatibles que se aplican en dos niveles diferentes (lo que, supongo, era más o menos la opinión expresada de Bohr). Ahora bien, quizá seamos capaces de abandonar tal actitud durante los próximos años, pero conforme las medidas se hagan más precisas y empiecen a sondear la frontera entre estos dos niveles, estaremos intentando conocer cómo trata en realidad la Naturaleza esta frontera. Quizá el modo en que se comportan algunos sistemas biológicos pudiera depender críticamente de lo que sucede en esta frontera. Supongo que la cuestión es si confiamos en encontrar una teoría matemática bella para tratar lo que ahora nos parece un horrible amasijo, o si la propia física es *realmente* tan solo un amasijo desagradable en este nivel. ¡Seguramente no! No hay ninguna duda de en qué lado están mis instintos en esta cuestión.

No obstante, de los comentarios de Nancy saco la impresión de que ella estaría dispuesta a aceptar un amasijo desagradable en las leyes de la física en esta etapa[19]. Quizá esto es lo que ella podría entender al decir que la biología no es reducible a la física. Por supuesto, podría haber perfectamente muchos parámetros desconocidos y complicados que desempeñan papeles importantes en este nivel en sistemas biológicos. Para tratar tales sistemas, incluso cuando son conocidos todos los principios físicos subyacentes, quizá sea necesario en la práctica adoptar todo tipo de conjeturas, métodos de aproximación, métodos estadísticos y quizá nuevas ideas matemáticas para proporcionar un tratamiento científico razonablemente efectivo. Pero, desde el punto de vista de la física estándar, aunque los detalles de un sistema biológico pudieran presentarnos un amasijo desagradable, esto no supone un amasijo en las propias leyes físicas subyacentes. Si las leyes físicas son completas a este respecto, entonces *las propiedades de la biología se derivarán de las de la física.*

Sin embargo, estoy defendiendo que las leyes físicas estándar no son completas a este respecto. Peor que esto, afirmo que no son completamente correctas en aspectos que podrían ser muy importantes para la biología. La teoría estándar permite una cierta apertura -en el proceso-R de la mecánica cuántica convencional-. En el punto de vista normal, esto simplemente da lugar a una aleatoriedad genuina, y es difícil ver cómo un nuevo principio *biológico* podría tener aquí un papel sin perturbar la autenticidad de esta aleatoriedad -lo que significaría cambiar la teoría física-. Pero yo estoy afirmando que las cosas son peores que todo esto. El procedimiento-R de la teoría estándar es *incompatible* con la evolución unitaria (U). Dicho claramente, el proceso de evolución-U de la teoría cuántica estándar es completamente inconsistente con los hechos observacionales manifiestos. En el punto de vista estándar uno evita esto mediante varios artificios con diferentes grados de plausibilidad, pero el hecho bruto permanece. En mi opinión, no hay duda de que esto es un problema físico, cualquiera que pudiera ser su relación con la biología. Posiblemente es un punto de vista coherente el que una Naturaleza *mosaico* pudiera vivir simplemente con esta situación -pero dudo mucho que nuestro mundo sea así en realidad-.

Más allá de este tipo de cosas, yo no entiendo sencillamente cómo podría ser una biología que no se derivara de la física. Lo mismo se aplica a la química. (Con esto no pretendo faltar el respeto a ninguna de estas dos disciplinas.) Algunas personas me han expresado algo parecido al decir que no pueden imaginar una física cuya acción sea no-computable. Este no es un sentimiento antinatural, pero el *modelo de juguete* de universo que describí en el capítulo 3 proporciona alguna idea de cómo podría ser una física no-computable. Si alguien puede darme análogamente una idea de lo que pudiera ser una *biología* que no se derive de *su física* correspondiente, entonces yo podría empezar a tomar esa idea en serio.

---

[19] Durante la discusión, Nancy Cartwright reiteró su posición a este respecto: «Roger cree que una física que no puede tratar sistemas abiertos es una mala física. Yo, por el contrario, creo que puede ser realmente muy buena física -si las leyes de la naturaleza son un mosaico, como yo imagino que pueden serlo-. Si el mundo está lleno de propiedades que no son reducibles a las de la física, pero que interaccionan causalmente con aquellas que sí lo son, entonces la física más precisa será necesariamente una física *ceterís paribus* que solo puede contar toda la historia cuando se trata de sistemas cerrados.

¿Cuál de estos puntos de vista es probable que sea el correcto? Eso, creo yo, es una cuestión metafísica, en el sentido de que cualquier respuesta a ella va mas allá de la evidencia empírica de la que disponemos, incluyendo la de la historia de la ciencia. Yo animo a evitar este tipo de metafísica siempre que sea posible, y, cuando las decisiones metodológicas requieren un compromiso en una dirección u otra, animo a cubrirnos con fuerza de nuestras apuestas. Allí donde debiéramos apostar, yo estimaría las probabilidades de forma muy diferente de aquellos que ponen toda su fe en la física. La ciencia moderna es un mosaico, no un sistema unificado. Si tenemos que hacer apuestas sobre la estructura de la realidad, creo que deberíamos proyectarlas a partir de la mejor representación que tenemos de dicha realidad -y esa es la ciencia moderna tal como existe, no como imaginamos que podría existir-.»

Permítanme volver a lo que yo considero que es la cuestión principal de Nancy Cartwright: ¿por qué creo que debemos buscar una nueva física para una explicación científica de la conciencia? Mi respuesta breve es que, de acuerdo con la discusión de Abner Shimony, sencillamente, yo no veo ningún lugar para la mentalidad consciente dentro de nuestra actual imagen del mundo físico -siendo la biología y la química parte de dicha imagen del mundo-. Más aún, no veo cómo podemos cambiar la biología para que no sea parte de dicha imagen del mundo sin cambiar también la física. ¿Aún querríamos seguir llamándola una imagen del mundo *basada en la física* si contiene elementos de protomentalidad en un nivel básico? Esta es una cuestión de terminología, pero es una cuestión con la que me siento razonablemente feliz al menos por el momento.

Respuesta a Stephen Hawking

Los comentarios de Stephen acerca de que él es un positivista podrían llevar a uno a esperar que también simpatizara con la imagen *mosaico* de la física. Pese a todo, él considera que los principios estándar de la mecánica cuántica U son inmutables, hasta donde puedo apreciar, en su propia aproximación a la gravedad cuántica. En realidad no veo por qué simpatiza tan poco con la genuina posibilidad de que la evolución unitaria fuera una aproximación a algo mejor. Yo mismo me siento cómodo con que sea una aproximación de algún tipo -del mismo modo que la teoría gravitatoria soberbiamente precisa de Newton es una aproximación a la de Einstein-. Pero creo que eso tiene muy poco que ver con la cuestión platonismo/positivismo como tal.

No estoy de acuerdo en que la decoherencia ambiental sola pueda des-superponer al gato de Schrodinger. Mi punto sobre la decoherencia ambiental consistía en que una vez que el entorno llega a estar inextricablemente enmarañado con el estado del gato (o con cualquier sistema cuántico bajo consideración), no parece que suponga ninguna diferencia práctica cuál es el esquema de reducción objetiva que decida uno seguir. Pero sin *algún* esquema para la reducción, incluso si es simplemente algún esquema provisional PTPP (para todo propósito práctico), el estado del gato seguiría siendo simplemente una superposición. Quizá, según la posición *positivista* de Stephen, él no se preocupa realmente de cuál sea el estado del gato unitariamente evolucionado, y preferiría una descripción de la *realidad* en términos de matriz densidad. Pero esto no nos evita, de hecho, el problema del gato, como demostré en el capítulo 2, al no haber nada en la descripción de la matriz densidad que asegure que el gato está o muerto o vivo, y no en alguna superposición de ambos.

En lo que se refiere a mi propuesta concreta de que la reducción objetiva (RO) es un efecto gravitacional cuántico, Stephen está en lo cierto al decir que «según las ideas físicas aceptadas, la curvatura [espacio-temporal] no será obstáculo para una evolución hamiltoniana», pero la dificultad está en que, sin la intervención de un proceso RO, las separaciones entre las diferentes componentes espacio-temporales pueden hacerse cada vez mayores (como sucede con el gato), y parecerán apartarse cada vez más de la experiencia. Sí, yo creo que las ideas aceptadas deben ser erróneas en esta etapa. Más aún, aunque mis ideas distan de ser enteramente detalladas con respecto a lo que yo creo que debe estar sucediendo a este nivel, he sugerido al menos un criterio que es, en principio, susceptible de prueba experimental.

Con respecto a la probabilidad de la relevancia de tales procesos para el cerebro, estoy de acuerdo en que esto parecería ser *muy improbable* -si no fuera por el hecho de que algo muy extraño está sucediendo realmente en el cerebro consciente que, en mi opinión (y también en la de Abner Shimony), está más allá de lo que podemos entender en términos de nuestra actual imagen del mundo físico-. Por supuesto, este es un argumento negativo, y uno debe ser muy cauto para no extralimitarse utilizándolo. Creo que es muy importante buscar en la neurofisiología real del cerebro, y también en otras áreas de la biología, de forma extremadamente cuidadosa, para tratar de ver qué está sucediendo en realidad.

Finalmente, está mi uso del argumento de Gödel. La base para utilizar este tipo de discusión consiste en que es algo que puede medirse desde el exterior (es decir, estoy interesado en la distinción A/C o B/C, como mencioné antes, y no en la distinción A/B no medible externamente). Más aún, con respecto a la selección natural, el punto preciso que yo estaba señalando consistía en que no era una habilidad específica para hacer matemáticas lo que se seleccionaba. Si hubiera sido así, habríamos quedado atrapados en la camisa de fuerza gödeliana, y no lo estamos. Toda la clave del argumento, a este respecto particular, reside en que lo que se seleccionaba era una capacidad general para comprender -que, como característica accidental, podía también aplicarse a la comprensión matemática-. Esta capacidad no tiene por qué ser una capacidad no-algorítmica (debido al argumento gödeliano), sino que se aplica a muchas otras áreas además

de las matemáticas. Yo no sé si las lombrices, pero estoy seguro de que los elefantes, los perros, las ardillas y muchos otros animales la comparten en gran medida.

Apéndice 1
El teorema de Goodstein y el pensamiento matemático

En el capítulo 3 di una demostración de una versión del teorema de Gödel en apoyo de mi afirmación de que la comprensión humana debe implicar ingredientes que no pueden ser simulados mediante procedimientos computacionales. Pero la gente suele encontrar dificultad para apreciar la relevancia del teorema de Gödel para nuestra forma de pensar, incluso en el caso del pensamiento matemático. Una razón para esto es que, de acuerdo con la forma en que el teorema se presenta normalmente, el enunciado *indemostrable* real que genera el procedimiento de Gödel parece no tener relevancia para ningún resultado matemático de interés.

Lo que el teorema de Gödel nos dice es que para cualquier procedimiento de *demostración* computacional *P* (suficientemente extenso), que estamos dispuestos a considerar como incuestionablemente válido, es posible construir una proposición aritmética precisa *G(P)* cuya verdad también debemos aceptar como incuestionablemente establecida pero que es inaccesible mediante el procedimiento de demostración original *P*. La dificultad abordada aquí es que el enunciado matemático real G(P) que nos proporciona la aplicación *directa* de las prescripciones de Gödel sería enormemente difícil de aprehender y no tendría ningún interés matemático intrínseco obvio, aparte del hecho de que sabemos que es verdadero pero no derivable utilizando *P*. En consecuencia, incluso los matemáticos se sienten con frecuencia felices de pasar por alto enunciados matemáticos como G(P).

Pese a todo, existen ejemplos de enunciados de Gödel que son fácilmente accesibles, incluso para aquellos que no tienen una familiaridad particular con la terminología o la notación matemática más allá de la utilizada en la aritmética ordinaria. Un ejemplo particularmente sorprendente llamó mi atención en 1996, durante una conferencia de Dan Isaacson. Se trata del resultado conocido como *Teorema de Goodstein*[20]. Creo que es instructivo exponer aquí explícitamente el teorema de Goodstein para que el lector pueda tener alguna experiencia directa de un teorema de tipo Gödel[21].

Para apreciar lo que afirma el teorema de Goodstein, consideremos cualquier número entero positivo, por ejemplo, 581. En primer lugar, expresamos este número como una suma de distintas potencias de 2:

$$581 = 512 + 64 + 4 + 1 = 2^9 + 2^6 + 2^2 + 2^\circ$$

(Esto es lo que habría que hacer para construir la representación *binaria* del número 581, a saber, 1001000101, donde los unos representan las potencias de 2 que están presentes en el desarrollo y los ceros, aquellas que están ausentes). Se advertirá que los exponentes en esta expresión (los números 9, 6 y 2) podrían representarse también de este modo: $9 = 2^3 + 2^\circ$, $6 = 2^2 + 2^1$, $2 = 2^1$; y obtenemos (incorporando $2^\circ = 1$ y $2^1 = 2$):

$$581 = (2^2)^{3+1} + (2^2)^2 + {}^2 + 2^2 + 1$$

Hay todavía un exponente en el orden siguiente, el 3, para el que esta representación puede ser adoptada una vez más ($3 = 2^1 + 2^\circ$), y podemos escribir:

$$581 = (2^2)^{2+1+1} + (2^2)^{2+2} + 2^2 + 1$$

[20] R. L. Goodstein (1944), «On the restricted ordinal theorem», *Journal of Symbolic Logic, 9*, págs. 33-41.

[21] Ver también R. Penrose (1997), «On understanding understanding», *International Studies in the Philosophy of Science*, 11, pág. 20.

Para números más grandes, quizá tengamos que llegar al tercer orden o a órdenes mayores de exponentes.

Apliquemos ahora a esta expresión una sucesión de operaciones simples, consistente en *alternar* sucesivamente las dos siguientes: *(a)* aumentar la base en 1 y *(b)* restar 1.

La base mencionada en *(a)* es simplemente el número 2 en las expresiones precedentes, pero podemos encontrar representaciones similares para bases mayores: 3, 4, 5, 6, y así sucesivamente. Veamos qué sucede cuando aplicamos (a) a la última expresión dada más arriba para 581, de modo que todos los doses se convierten en treses. Obtenemos:

$$((3^3)^{3+1})x3 + (((3)^3)^3)x3 + 3^3 + 1$$

(que es, de hecho, un número de 40 dígitos, cuando se escribe en la forma normal, que empieza por 133027946...). A continuación, apliquemos *(b)*, para obtener:

$$((3^3)^{3+1})x3 + (((3)^3)^3)x3 + 3^3$$

(que, por supuesto, sigue siendo un número de 40 dígitos, que empieza por 133027946...). Apliquemos ahora *(a)* una vez más, para obtener:

$$((4^4)^{4+1})x4 + (((4)^4)^4)x4 + 4^4$$

(que es ahora un número de 618 dígitos, que empieza por 12926802...). La operación *(b)* de restar 1 proporciona ahora:

$$(4^4)^{4+1}x4 + (4^4)^4x4 + 3 \times 4^3 + 3 \times 4^2 + 3 \times 4 + 3$$

(donde los treces surgen de forma análoga a los nueves que aparecen en la notación ordinaria de base 10 cuando restamos 1 de 10.000 para obtener 9.999). La operación (a) da lugar entonces a:

$$(5^5)^{5+1}x5 + (5^5)^5x5 + 3 \times 5^3 + 3 \times 5^2 + 3 \times 5 + 3$$

(que tiene 10.923 dígitos y empieza por 1274...). Nótese que todos los *coeficientes* 3 que aparecen aquí son necesariamente menores que la base (ahora 5) y no son afectados por el incremento en la base. Aplicando *(b)* una vez más, obtenemos:

$$(5^5)^{5+1}x5 + (5^5)^5x5^5 + 3 \times 5^3 + 3 \times 5^2 + 3 \times 5 + 2$$

y continuamos esta alternancia (a), *(b), (a), (b),* (a), *(b)* hasta donde podamos. Los números parecen ser siempre crecientes, y sería natural suponer que esto continúa indefinidamente. Sin embargo, no es así; pues el notable teorema de Goodstein nos dice que, independientemente del número entero positivo del que partamos (aquí 581), ¡siempre terminaremos en cero!

Esto parece extraordinario. Pero, de hecho, es cierto y para hacerse a la idea, yo recomiendo que el lector lo ensaye partiendo primero de 3 (donde $3 = 2^1 + 1$, de modo que nuestra secuencia es 3, 4, 3, 4, 3, 2, 1, 0); pero luego, y más importante, partiendo de 4 (en cuyo caso tenemos $4 = 2^2$, de modo que obtenemos una secuencia que empieza de forma aparentemente obediente con 4, 27, 26, 42,41,61, 60, 84..., pero que, después de alcanzar un número con 121.210.695 dígitos, decrece finalmente hasta cero).

Lo que es bastante más extraordinario es que el teorema de Goodstein es realmente un teorema de Gödel para el procedimiento que aprendemos en la escuela denominado *inducción matemática*[22]. Recordemos que la inducción matemática proporciona una manera de demostrar que ciertos enunciados matemáticos S(n) son válidos para cualquier n = 1,2, 3, 4, 5... El procedimiento consiste en demostrar, primero, que el enunciado es válido para $n = 1$ y demostrar luego que si es válido para n, entonces debe ser también válido para $n + 1$. Un ejemplo familiar es el enunciado:

$$1 + 2 + 3 + 4 + 5 + ... + n = (1/2)\, n\, (n+1)$$

Para demostrar esto por inducción matemática, establecemos primero que es cierto para $n = 1$ (obvio) y confirmamos luego que si la fórmula funciona para n, entonces también funciona para $n + 1$, lo que es ciertamente verdadero porque tenemos:

$$1+2 + 3+4 + 5 + ... + n + (n+1) = (1/2)\, n\, (n+1) + (n+1) = (1/2)\, (n+1)\, ((n+1)+1)$$

Lo que Kirby y Paris demostraron, de hecho, era que si *P* representa el procedimiento de inducción matemática (junto con las operaciones aritméticas y lógicas ordinarias), entonces podemos volver a expresar G(P) en la forma del teorema de Goodstein. Este nos dice que si creemos que el procedimiento de inducción matemática es digno de confianza (lo que difícilmente es una hipótesis dudosa), entonces debemos creer también en la verdad del teorema de Goodstein, ¡pese al hecho de que no es demostrable por la sola inducción matemática!

La *indemostrabilidad,* en este sentido, del teorema de Goodstein no nos impide ciertamente ver que es, de hecho, verdadero. Nuestra intuición nos capacita para trascender los procedimientos limitados de demostración que anteriormente nos habíamos permitido. De hecho, la forma en que el propio Goodstein demostró su teorema consistía en utilizar un ejemplo de lo que se denomina *inducción transfinita.* En el contexto presente, esto proporciona un modo de dar forma a una intuición que puede ser obtenida directamente si uno mismo se familiariza con la razón por la que el teorema de Goodstein es, de hecho, verdadero. Esta intuición puede obtenerse en general examinando algunos casos individuales del teorema de Goodstein. Lo que sucede es que la modesta operación *(b) roe* sin descanso hasta que las torres de exponentes finalmente se vienen abajo, una a una, hasta que no queda ninguna, incluso si esto necesita un número de pasos increíblemente grande.

Lo que todo esto demuestra es que la cualidad de *comprensión* es algo que nunca podrá ser recogido en un conjunto específico de reglas. Además, la comprensión es una cualidad que depende de nuestra conciencia, de modo que, cualquier cosa que sea la responsable del conocimiento consciente, parece estar interviniendo esencialmente cuando la comprensión está presente. Así pues, nuestra conciencia parece ser algo que implica elementos que no pueden ser recogidos en reglas computacionales de ningún tipo; hay, de hecho, razones muy fuertes para creer que nuestras acciones concientes son esencialmente procesos no computacionales.

Ciertamente, hay posibles vías de escape a esta conclusión, y los defensores del punto de vista filosófico computacional con respecto a la mentalidad conciente tendrían que recurrir a una o más de estas. Básicamente, estas vías de escape son que nuestra capacidad para la comprensión (matemática) podría ser el resultado de algún procedimiento de cálculo que es incognoscible debido a su complicación, o quizá cognoscible en principio pero no cognosciblemente correcto, o podría ser inexacto y sólo aproximadamente correcto. En *Las sombras de la mente,* capítulos 2 y 3, abordo todas estas posibles vías de escape con considerable detalle, y recomendaría esta discusión a cualquier lector interesado en seguir estas cuestiones con más atención. Algunos lectores podrían encontrar útil, primero, examinar mi informe en *Psyche, beyond the doubting of a shadow*[23].

---

[22] Esto fue demostrado por L. A. S. Kirby y J. B. Paris en «Accesible independence results for Peano arithmetic», *Bulletin of the London Mathematical Society,* 14, 1982, págs. 285-293.

[23] La referencia y dirección www son citados en la nota a pie de página de la pág. 136. La referencia impresa más completa es: *Psyche,* 2, 1996, págs. 89-129.

Apéndice 2

Experimentos para poner a prueba la reducción de estado inducida gravitatoriamente

En el capítulo 2, esbocé una propuesta según la cual una superposición cuántica de dos estados, que difieren en un desplazamiento de masa significativo, debería reducirse espontáneamente -sin necesidad de realizar ninguna medida externa sobre el sistema- a un estado o el otro. Según esta propuesta concreta, esta *reducción de estado objetiva* (RO) debe ocurrir en una escala de tiempo de, aproximadamente, $T = h/E$, donde E es una energía gravitatoria caracterizada por el desplazamiento entre los dos estados. En el caso de un desplazamiento rígido, podemos considerar que esta energía E es la energía que se necesitaría para desplazar una copia del objeto hasta llevarla fuera del campo gravitatorio de su otra copia, siendo esto equivalente a tomar E como la autoenergía gravitatoria de la *diferencia* entre los campos gravitatorios de las distribuciones de masa en los dos estados.

Recientemente se han avanzado dos ideas relativas a esta cuestión, una teórica y la otra (propuesta) experimental. Ambas tienen importancia a propósito de la queja que planteaba Stephen Hawking (en la pág. 134) al decir que yo «no había presentado una teoría detallada que nos capacitase para calcular cuándo ocurriría una reducción objetiva» y mi respuesta (en la pág. 141) a su comentario, así como mi comentario anterior (en la pág. 107) relativo a los posibles experimentos.

Desde el lado teórico se ha reconocido durante algún tiempo que hay cierta incompleción en mi propuesta, tal como se presenta en este libro y en el epígrafe 6.12 de *Las sombras de la mente* (y una dificultad emparentada con una propuesta íntimamente relacionada presentada por Diósi, 1989), donde no hay que introducir ningún parámetro de escala básico aparte de la constante gravitatoria G (y h y c). Esta incompleción surge del hecho de que no hay un enunciado preciso acerca de qué estados van ser los estados privilegiados a los que un estado general debería reducirse. Si los estados privilegiados fueran *estados de posición*, en los que cada partícula individual tiene una localización *puntual* completamente definida, entonces deberíamos obtener un valor infinito para la energía gravitatoria relevante E y, por consiguiente, una reducción instantánea de cualquier estado, en evidente violación de muchos efectos mecanocuánticos bien confirmados. Pero sin estados privilegiados de un tipo u otro, uno no puede decir qué estados van a ser considerados como *superposiciones* inestables y cuáles van a ser considerados como aquellos estados (los privilegiados) a los que se supone que van a desexcitarse tales superposiciones. (Recordemos que esta desexcitación debe tener una vida media h/E según el esquema **RO.** Para una masa finita, concentrada en un punto, esto conduciría a $E = \infty$). En la formulación original de Diósi (1989) hay un problema relacionado, a saber, la no conservación de la energía, que, como apuntaron Ghirardi, Grassi y Rimini, llevaría a una grave incompatibilidad con la observación. Con la introducción de un parámetro adicional -una longitud fundamental λ- estos autores fueron capaces de eliminar esta incompatibilidad, pero no hay ninguna justificación *a priori* para escoger algún valor particular para λ,[24]. De hecho, en este esquema modificado, un proceso de reducción de estado localizaría a alguna partícula individual en una región cuyo diámetro es del orden de λ, y no en un punto preciso.

En el esquema que estoy proponiendo, no debería haber ningún parámetro adicional, tal como λ. Todo debería estar fijado por las constantes fundamentales (de relevancia) que ya tenemos, a saber, G, h *y* c (y c no es en sí misma de relevancia en el régimen no relativista). ¿Cómo, entonces, van a ser especificados los *estados privilegiados?* La idea consiste en que, suponiendo que las velocidades son pequeñas comparadas con c *y* los potenciales gravitatorios son también pequeños, dichos estados van a ser soluciones *estacionarias* de lo que yo llamo la *ecuación de Schrodinger-Newton.* Esta ecuación es simplemente la ecuación de Schrodinger (no relativista) para una función de onda ψ pero en donde existe un término adicional dado por un potencial gravitatorio newtoniano Φ, donde la fuente de Φ es el valor esperado de la distribución de masa determinada por ψ. En general, esto lleva a un complicado sistema de ecuaciones en derivadas parciales no lineales y acopladas, que están siendo aún investigadas. Incluso en el caso de una única partícula puntual, no es una cuestión trivial extraer las soluciones estacionarias de esta ecuación que se comporten adecuadamente, incluyendo el infinito. Pero trabajos recientes demuestran que las soluciones

[24] Ghirardi, G. C., Grassi, R. y Rimini (1990), «A. Continuous-spontaneus-reduction model involving gravity», *Physical Review*, A42, págs. 1.057-1.065.

requeridas existen de hecho para este caso de una única partícula puntual, y esto da algún apoyo matemático a la propuesta[25].

El punto crucial, por supuesto, es si un esquema de esta naturaleza está de acuerdo con lo que realmente sucede en una superposición cuántica macroscópica. Es interesante el hecho de que ciertas propuestas para poner a prueba experimentalmente esta cuestión pueden ser factibles. Aunque técnicamente muy difíciles, estos experimentos propuestos no parecen requerir nada más que lo que ya puede conseguirse en principio con la tecnología actual. La idea consiste en colocar un cristal minúsculo, quizá no mucho mayor que una mota de polvo, en una superposición cuántica de dos localizaciones mínimamente desplazadas, y averiguar si esta superposición puede mantenerse coherentemente durante una fracción apreciable de segundo sin que su estado superpuesto se desexcite espontáneamente a un estado o el otro. De acuerdo con mi esquema propuesto, esbozado más arriba, tal desexcitación debería tener lugar, mientras que el punto de vista convencional sostendría que la superposición debería mantenerse indefinidamente, a menos que intervenga alguna otra forma de decoherencia para contaminar el estado.

Permítanme describir el esquema general de un montaje experimental que podría utilizarse para ello[26]. La disposición experimental básica se indica en la figura. He ilustrado el montaje con un fotón como partícula incidente. Sin embargo, debería quedar claro que esto se hace fundamentalmente para facilitar la descripción. La versión para un laboratorio en tierra de este experimento podría perfectamente realizarse mejor utilizando algún otro tipo de partícula incidente, tal como un neutrón o un átomo neutro del tipo apropiado. La razón para esto es que el fotón que necesitaríamos utilizar en el experimento -si es realmente un fotón- tiene que ser un fotón de rayos X, y la construcción de la cavidad necesaria para un fotón semejante constituiría un desafío técnico muy considerable. (En la versión para un laboratorio espacial, la *distancia* entre las dos plataformas espaciales que se utilizarían haría el papel de la cavidad). Para conveniencia de la descripción, en lo que sigue me referiré a la partícula simplemente como un *fotón*, cualquiera que sea la partícula incidente utilizada.

Una fuente de fotones dirige un único fotón hacia un divisor de haz. El divisor de haz realiza entonces una división del estado cuántico del fotón en dos partes de igual amplitud. Un brazo de la superposición resultante del estado del fotón (la parte reflejada) va a mantenerse durante, digamos, una décima de segundo, aproximadamente, sin pérdida de coherencia de fase. En el experimento terrestre, esto se conseguiría manteniendo al fotón en algún tipo de cavidad; en el experimento espacial, el fotón se transmite a un espejo de rayos X en una plataforma espacial independiente, quizá a una distancia de un diámetro terrestre. En el otro brazo del estado del fotón, el fotón incide en un pequeño cristal -que contiene, digamos, unos 10^{15} núcleos- donde es reflejado, impartiendo en este proceso una fracción significativa de su momento al cristal. En el experimento terrestre, esta parte del estado fotónico, reflejado en el cristal, se mantiene en una cavidad similar (o quizá la misma) que la otra parte del estado del fotón; en el experimento espacial, esta segunda parte del estado del fotón es también enviada al espejo en la plataforma espacial. El cristal debe ser tal que el momento entero del impacto del fotón en el mismo debe ser compartido por todos los núcleos del cristal, que actúa como un cuerpo rígido (como sucede con un cristal Móssbauer) sin que haya una probabilidad significativa de excitación de modos vibracionales internos. El cristal está sometido a algún tipo de fuerza recuperadora -indicada en la figura por un muelle- de intensidad tal que vuelve a su posición original en, digamos, una décima de segundo. En ese instante, en

[25] Ver Moroz, I., Penrose, R. y Tod, K. P. (1998), «Spherically-symmetric solutions on the Schrodinger-Newton equations», *Classical and Quantum Gravity*, 15. Moroz, L. y Tod, K. P. An analytic approach to the Schrodinger-Newton equation, próximo a publicarse.

[26] Estoy agradecido a varios colegas por sus sugerencias en relación con esto. Más concretamente, Johannes Dapprich sugirió la idea de que un pequeño cristal (tipo Móssbauer) podría ser el objeto apropiado para colocar en una superposición lineal de dos localizaciones ligeramente diferentes. Una ayuda considerable acerca de cuestiones de viabilidad, y sugerencias específicas sobre las escalas apropiadas para el experimento, vinieron de Antón Zeilinger y varios miembros de su grupo experimental en el Instituto de Física Experimental de la Universidad de Innsbruck. La versión espacial de este experimento fue el resultado de discusiones con Anders Hansson. Ver Penrose, R. (1998), «Quantum Computation, entanglement and state reduction», *Philosophical Transactions of the Royal Society, London*, 356, págs. 1927-1939, para una descripción preliminar de la versión en tierra del experimento.

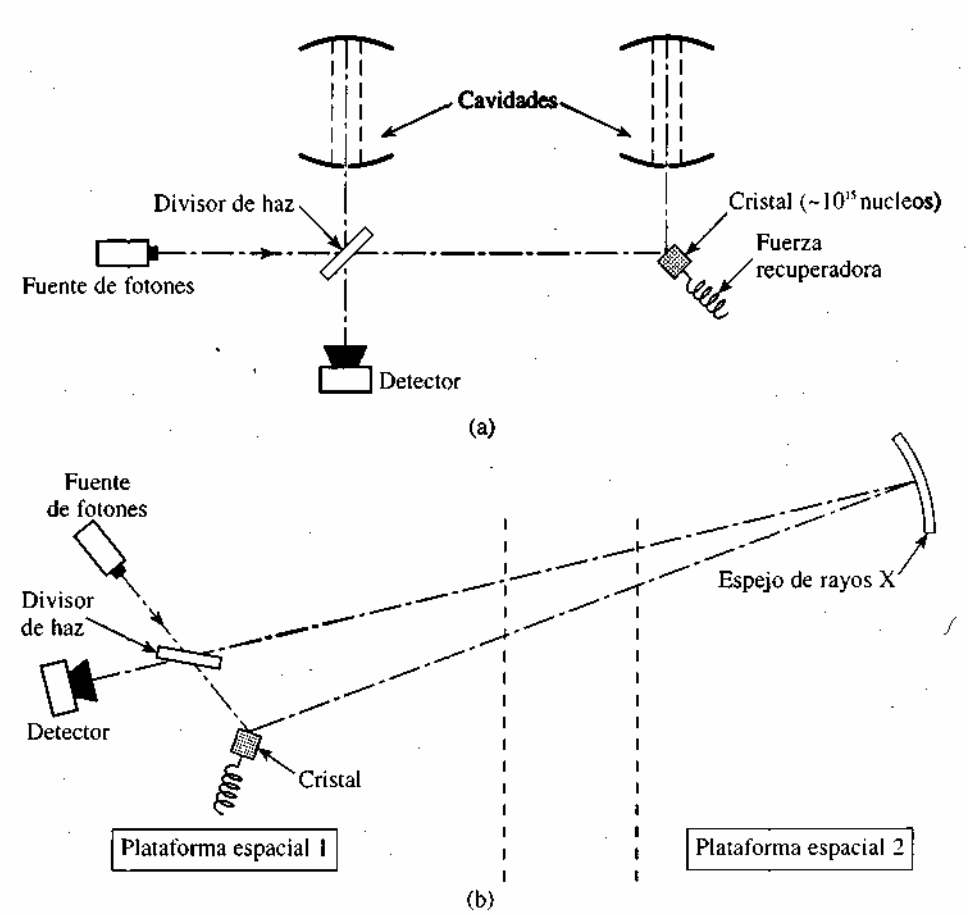


Fig. 1. (*a*) Experimento sugerido para un laboratorio en tierra; (*b*) experimento sugerido para un laboratorio espacial.

la versión terrestre, la parte del estado del fotón que incidió en el cristal es liberada de la cavidad de modo que invierte su camino, cancelando al hacerlo la velocidad de retorno del cristal. La otra parte del estado del fotón es entonces liberada también, con precisión cronométrica, de modo que las dos partes llegan juntas al divisor de haz original. En la versión del laboratorio espacial, el espejo en la plataforma espacial refleja cada parte del estado del fotón al lugar de donde procedía la otra, en la plataforma espacial principal, y el resultado es similar. En cualquiera de las versiones, con tal de que no haya habido pérdida de coherencia de fase en todo el proceso, las dos partes del estado del fotón se combinan coherentemente en el divisor del haz y salen por el mismo camino por el que entraron, de modo que un detector colocado en el haz de salida alternativa del divisor de haz no detectaría nada.

Ahora bien; según mi propuesta, la superposición de las dos localizaciones del cristal, que persiste durante aproximadamente una décima de segundo en las descripciones precedentes, sería inestable, con un tiempo de desexcitación de ese orden. Esto supone que la función de onda del cristal es tal que el valor esperado de la distribución de masa de las localizaciones de los núcleos está fuertemente concentrado alrededor de sus posiciones nucleares medias. Así pues, según esta propuesta, habría una gran probabilidad de que las posiciones superpuestas del cristal (un «gato de Schródinger») se reduzcan espontáneamente a una localización u otra. El estado del fotón está inicialmente enmarañado con el del cristal, de modo que la reducción espontánea del estado del cristal entraña una reducción simultánea del estado del fotón. En esta circunstancia, el fotón ahora ha «seguido un camino o el otro» y ya no es una superposición de los dos, de modo que se pierde la coherencia de fase entre los dos haces y hay una probabilidad significativa (calculable) de que el detector encuentre el fotón.

Por supuesto, en cualquier experimento real de esta naturaleza tiene que haber probablemente muchas otras formas de decoherencia que podrían destrozar la interferencia entre los dos haces de vuelta. Si todas estas otras formas de decoherencia pudieran reducirse hasta un grado suficientemente pequeño, entonces, variando los parámetros implicados (tamaño y naturaleza específica del cristal, la distancia que se desplaza en relación al espaciado de la red, etc.), sería posible identificar la rúbrica particular del tiempo de

decoherencia inherente en el esquema RO que estoy defendiendo. Se pueden considerar muchas modificaciones a este experimento propuesto. (En una de estas, sugerida por Lucien Hardy, se utilizan dos fotones, y puede haber algunas ventajas en la versión en tierra por el hecho de que no es necesario mantener coherentemente durante una décima de segundo a los fotones individuales). Me parece que hay expectativas razonables de poner a prueba, en un futuro no demasiado lejano, no solo mi propio esquema RO**,** sino también otras diversas propuestas para la reducción del estado cuántico que han sido presentadas en la literatura.

El resultado de este experimento podría tener importantes consecuencias para los fundamentos de la mecánica cuántica. Podría muy bien tener una profunda influencia en el uso de la mecánica cuántica en muchas áreas de la ciencia, tales como en biología, donde no tiene por qué haber una división precisa entre *sistema cuántico y observador*. Más concretamente, las sugerencias que Stuart Hameroff y yo hemos presentado concernientes a los procesos físicos y biológicos que tienen lugar en el cerebro para acomodar el fenómeno de la conciencia dependen crucialmente de la existencia y la escala de los efectos para cuya comprobación están diseñados estos experimentos. Un resultado negativo concluyente en dichos experimentos descartaría nuestra propuesta.

Contratapa 1:

Las originales y provocativas ideas de Roger Penrose acerca de la física del Universo en todas sus dimensiones (desde la física cuántica hasta la física de la mente) han sido temas de discusión y controversia. En sus dos libros anteriores, *La nueva mente del emperador y Las sombras de la mente,* estas ideas se exponían de manera clara y rigurosa. En este, Roger Penrose resume y actualiza su pensamiento acerca de estos asuntos, y presenta un compendio fundamental de los problemas aún pendientes. Además, introduce conceptos nuevos que pueden ser de gran utilidad para comprender el funcionamiento del cerebro y la naturaleza de la mente humana. Estas ideas son debatidas a continuación por tres distinguidos expertos con diferentes formaciones —Abner Shimony y Nancy Cartwright como Filósofos de la Ciencia, y Stephen Hawking como Físico Teórico y Cosmólogo—. Finalmente, Roger Penrose responde a las elaboradas críticas.

Este volumen proporciona una introducción accesible, esclarecedora y estimulante a la visión de Roger Penrose sobre la física teórica del siglo XXI. Su entusiasmo, intuición y buen humor resplandecen en esta brillante exposición de los problemas de la física moderna.

Contratapa 2:

Roger Penrose, *Rouse Ball Professor* de Matemáticas en el *Mathematical Institute* de la Universidad de Oxford, es uno de los más importantes investigadores en muchas áreas de las matemáticas y de la física matemática y teórica. Es muy conocido por sus trabajos sobre agujeros negros, gravedad cuántica, teselación apariódica y, más recientemente, la ciencia de la mente. Además de una larga lista de publicaciones técnicas, ha publicado monografías sobre investigación avanzada como *Techniques of Differential Topology in Relativity* y *Spinors and Space-Time,* así como dos importantes libros de divulgación, *La nueva mente del emperador* y Las *sombras de la mente.* En 1994 recibió el título de *Sir.*

Contraportada:

Este volumen proporciona una introducción accesible, esclarecedora y estimulante a la visión de Roger Penrose sobre la futura física teórica del siglo XXI. Se resumen y actualizan las originales y controvertidas ideas expuestas en sus populares libros *La nueva mente del emperador* y *Las sombras de la mente,* e incluye incisivas contribuciones de Abner Shimony, Nancy Cartwrighty Stephen Hawking. En dos apéndices finales el autor explica de forma clara y sencilla por un lado el Teorema de Goodstein y el pensamiento matemático, y por otro nos muestra experimentos para poner a prueba la reducción de estado inducida gravitatoriamente. Un gran libro escrito por una de las mentes más originales y provocativas de la ciencia.