



Reseña

No cabe duda de que la ciencia desempeña un papel decisivo en nuestra cultura. Sin embargo, a menudo, las más iluminadoras teorías científicas tienen su fundamento en ecuaciones que, para muchos de nosotros, han sido siempre un coto vedado. Ya su aspecto se nos antoja un obstáculo insuperable a la hora de entenderlas, y han llegado a simbolizar el misterio y el miedo que inspira la ciencia moderna. *Fórmulas elegantes* intenta superar esta fractura presentando algunas de las grandes ecuaciones de la ciencia moderna al lector no especializado.

Para ello, Graham Farmelo ha reunido a un extraordinario equipo de científicos y divulgadores que han puesto todo su entusiasmo y habilidad en la tarea que les encomendó: desmenuzar y analizar, cada uno, una ecuación, explicando no sólo el significado de los términos y el alcance de la realidad que enuncian, sino también las circunstancias en que se concibieron. Así, *Fórmulas elegantes* consigue enseñar deleitando y abrirnos los ojos a la belleza e importancia de esas breves sucesiones de símbolos que resumen verdades eternas.

Índice

Prefacio

1. [Una revolución sin revolucionarios](#)
2. [La ecuación del sextante. \$E = mc^2\$](#)
3. [El redescubrimiento de la gravedad](#)
4. [Erótica, estética y la ecuación de onda de Schrödinger](#)
5. [Un poco de magia](#)
6. [Comprender la información, bit a bit](#)
7. [Simetrías ocultas](#)
8. [Un espejo en el cielo](#)
9. [Las ecuaciones de la vida](#)
10. [El mejor tiempo posible para vivir](#)
11. [Un cuento de hadas medioambiental](#)

Epílogo

Prefacio

Fórmulas elegantes

La ciencia es para los que aprenden; la poesía, para los que saben.

Joseph Roux, Meditaciones de un párroco, parte 1, n.º 71 (1886)

Durante una entrevista en la radio, concedida en mayo de 1974 para promocionar su colección *High Windows*, Philip Larkin decía que un buen poema es como una cebolla. Por fuera, ambos son agradablemente suaves y misteriosos y se van haciendo aún más suaves y misteriosos a medida que desprendemos sus sucesivas capas. Su ambición era crear la cebolla perfecta.

La poesía de la ciencia está contenida, en cierto modo, en las grandes ecuaciones y, como los ensayos de este libro demuestran, las ecuaciones también pueden ser *peladas*. Pero sus capas representan sus atributos y consecuencias, y no sus significados.

A pesar de los intentos de poetas y críticos literarios, nadie ha dado con una definición de poema que esté libre de controversia. No es el caso de los matemáticos cuando deben definir el término «ecuación». Una ecuación es, básicamente, la expresión de un equilibrio perfecto. Para los matemáticos puros —desconectados, normalmente, de la ciencia— una ecuación es una declaración abstracta, sin relación alguna con hechos concretos del mundo real. De modo que cuando esos matemáticos se enfrentan a una ecuación

del tipo de $y = x + 1$, ven la x y la y como si fueran símbolos totalmente abstractos y no representaciones de cosas que existen en la realidad.

Sería posible imaginar un universo en el que las ecuaciones matemáticas no tuvieran nada que ver con los fenómenos de la naturaleza. Lo curioso es que no es así. Los científicos plasman sistemáticamente sus leyes mediante ecuaciones en las que los símbolos representan magnitudes que los experimentadores pueden medir. Es precisamente esta representación simbólica lo que ha hecho de las ecuaciones matemáticas una de las armas más potentes del arsenal científico.

La más conocida de las ecuaciones científicas es $E = mc^2$, enunciada por primera vez por Einstein en 1905. Como muchas de las grandes ecuaciones, establece una igualdad entre cosas que, *a priori*, parecen ser por completo diferentes¹ —energía, masa y la velocidad de la luz en el vacío—. Mediante esta ecuación, Einstein predecía que, dada una masa (m), si la multiplicamos dos veces por la velocidad de la luz en el vacío (representada por la letra c), el resultado es exactamente igual a su energía correspondiente (E). Como cualquier otra ecuación, $E = mc^2$ equilibra dos magnitudes como si se tratara de los brazos de una balanza, con el signo = como punto de apoyo. Pero así como una balanza equilibra pesos, la

¹ Citemos aquí una oportuna analogía entre la ecuación científica y el poema. En la introducción a su versión de *Beowulf* Seamus Heaney comenta que este gran poema «responde perfectamente a la moderna concepción de trabajo de imaginación creativa en el sentido de acomodar realidades en conflicto en el seno de un orden nuevo» (Faber & Faber, 1999, pág. 17). De una gran ecuación científica se puede decir algo parecido, ya que acomoda en un orden nuevo las aparentemente distintas magnitudes que en ella intervienen.

mayoría de las ecuaciones equilibran otras magnitudes; $E = mc^2$, por ejemplo, equilibra energías. Nuestra celeberrima ecuación comenzó su andadura como una mera especulación einsteniana y sólo décadas más tarde pasó a formar parte del acervo científico, una vez los experimentadores constataron que estaba de acuerdo con las observaciones. Convertida en un tótem del siglo XX, $E = mc^2$ es hoy una de las pocas cosas sobre ciencia que cualquier participante en un concurso de televisión se supone que conoce.²

Como todas las grandes ecuaciones científicas, $E = mc^2$ es, en muchos aspectos, similar a un poema. Al igual que un buen soneto se echa a perder si cambiamos simplemente una palabra o un signo de puntuación, no cabe alterar lo más mínimo una ecuación como la citada sin convertirla en algo inútil. $E = 3mc^2$, por ejemplo, no tiene relación alguna con el comportamiento de la naturaleza.

Las grandes ecuaciones comparten también con la poesía cierta cualidad especial: la poesía es la forma del lenguaje más concisa y cargada de significado, del mismo modo que las grandes ecuaciones científicas son la forma más sucinta de expresar el aspecto de la realidad física que describen. $E = mc^2$ es enormemente potente: sus escasos símbolos concentran un conocimiento aplicable a toda conversión de energía, desde las que tienen lugar en las células de todos los seres vivos hasta las que se producen en una explosión

² Roland Barthes, «The brain of Einstein», *Mythologies*, Vintage, 1993, págs. 68-70. En este ensayo el autor realiza un breve y perspicaz análisis de la simbología einsteniana (incluyendo algunos comentarios sobre $E = mc^2$).

cósmica lejana. Y lo que es más: al parecer, la ecuación lleva siendo válida desde el origen de los tiempos.

Del mismo modo que el estudio detallado de una gran ecuación lleva a los científicos a descubrir progresivamente cosas que no percibieron al principio, la repetida lectura de un buen poema desencadena invariablemente nuevas emociones y asociaciones. Las grandes ecuaciones suponen, para una mente dispuesta, un estímulo tan rico como la poesía. Y al igual que Shakespeare nunca pudo prever los múltiples significados que los lectores darían a «¿Debería compararte con un día de verano?» (soneto 18), Einstein no imaginó las miles de consecuencias que se derivarían de sus ecuaciones de la relatividad.

Sin embargo, existen grandes diferencias entre las ecuaciones científicas y la poesía. Un poema está escrito en un idioma concreto y pierde buena parte de su magia al ser traducido, mientras que una ecuación se expresa en el lenguaje universal de las matemáticas: $E = mc^2$ es lo mismo en inglés que en swahili. Por otra parte, los poetas buscan significados e interacciones múltiples entre palabras, en tanto que los científicos tratan de que sus ecuaciones transmitan un significado lógico único.³

El significado de una gran ecuación científica nos suele proporcionar lo que se denomina una ley de la naturaleza. Una

³ Para más información, véase el ensayo del poeta e inmunólogo checo Miroslav Holub, «Poetry and science», *The Dimension of the Present Moment*, Londres, Faber & Faber, 1990, págs. 132-133. Merece también la pena leer las reflexiones personales de la poetisa Lavinia Greenlaw, «Unstable regions: poetry and science», en: Francis Spufford y Jenny Uglow, eds., *Cultural Babbage*, Londres, Faber & Faber, 1996.

famosa analogía debida al físico Richard Feynman sirve para clarificar esta relación entre ecuaciones y leyes.⁴ Imaginemos a alguien que presencia un juego que se desarrolla sobre un tablero de ajedrez. Si nadie le ha enseñado antes las reglas, podría tratar de deducirlas a partir de los movimientos de piezas que observa. Imaginemos ahora que los jugadores no están jugando su partida en un tablero de ajedrez normal, sino que mueven las piezas siguiendo un conjunto de reglas mucho más complicado y sobre un tablero con un número de casillas enorme. Para deducir las reglas del juego, el observador tendría que examinar partes de él de manera extraordinariamente cuidadosa, buscando pautas y reuniendo pistas repetitivas. Ése es, en esencia, el reto de los científicos. Los científicos observan de cerca la naturaleza —los movimientos de las piezas sobre el tablero—, tratando de descubrir sus leyes ocultas. Decenas de pensadores se han rendido ante el enigma de por qué la mayoría de las leyes fundamentales de la naturaleza pueden ser expresadas mediante una simple ecuación. ¿Por qué cabe expresar tantas leyes como un imperativo absoluto, el que dos magnitudes aparentemente inconexas (las partes izquierda y derecha de la ecuación) sean exactamente iguales? En realidad, tampoco está claro por qué existen las leyes fundamentales.⁵ Según cierta afirmación popular, se debe a que Dios es matemático, una idea inútil que contesta a una cuestión profunda con una proposición

⁴ Richard Feynman, *The Character of Physical Law*, Londres, Penguin, 1992, págs. 35-36. [Trad, esp.: *El carácter de la física*, Barcelona, Tusquets Editores, col. Metatemas 65, 2000].

⁵ Eugene Wigner, «The unreasonable effectiveness of mathematics», *Communications on Pure and Applied Mathematics*, 1960, 13, págs. 1-14.

imposible de verificar. Aun así, el designio divino ha sido, hasta no hace mucho, una explicación común para la eficacia de las ecuaciones científicas. Basta con ver la inscripción en el monumento a María Mitchell (1818-1889), la primera astrónoma profesional estadounidense, ubicado en el Bronx Hall of Fame, según la cual «Cada fórmula que expresa una ley de la naturaleza es un himno de alabanza a Dios».

Aún más polémica que la procedencia de las ecuaciones científicas es la cuestión de si éstas son inventadas o descubiertas.⁶ El astrofísico indoamericano Subrahmanyan Chandrasekhar hablaba por boca de los más grandes teóricos cuando afirmaba que, cada vez que descubría un hecho nuevo o era presa de una nueva intuición, parecía ser algo «que siempre había estado allí y que él, simplemente, había tenido la suerte de dar con ello». Bajo esta óptica, las ecuaciones que subyacen en los fenómenos del universo en cierto sentido «están ahí», ajenas a la humana existencia, de modo que los científicos no serían sino arqueólogos cósmicos que tratan de desenterrar unas leyes que han permanecido escondidas desde el principio de los tiempos. El origen de las leyes sigue siendo un completo misterio.

De los cientos de miles de investigadores que han poblado el mundo, muy pocos pueden presumir de que una ecuación importante lleve su nombre. Dos científicos particularmente expertos en el descubrimiento de ecuaciones fundamentales y

⁶ Esta espinosa cuestión es analizada de forma clara en *The Universe that Discovered Itself* de John Barrow (Oxford University Press, 2000), capítulo 5.

notablemente conscientes del papel de las matemáticas en la ciencia fueron Albert Einstein y el brillante físico teórico británico Paul Dirac. Sin que las matemáticas fueran su especialidad, ambos destacaron por su habilidad para crear ecuaciones tan fecundas como los más grandes poemas. Y los dos también estaban convencidos de que las ecuaciones fundamentales de la ciencia tenían que ser bellas.⁷

La idea puede resultar extraña. El concepto subjetivo de belleza es mal acogido en los círculos intelectuales y, desde luego, no tiene cabida en las críticas académicas de arte.⁸ Sin embargo, la palabra acude automáticamente a nuestros labios —incluso a los de los críticos más pedantes— cuando contemplamos la sonrisa de un niño, la imponente estampa de una montaña o las exquisitas formas de una orquídea. ¿Qué queremos decir al afirmar que una ecuación es bella?⁹ Básicamente, que esa ecuación puede evocar en nosotros sensaciones similares a las que otras cosas que la mayoría de nosotros describimos como bellas producen. De manera similar a una gran obra de arte, una ecuación bella cuenta entre sus atributos mucho más que con el simple atractivo —poseerá

⁷ Varios grandes físicos teóricos han escrito sobre la importancia de la belleza en sus respectivas materias. Véase: Steven Weinberg, *Dreams of a Final Theory*, Random House, 1993, capítulo 6; Subrahmanyan Chandrasekhar, *Truth and Beauty: Aesthetics and Motivations in Science*, University of Chicago Press, 1987, y Chen Ning Yang, «Beauty in theoretical physics», en: Deane W. Curtin, ed., *The Aesthetic Dimension of Science*, Nueva York, Philosophical Library, 1980.

⁸ Y, sin embargo, «la belleza ha vuelto», según afirma el crítico arquitectónico posmodernista Charles Jencks en su extenso análisis «What is beauty?», *Prospect*, 2001, agosto-septiembre, págs. 22-27.

⁹ Un completo análisis del concepto de belleza en la ciencia nos lo proporciona James W. McAllister en *Beauty and Revolution in Science*, Cornell University Press, 1996.

universalidad, simplicidad, inevitabilidad y una especie de fuerza elemental—. Pensemos en obras maestras como *Manzanas y peras*, de Paul Cézanne, la cúpula geodésica de Buckminster Fuller, la interpretación de Lady Macbeth realizada por Judi Dench o la versión de Ella Fitzgerald de *Manhattan*. En mi primera contemplación de cada una de ellas sentí que estaba ante algo monumental en su concepción, fundamentalmente puro, libre de todo elemento inútil y ejecutado tan exquisitamente que su fuerza disminuiría si intentásemos cambiar cualquiera de sus atributos.

Una cualidad adicional de una gran ecuación científica es que posee una belleza útil. Ha de ajustarse a los resultados de todo experimento relevante y, además, predecir resultados de experimentos que nadie haya realizado aún. Este aspecto de la efectividad de una ecuación es semejante a la belleza de una máquina de precisión como la que imaginamos cuando en el filme de Stanley Kubrick *La chaqueta metálica* el recluta Gomer Pyle se pone a hablar de su rifle. El embrutecido *marine* alaba su meticulosa fabricación, deleitándose en las cualidades que lo hacen ideal para su letal propósito. No sería tan bello si no funcionase.

El concepto de belleza era especialmente importante para Einstein, el científico más preocupado por la estética de todo el siglo XX. Según su hijo mayor Hans, «Su carácter se parecía más al de un artista que al de un científico al uso. Por ejemplo, su mayor aspiración para una buena teoría no era que resultara correcta o exacta, sino que fuera bella». En cierta ocasión, llegó a afirmar que

«las *únicas* teorías físicas que estamos dispuestos a aceptar son las que resultan bellas», dando por supuesto que una buena teoría, a la postre, tenía que concordar con los experimentos. Dirac iba incluso más allá que Einstein en su convencimiento de que la belleza matemática era un criterio para establecer la calidad de una teoría fundamental,¹⁰ declarando que la cuestión era para él «una especie de religión». En los últimos años de su vida dedicó mucho tiempo a viajar alrededor del planeta y a dar conferencias sobre el origen de la gran ecuación que lleva su nombre, haciendo hincapié en que la búsqueda de la belleza había sido siempre su norte y una continua fuente de inspiración. Requerido para que expresara en pocas palabras su filosofía de la física durante un seminario dado en Moscú en 1955, escribió en letras mayúsculas sobre la pizarra: «Las leyes físicas han de ser matemáticamente bellas».

Para el resto de los mortales, ese esteticismo es un arduo e improductivo credo. El hecho es que, para la mayoría de los científicos, la belleza no es un concepto que les preocupe demasiado ni que les sirva de guía en su trabajo diario. Es cierto que las ecuaciones que usan poseen una belleza subyacente y las soluciones correctas de esas ecuaciones son más bellas que feas. Pero la belleza puede resultar engañosa. La ciencia está salpicada de restos de teorías que una vez parecieron bellas, pero que se demostró que estaban equivocadas —la naturaleza opinaba de otra manera—. A la hora de validar una nueva teoría, el criterio

¹⁰ Dirac enunció su credo estético en «Pretty mathematics», *International Journal of Theoretical Physics*, 1982, 21 (8-9), págs. 603-605.

fundamental para la mayor parte de los científicos es comprobar que se ajusta a los experimentos.

La idea de que la ciencia avanza por medio de una combinación de experimentos y teoría basada en las matemáticas es relativamente moderna. Tuvo su origen en Florencia hace tan sólo trescientos cincuenta años —ayer, respecto a la historia de la raza humana—. Su progenitor fue Galileo, el primer científico moderno, quien observó que la ciencia avanza mejor cuando trabaja sobre una estrecha franja de fenómenos y asume que los resultados serán leyes que cabe describir mediante términos matemáticos precisos.¹¹ El suyo fue uno de los más grandes y productivos descubrimientos de toda la historia de las ideas.

La ciencia se ha ido haciendo cada vez más matemática desde los tiempos de Galileo. Las ecuaciones son, actualmente, una herramienta científica de primer orden y es casi un artículo de fe para la mayoría de los teóricos —y, desde luego, para la mayor parte de los físicos— que existe una ecuación fundamental que describe el fenómeno que están estudiando o que, algún día, alguien hallará la ecuación idónea. En cualquier caso, y tal como le gustaba decir a Feynman, podría resultar al final que las leyes fundamentales de la

¹¹ En su *Opere Il Saggiatore* (1656), Galileo escribió: «[El universo] no podrá ser leído hasta que hayamos aprendido su idioma y nos hayamos familiarizado con los caracteres en los que está escrito: las matemáticas». Como en otras ocasiones, Platón ya se había anticipado; había dicho que «el mundo es la epístola que Dios le ha escrito a la humanidad» y que aquella «estaba escrita en caracteres matemáticos». Es interesante observar que Galileo y los otros fundadores de la física moderna no escribieron ecuaciones, sino proporciones. Éstas eran, no obstante, equivalentes en muchos aspectos a las ecuaciones. Transcurrirían aún algunas décadas antes de que las ecuaciones se convirtieran en la forma preferida para una expresión matemática. Véase: I. Bernard Cohen, *Revolution in Science*, Harvard University Press, 1985, págs. 139-140.

naturaleza no precisen ser expresadas mediante las matemáticas, sino que se ajusten mejor a otros lenguajes, tales como las reglas que gobiernan una partida de ajedrez.

De momento parece que las ecuaciones son la manera más eficaz de expresar la mayoría de las leyes científicas fundamentales. Pero las ecuaciones no preocupan por igual a todos los científicos, muchos de los cuales se las arreglan con las herramientas matemáticas más rudimentarias. A este respecto, es muy ilustrativa una historieta en la que un matemático, un físico, un ingeniero y un biólogo son preguntados por el valor numérico de π . El matemático contesta secamente que «es igual a la circunferencia de un círculo dividida por su diámetro». El físico, en cambio, afirma que π vale «3,141593 más menos 0,000001». El ingeniero dice que vale «alrededor de 3» y, finalmente, el biólogo pregunta: «¿Qué es π ?».

Por supuesto, se trata de un chiste. Algunos físicos tienen unos conocimientos matemáticos escasos, algunos ingenieros aplican brillantemente las matemáticas a la tecnología y algunos biólogos teóricos son matemáticos de primera. Pero, como toda caricatura, tiene una parte de verdad. Los ingenieros suelen tener un enfoque utilitario de las matemáticas y le dan mucho valor a hacer buenas aproximaciones. Y, de entre todas las ciencias, la física es la más matemática y la biología, la menos. Desde los tiempos de Galileo, los físicos han tratado siempre de simplificar las cosas, subdividiendo las complejidades del mundo real en sus componentes más simples. Este reduccionismo no siempre es aplicable para los biólogos, cuyo

objeto de estudio es el mundo vivo tan inmensamente complejo, con sus comunidades de organismos interrelacionados, cada uno de los cuales presenta una estructura enormemente compleja en términos moleculares. Y no olvidemos que la teoría biológica unificadora por excelencia es, en la superficie al menos, no matemática: *El origen de las especies*, el tratado en el que Darwin describe su teoría de la evolución mediante la selección natural, no contiene una sola ecuación. Lo mismo sucede con la teoría geológica de la deriva continental, en cuyos primeros trabajos (publicados poco después de la primera guerra mundial) no hallamos ecuación alguna.

Los ensayos de la presente colección reflejan la importancia de las matemáticas en las diferentes —aunque, a veces, solapadas— áreas de la ciencia a partir de 1900. La física está especialmente bien representada. Se analizan tres importantes aportaciones einstenianas (incluyendo $E = mc^2$ y la ecuación de la relatividad general) y otras grandes ecuaciones que han transformado nuestra visión del mundo subatómico. La ecuación de Dirac ocupa un lugar de honor: no sólo cumplió su misión de describir el comportamiento del electrón, sino que, inesperadamente, predijo la existencia misma de la antimateria: nada más y nada menos que la otra mitad del universo. No es extraño que el propio Dirac comentara: «Mi ecuación es más inteligente que yo».

Las ecuaciones de la física subatómica constituyen la base de lo que se denomina el «modelo estándar», un nombre demasiado prosaico para la teoría actual de las partículas fundamentales y sus

interacciones (la cual, ironías del destino, deja fuera la fuerza más conocida de todas, la gravedad). En el Epílogo se contemplan juntas las ramas que han dado lugar al modelo, uno de los hitos intelectuales del siglo XX.

Dos de los ensayos echan un vistazo a sendas ecuaciones de la biología moderna. El primero explica el modo en el que las ideas evolutivas pueden ser expresadas matemáticamente, dando lugar a una perspectiva rica y diversa del mundo vivo, desde el comportamiento nupcial del ciervo rojo hasta la proporción entre machos y hembras en los avisperos. El segundo ensayo se refiere al denominado mapa cuadrático, una ecuación engañosamente simple de ecología teórica que puede ser utilizada para explicar las variaciones en la población de peces en el estanque de un jardín, la fluctuación del número de perdices en un coto de caza y una multitud de problemas similares. Esta ecuación desempeñó un papel crucial en la historia de la teoría del caos, pues resultó encarnar de forma notable el comportamiento caótico —es extremadamente sensible a las condiciones iniciales—. En buena parte, gracias a esta ecuación —una ecuación tan simple que los niños la pueden estudiar en la escuela— en la década de 1970 los científicos llegaron a la conclusión de que algunas ecuaciones que parecen predecir el futuro a partir de sucesos pasados son por completo inservibles para hacer tales predicciones, contrariamente a lo que la ciencia había creído hasta entonces.

Otras dos ecuaciones incluidas en este volumen se refieren a las ciencias de la información y a la búsqueda de inteligencia extraterrestre. El primero de los ensayos contempla las ecuaciones debidas al decano de los teóricos de la información, Claude Shannon, quien fue pionero en crear el aparato matemático que soporta lo que hoy conocemos como la revolución de las comunicaciones. Las ecuaciones de Shannon son aplicables a cualquier tipo de transferencia de información, incluyendo Internet, la radio y la televisión.

La búsqueda de inteligencia extraterrestre (*Search for Extraterrestrial Intelligence*, SETI) no parece un tema que pueda dar lugar a una ecuación. ¿Cómo va a haber una ecuación para algo que quizá no exista? La respuesta es que la ecuación fundamental de SETI —creada por el astrónomo norteamericano Frank Drake— no efectúa predicciones; en lugar de ello, estructura nuestro modo de pensar sobre la probabilidad de que existan civilizaciones que se puedan comunicar con nosotros. Aunque no sea bella en el sentido de las de Dirac o Einstein, la fórmula de Drake ha aportado un poco de coherencia a un campo en el que la confusión abunda.

Los científicos no emplean sólo ecuaciones de tipo matemático. Los químicos, por ejemplo, usan ecuaciones que no están constituidas sólo por símbolos matemáticos, sino que incluyen letras que representan átomos, moléculas y sus parientes subatómicos. Muchas actividades industriales se basan en ecuaciones de esta clase, ecuaciones que describen interacciones cuyos detalles

podemos inferir, pero que difícilmente podríamos observar a simple vista. Hemos recogido un conjunto especial de reacciones químicas en este libro, en representación de esta rama de la ciencia. Esas ecuaciones maravillosamente sencillas constituyeron la base para comprender las causas de la reducción de la capa de ozono, la presencia de compuestos químicos denominados CFC (clorofluorocarburos) en la atmósfera terrestre. A comienzos de la década de 1980, esas simples ecuaciones alertaron a la humanidad sobre el riesgo de una catástrofe ambiental.

Los autores de este libro son prestigiosos científicos, historiadores y escritores. Han analizado todas y cada una de las facetas de su ecuación —las capas de la cebolla de Larkin— y han puesto de relieve las más fascinantes, evitando en lo posible entrar en excesivo detalle matemático. El resultado es un conjunto único de reflexiones personales sobre algunas de las ecuaciones básicas de la ciencia moderna, unas ecuaciones que debido a su concisión, potencia y simplicidad fundamental pueden ser contempladas como auténtica poesía del siglo XX.

En mi colección de poesía, en la estantería que hay justo encima de mi mesa, se halla un ejemplar de *High Windows*. Lo leí por primera vez cuando era un bisoño estudiante de física subatómica y trataba a duras penas de entender sus ecuaciones fundamentales y de apreciar su belleza. La colección me la regaló un amigo, un estudiante de literatura inglesa seguidor de las ideas de Larkin, pocos días después de que fuera publicada. Su mensaje fue el

mismo que quisiera transmitir ahora al lector: «Que le aprovechen las cebollas».

Graham Farmelo

Agosto de 2001

Parte 1

Una revolución sin revolucionarios

La ecuación de Planck-Einstein para la energía de un cuanto.

Graham Farmelo

§. I

*Festejamos las revoluciones
cuando ya no son peligrosas.*

Pierre Boulez,

*13 de enero de 1989, durante la
conmemoración del bicentenario de
la Revolución francesa*

El siglo XX eligió entre sus celebridades a algunos personajes verdaderamente indignos, pero tuvo un exquisito gusto a la hora de escoger a su científico favorito. Albert Einstein, tan brillante en la identificación de problemas científicos a la vez que los resolvía, hizo más que ningún otro por el avance del conocimiento humano en el siglo científicamente más productivo de la historia. Lástima que su obra más revolucionaria permanezca hoy en día ampliamente ignorada.

Si se le pregunta al hombre de la calle cuál fue la más famosa contribución de Einstein a la ciencia, con toda certeza responderá que la teoría de la relatividad. Un trabajo brillante, sin duda, pero no revolucionario, como el propio Einstein solía subrayar. Partiendo

de la obra de Newton y Galileo, Einstein produjo una nueva teoría del espacio, el tiempo y la materia que engranaba suavemente con las teorías de aquéllos. Sólo en un punto se había apartado radicalmente de sus predecesores: al proponer la extraordinaria idea de la energía de la luz.

El sentido común nos dice que la luz entra en nuestros ojos mediante un flujo continuo. A finales del siglo XIX, los científicos parecían confirmar esta idea intuitiva mediante su universalmente aceptado modelo de onda, según el cual la luz es transmitida suavemente, como la energía de las olas chapoteando en el muelle. Pero, como Einstein decía, «el sentido común es el conjunto de todos los prejuicios adquiridos antes de los dieciocho años». El gran físico afirmó en 1905, cuando trabajaba como examinador de patentes en Berna, que la idea era incorrecta y que la energía de la luz no se propagaba de forma continua, sino en cantidades discretas que denominó «cuantos». Poco después, se aventuró a especular que las energías de los átomos de un sólido estaban también cuantizadas — sólo eran posibles ciertos valores de energía—. Esta cuantización de la energía iba en contra del sentido común. La energía del movimiento de la manzana que había caído en el jardín de Newton parecía incrementarse gradualmente y no a saltos.

Einstein vio con mayor claridad que nadie que el mundo submicroscópico está lleno de cuantos: la naturaleza es básicamente granular, en vez de continua. Aunque en la época en la que llegó a estas conclusiones trabajaba solo, Einstein no lo

consiguió por inspiración divina. Se basó en los artículos escritos por un físico veintiún años mayor que él, Max Planck, que por aquel entonces trabajaba en Berlín y era el decano de los físicos germanos. Planck había sido el primero en hablar de cuantos de energía a finales de 1900, aunque no está claro si era consciente de las implicaciones de la idea.

Cierta ecuación engañosamente simple inquietó de manera especial a los pioneros de la física cuántica. Escrita en primer lugar por Planck, pero interpretada en todo su alcance sólo años después por Einstein, la ecuación relaciona la energía E de un cuanto con su frecuencia, f : $E = hf$, donde h es una cantidad constante que acabaría recibiendo el nombre de su descubridor. La expresión se convirtió en la primera ecuación científica importante del nuevo siglo (el káiser Guillermo II había decretado que 1900 era el primer año del siglo XX y no el último del XIX). Los alumnos de instituto la aprenden hoy de memoria y sin quebradero de cabeza alguno, pero a los primeros físicos cuánticos les costó casi veinticinco años descifrar su significado. Durante ese tiempo, los trabajos de Einstein en torno a las ideas subyacentes en $E = hf$ le condujeron a ser la primera persona en predecir la existencia de una partícula fundamental y, en unión de otros científicos, a poner las bases de una teoría cuántica totalmente articulada, la que posiblemente sea la idea científica más revolucionaria del siglo.

Albert Einstein y Max Planck dominan la historia de esta ecuación tan enormemente fecunda desde el punto de vista intelectual. A

primera vista, ambos personajes eran muy distintos. Planck era alto, delgado y calvo, mientras que Einstein era fornido, de estatura ligeramente superior a la media y dotado de una resplandeciente melena. Planck era sociable con sus colegas, en tanto que Einstein mantenía cierta distancia intelectual con ellos. Planck era nacionalista; Einstein, declaradamente liberal y cosmopolita. Planck tenía inclinaciones políticas de derechas; Einstein, de izquierdas. Planck era un puntilloso administrador; Einstein huía en cuanto podía de los papeles. Planck era un hombre hogareño; la vida familiar de Einstein fue casi siempre anómala.

Pero también tenían muchas cosas en común. Los dos eran físicos teóricos, una raza de científicos relativamente nueva con un marcado interés por explicar la naturaleza en términos de principios universales. Adictos ambos al trabajo, valoraban cualquier nuevo resultado experimental, pero eran más felices trabajando en el laboratorio mental que alojaban sus cerebros. Los dos pensaban que los principios científicos existen con independencia del ser humano y que hay muchos otros principios que aguardan ser descubiertos. Como todos los buenos científicos, Planck y Einstein abordaban sus trabajos de una forma muy prudente. Eran muy cautelosos con los nuevos resultados experimentales, remisos a innovaciones que contradijeran las teorías establecidas y conscientes de que, para que una nueva teoría pueda ser tomada en serio, debe reproducir los éxitos de sus predecesoras e, idealmente, hacer ella misma nuevas predicciones.

Para ambos, la física era su primer amor y la música, el segundo. Einstein era ferviente admirador de Bach, Mozart y Haydn y le gustaba tocar el violín, que llevaba con él a todas partes. Las opiniones sobre la calidad de su técnica varían: según el gran pedagogo del instrumento Shinichi Suzuki, su sonido era «de una hermosa delicadeza», pero algún otro experto afirmó que Einstein «movía el arco como un leñador». Cualquiera que sea la verdad sobre sus dotes musicales, Einstein no era benévolo con los críticos: «Disfrutaba más con las disputas musicales que con las científicas», según afirmaba uno de sus conocidos. Planck era un músico mucho más refinado y afable, un pianista suficientemente bueno como para formar dúo con el genial violinista Joseph Joachim en sus últimos años. A Planck le gustaba la música de un buen amigo y colaborador de Joachim, Johannes Brahms, y también las de Bach y Schubert.

Al poner los cimientos de la teoría cuántica, Planck, Einstein y sus colegas se hallaban en la vanguardia del movimiento más amplio del modernismo, pues reinventaban conscientemente su propia materia de estudio y exploraban los métodos y límites de las técnicas clásicas.¹² En este sentido, eran como Igor Stravinski en San Petersburgo, Virginia Woolf en Londres, Pablo Picasso en París o Antonio Gaudí en Barcelona. Pero, a diferencia de los artistas, Planck y Einstein eran modernistas a pesar de ellos mismos: no

¹² Para un excelente análisis del modernismo en un amplio contexto cultural, véase: Thomas Vargish y Delo Mook, *Inside Modernism: Relativity Theory, Cubism, Narrative*, Yale University Press, 1999.

trataban de remover los cimientos de la ciencia como fin último. Mientras los artistas eran libres de crear nuevas formas para reemplazar las que parecían anticuadas, los científicos no tenían más remedio que crear nuevas teorías para sustituir las que habían demostrado ser irremediablemente defectuosas. Fue una minúscula pero inquietante disparidad entre la teoría y los experimentos lo que condujo a la revolución cuántica. La cuestión empezó a cocerse en algunos fogones de Berlín.

§. II

Berlín nunca ha sido un destino preferente para los finos *gourmets*. Éstos admiten, no obstante, que en la actualidad al menos es posible tomar un café exprés decente en alguna de las cafeterías Einstein® esparcidas por la ciudad. Aunque el nombre de esos pulcros establecimientos no se deba al gran científico, los rótulos de sus fachadas nos recuerdan el tiempo, hace un siglo, en el que Berlín no sólo era la ciudad europea más rica y de crecimiento más rápido,¹³ sino la capital mundial de la física.

Poco antes de que concluyera la guerra franco-prusiana en 1871, Bismarck convirtió Berlín en la capital del victorioso nuevo Reich. Algunos de los más grandes experimentadores del mundo vivían en esta ciudad, un suculento guiso cultural desde los días de gloria del erudito déspota prusiano Federico el Grande, un siglo atrás. Berlín era también el cuartel general de una élite de físicos teóricos,

¹³ Roland Taylor, *Berlin and its Culture*, Yale University Press, 1997.

miembros de una nueva disciplina que los experimentadores más prestigiosos habían comenzado a emplear a finales de la década de 1860 para enseñar las teorías, abrumadoramente matemáticas, que cada vez estaban más en boga. La comunidad se hallaba, en un principio, formada exclusivamente por hombres, y sólo fue a partir del verano de 1908 que las universidades de Berlín comenzaron a admitir a mujeres.¹⁴ Un siglo después, el panorama había cambiado poco: la inmensa mayoría de los físicos teóricos continuaba estando constituida por hombres.

Miembro destacado de la nueva comunidad científica berlinesa, Planck concibió el concepto de cuanto de energía y escribió la ecuación $E = hf$. Para comprender la obra de Planck, necesitamos examinar dos grandes teorías que cautivaron la imaginación de los físicos en la segunda mitad del siglo XIX. La primera consiste en el tratamiento matemático unificado de la electricidad, el magnetismo y la óptica, establecido en 1864 por James Clerk Maxwell, un brillante y versátil físico escocés. Mediante el conjunto de ecuaciones que lleva su nombre, Maxwell demostró que la luz visible es una onda electromagnética que viaja a través de un omnipresente éter, de manera similar a como las ondas sonoras se propagan en el aire. Como toda onda, la electromagnética posee una longitud de onda y su correspondiente frecuencia. La longitud de onda es simplemente la distancia entre dos picos consecutivos, y la

¹⁴ El lugar de la mujer en la vida universitaria alemana de la época aparece descrito en el capítulo 2 de *Lise Meitner: A Life in Physics*, de Ruth Lewin Sime (University of California Press, 1996).

frecuencia, el número de veces que la onda oscila arriba y abajo cada segundo. En el extremo rojo del espectro, las ondas de luz tienen una longitud de onda de siete diezmilésimas de milímetro y se mueven arriba y abajo 430 billones de veces por segundo, mientras que en el extremo violeta la longitud de onda es más corta y la frecuencia, más alta. La teoría de Maxwell explicaba correctamente por qué existen ondas electromagnéticas fuera del espectro visible, con frecuencias más altas o más bajas. La luz es, simplemente, una parte del espectro de radiación electromagnética.

Otro de los muchos intereses de Maxwell era la termodinámica, la segunda gran teoría física que alcanzó su madurez a finales del siglo XIX. La teoría trataba de las diferentes formas de energía y hasta qué punto pueden ser convertidas unas en otras; por ejemplo, el movimiento de un volante en calor. Se ocupaba sólo de la materia como un todo y no decía nada sobre el comportamiento individual de los átomos que la constituyen. Las máquinas de vapor que impulsaron la revolución industrial en la Europa occidental sirvieron inicialmente para estimular las investigaciones teóricas sobre la termodinámica. A mediados del siglo XIX, los avances teóricos llevaron a mejorar la tecnología, lo que a su vez condujo a perfeccionar los aparatos diseñados para ensayar la teoría.

La termodinámica y el electromagnetismo, junto con los trabajos de Newton sobre las fuerzas, forman parte de lo que hoy conocemos como «física clásica». No es que sus inventores pensaran que estaban trabajando en el marco de una tradición inmemorial —

creían, simplemente, que estaban haciendo física—, sino que la aparición de la teoría cuántica de Planck, Einstein y sus colegas hizo que, *a posteriori*, se les aplicara esa etiqueta.

Uno de los más grandes físicos clásicos alemanes fue Rudolf Clausius, tal vez el primer físico teórico.¹⁵ Hombre polémico, fue el primero en proponer un enfoque matemático para la termodinámica que se concentrara en la búsqueda de unos pocos principios o axiomas fundamentales. Era crucial, según él, que éstos fueran consistentes desde el punto de vista lógico y que llevaran a resultados que concordaran con los experimentos. Esta aproximación de arriba-abajo contrastaba abiertamente con el estilo tradicional, paso a paso, de hacer física matemática, e implicaba escribir ecuaciones para describir un fenómeno antes de comprobar si respondían a los resultados experimentales.

Otros habían establecido ya que la energía ni se crea ni se destruye: la primera ley de la termodinámica. Pero, en 1850, Clausius se hallaba entre los que pergeñaron la que acabaría siendo conocida como segunda ley, la cual más o menos afirma que el calor no fluye de forma espontánea de algo frío a algo caliente. Esto es perfectamente creíble: un cappuccino frío nunca se calentará si lo dejamos solo. Las dos leyes parecen ser absolutas, es decir, universalmente válidas, las comprobemos o no.

¹⁵ Christa Jungnickel y Russell McCormach nos ofrecen una exhaustiva historia de los inicios de la moderna física teórica en *Intellectual Mastery of Nature*, University of Chicago Press, 1986, vols. 1 y 2.

Aunque ambas son en apariencia simples, Clausius tuvo que emplear toda su artillería intelectual para formularlas de manera rigurosa. Su precisión matemática y lingüística, unida a la meridiana claridad de sus razonamientos, fascinaron a Max Planck cuando era un impresionable graduado.¹⁶ Nacido en 1858 en el seno de una patriótica e influyente familia de eruditos, abogados y funcionarios públicos, estaba profundamente imbuido de valores conservadores. Ciertos recuerdos de infancia asociados a la inestabilidad política de la época le acompañarían toda su vida, así como el hecho de haber presenciado con ocho años el desfile victorioso de las tropas prusianas y austríacas, tras vencer a Dinamarca, en su ciudad natal de Kiel. Diligente estudiante universitario, aunque no especialmente brillante, Planck recibió una amplia educación en matemáticas, física, filosofía e historia. Estudió asimismo música, su principal *hobby*, y llegó a componer una opereta para ser interpretada en las veladas musicales que tenían lugar en la residencia de sus profesores.

Indeciso sobre qué materia acometer, eligió finalmente la física, a pesar de las advertencias de su profesor en la Universidad de Múnich, Philip von Jolly. En lo que hoy se conoce como uno de los más grandes fiascos en la historia de los consejeros universitarios, Von Jolly trató de disuadir a un Planck de veintiún años para que emprendiese la carrera de físico porque, según él, tras el

¹⁶ John Heilbron, *The Dilemmas of an Upright Man*, Harvard University Press, 2000. Se trata de un amplio y cuidadoso relato de la vida de Planck, que hace especial incidencia en los aspectos políticos. Esta edición contiene un valioso epílogo que clarifica las conclusiones a las que Heilbron llegaba en la edición original de 1986.

descubrimiento de las dos leyes de la termodinámica, todo lo que quedaba por hacer a los físicos teóricos era «atar los cabos sueltos». Con su proverbial conservadurismo, Planck replicó cándidamente que sólo deseaba profundizar en los fundamentos legados por sus predecesores y que no tenía deseo alguno de hacer nuevos descubrimientos. Como ya veremos, su primer deseo tuvo que lograrlo a costa del segundo.

Planck se enamoró de la termodinámica cuando, trabajando en su tesis doctoral, quedó fascinado por la potencia y generalidad de sus leyes. Le incomodaban, no obstante, dos aspectos cuyo paladín era un gran físico teórico austríaco, el apasionado y depresivo Ludwig Boltzmann. En primer lugar, Planck no estaba convencido de que la materia estuviese constituida por átomos: nadie había observado realmente uno, así que, tal vez, se trataba tan sólo de una abstracción oportuna. Era escéptico también sobre el argumento de Boltzmann de que la segunda ley de la termodinámica resultaba cierta sólo estadísticamente: que era abrumadoramente probable — pero no totalmente seguro— que el calor fluyera desde algo caliente hacia algo más frío. Esta ausencia de certidumbre era contraria a la pasión de Planck por los absolutos, por lo incontestable, por la seguridad.

No obstante, cierto absoluto había captado su atención. Se refería a un problema tan sutil que pocos científicos fuera de Berlín se habían ocupado de él, algo que cualquiera ajeno al mundillo científico habría pasado por alto por considerarlo ridículo.

Imaginemos una cavidad herméticamente cerrada, como un horno eléctrico, pero sin puerta ni ventana alguna. Supongamos ahora que la cavidad se encuentra a una temperatura estable y uniforme. Las paredes de la cavidad generan una radiación electromagnética que continuamente rebota por todo el interior de aquélla, siendo absorbida y re-emitida o reflejada directamente. Los experimentadores observan esta radiación mediante un diminuto agujero practicado en una de las paredes, con lo que una pequeñísima parte de la radiación se escapa. Asimismo, una fracción de la radiación presente en los alrededores de la cavidad entra en ella, aunque de inmediato es absorbida, re-emitida y reflejada por todo el interior de la cavidad, lo cual la lleva a tener las mismas características que la radiación que existía previamente. Dado que toda la radiación procedente del exterior que pasa a través del agujero es «absorbida», el agujero parece negro cuando es observado a temperatura ambiente y la radiación emergente —una muestra de la que hay dentro de la cavidad— suele recibir el nombre de «radiación de un cuerpo negro».¹⁷ La cuestión que fascinaba a los físicos era la siguiente: dada una temperatura de la cavidad, ¿cuál es la intensidad de la radiación para cada color o, más exactamente, para cada longitud de onda? Fue el tratar de responder a esta pregunta lo que condujo a Planck a la ecuación $E = hf$.

¹⁷ Si vamos calentando el interior de la cavidad hasta alcanzar temperaturas lo suficientemente elevadas, la radiación emergente no se verá de color negro, sino rojo oscuro, después naranja y, más tarde, amarillo. Así pues, la denominación «radiación del cuerpo negro» es un tanto equívoca y debería ser sustituida simplemente por «radiación de la cavidad».

Uno de los asesores de Planck ya había probado que cualquiera que resultara ser la ley para la intensidad de la radiación, no dependería ni del tamaño de la cavidad, ni de su forma, ni del material de las paredes. Esta ley sería un ejemplo clásico de lo que Planck denominaba un «absoluto», algo que «mantendría necesariamente su importancia en el marco de cualquier época y cultura, incluso para las ajenas a la Tierra y al Hombre». La radiación de la cavidad no tenía únicamente un interés académico: era trascendental para la industria alemana del alumbrado, una de las muchas florecientes ramas de la economía del país en la época en que las tecnologías eléctrica y química revolucionaban el capitalismo. En su permanente búsqueda de nuevas fuentes de iluminación que produjeran la máxima luz y el mínimo calor posibles, los ingenieros que trataban de diseñar lámparas eléctricas cada vez más eficientes necesitaban saber cuánta radiación emitían sus filamentos. Cuantos más datos tuvieran sobre la radiación de la cavidad, mejor preparados estarían para producir lámparas más eficaces, del tipo de la inventada en 1897 por el norteamericano Thomas A. Edison. Éste era uno de los temas de investigación en el lujosamente equipado Physikalisch-Technische Reichsanstalt (Instituto Imperial de Física y Tecnología) de Charlottenburgo, justo a las afueras de Berlín y a apenas cinco kilómetros de la universidad en la que Planck trabajaba desde 1889. Financiado conjuntamente por el gobierno alemán y el industrial Werner von Siemens, el Reichsanstalt fue creado a raíz de la guerra franco-prusiana (1870-

1871) con el fin de perfeccionar las técnicas de medida y de establecer normas en las que pudieran basarse los científicos e ingenieros. Conscientes de los potenciales beneficios económicos para el nuevo Reich, los fundadores del Reichsanstalt se propusieron crear unos laboratorios de investigación sin parangón en la época, cuyo objetivo fueran las aplicaciones prácticas para la industria alemana.¹⁸ Hasta el diseño clásico de sus edificios de ladrillos amarillos, ubicados en medio de casi cuatro hectáreas de inmaculadas zonas verdes, nos habla de las ambiciones imperiales de la institución.

Los físicos alemanes llevaban treinta años investigando la radiación de la cavidad. Sus conocimientos podían ser resumidos en leyes matemáticas simples que predecían la intensidad de la radiación para cada longitud de onda. Mientras dos equipos de investigadores del Reichsanstalt trabajaban en el tema, Planck intentaba descifrar la más lograda de las leyes de radiación, escrita en marzo de 1896 por su buen amigo Wilhelm Wien, uno de los mejores físicos del Reichsanstalt. Wien era todo un personaje: hijo de un terrateniente de la Prusia del este, había planeado repartir su tiempo entre la física y la agricultura hasta que una desastrosa cosecha obligó a su padre a vender la granja y a hacer lo necesario para que el joven Wien se dedicara de lleno a su carrera científica. Wilhelm Wien, por otra parte, era un chauvinista y un antisemita reaccionario. Pocos días después de finalizar la primera guerra mundial, encabezó un

¹⁸ La historia definitiva del Reichsanstalt está recogida en la obra de David Cahan, *An Institute for an Empire*, Cambridge University Press, 1989.

grupo de voluntarios, veteranos de guerra fundamentalmente, dispuestos a abatir a tiros a comunistas y a otros militantes de izquierdas en las calles de Würzburg y Múnich para prevenir lo que él mismo denominaba «la bolchevización de Alemania».

La ley de Wien para la radiación de la cavidad predecía correctamente la intensidad de la radiación para cada color en un amplio margen de temperaturas. Planck deseaba explicar la ley de su colega por medio de la termodinámica y el electromagnetismo y estaba convencido al principio de que, a tal efecto, no tenía por qué asumir la existencia de átomos o emplear una versión de la segunda ley de la termodinámica que implicara probabilidades en lugar de certidumbres. Sin embargo, a comienzos del verano de 1899 ya había tenido que renunciar a ambas premisas. A regañadientes, concluyó que sólo podía justificar la radiación de la cavidad si aceptaba que los átomos existían y adoptaba la formulación estadística de Boltzmann. Cuando estaba corrigiendo las pruebas de imprenta del artículo que presentaba la teoría, los experimentadores —«las fuerzas de choque de la ciencia», como él las llamaba— le trajeron preocupantes noticias: la ley de Wien parecía estar en dificultades. Subestimaba claramente la intensidad de la radiación, especialmente en longitudes de onda grandes, que eran las que los nuevos aparatos les estaban permitiendo empezar a investigar (*figura 1.1*).

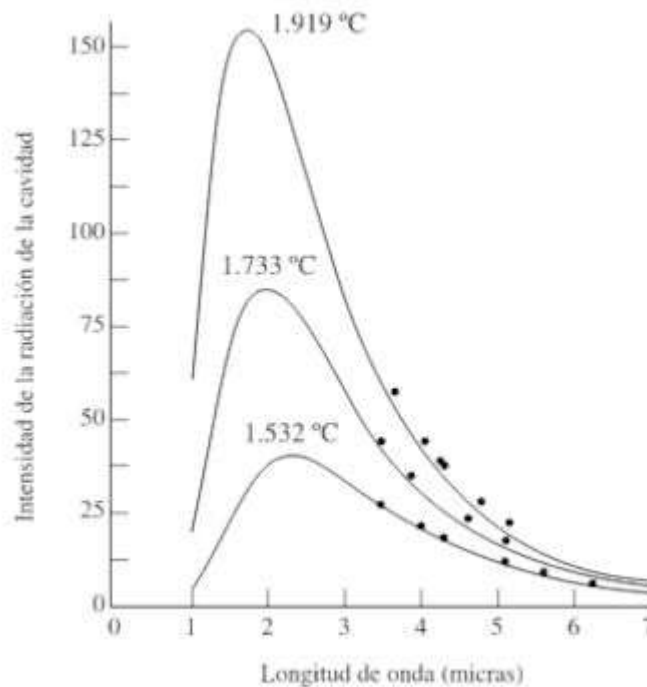


Figura 1.1. Los puntos de esta gráfica fueron la semilla de la teoría cuántica. Planck observó que no concordaban con las predicciones de la ley de Wien, mostradas aquí como líneas continuas. El desarrollo teórico de una nueva ley capaz de ajustarse a todos los datos le condujo a proponer la idea de los cuantos de energía.

El domingo 7 de octubre de 1900, el experimentador del Reichsanstalt Heinrich Rubens y su esposa visitaron a Planck y a su familia en su hermosa villa de Grünewald, el elegante suburbio de Berlín en el que residía el profesorado. Los dos físicos charlaron sobre temas de trabajo y, apenas los Rubens abandonaron la casa, Planck puso manos a la obra para encontrar una ley mejor. Aquella noche, en su estudio —seguramente, de pie ante su elevado escritorio, como solía—, creó una versión modificada de la ley de

Wien que podía explicar los datos registrados por los experimentadores. Planck escribió enseguida una tarjeta postal a Rubens, hablándole de su nueva ley para la radiación de la cavidad que, doce días después, presentaría por primera vez en público en una reunión oficial de sus colegas de Berlín, Rubens incluido. Más tarde, el propio Rubens volvió a su laboratorio y a la mañana siguiente tuvo el placer de confirmarle a Planck que su ley había explicado correctamente los nuevos datos. Hasta la fecha, nadie ha encontrado una ley que se ajuste mejor a la intensidad de la radiación de la cavidad.

Desde el mismo día en que trasladó al papel su ley de la radiación de la cavidad, Planck comenzó a tratar de explicarla en términos de lo que realmente tenía lugar en el interior de un horno, al principio usando las leyes de la física clásica.¹⁹ Pronto se dio cuenta de que no tenía otra opción que hacer uso de los razonamientos estadísticos de Boltzmann, que hasta entonces había abominado, para entender cómo interacciona la radiación con los átomos de las paredes del horno. Partió de la definición estándar de las paredes, es decir, de la idea de que éstas consisten —como cualquier sólido— en un conjunto de átomos que vibran en torno a posiciones fijas, con una energía media que se incrementa a medida que calentamos el horno. Pero, en este caso, la manera en la que Boltzmann trataba las energías de los átomos no servía, con lo que a Planck no le

¹⁹ Contrariamente a la idea generalizada entre los físicos actuales, Planck comenzó sus trabajos sin saber que los datos de la radiación de la cavidad suponían una crisis para la física clásica. Este punto es analizado con detalle por Martin Klein, *Max Planck and the Beginnings of Quantum Theory*, Archives for History of Exact Science, 1962, vol. 1, págs. 459-479.

quedó más remedio que abandonar algunas de las premisas provenientes de la física clásica, que hasta aquel momento habían sido para él inamovibles, y hacer algo que le desagradaba profundamente: improvisar. Era como si, de repente, Artur Rubinstein tuviera que transformarse en Duke Ellington.

Durante las ocho semanas más agotadoras de su vida, encontró que sólo podía deducir su ley si modificaba drásticamente las técnicas estadísticas de Boltzmann y adoptaba una idea particularmente extraña. Tenía que dividir la energía total de los átomos que vibraban en las paredes del horno a cada frecuencia en cantidades discretas, poseyendo cada una de ellas la energía dada por la ecuación $E = hf$. Era la primera manifestación del cuanto de energía, el primer indicio de que la energía a nivel molecular es fundamentalmente distinta de la energía a la escala a la que estamos acostumbrados.

El concepto de cuanto de energía chocaba abruptamente con la idea de energía que tenía cualquier científico de la época. Se suponía que la energía, como el agua, podía existir en cualquier cantidad — podemos extraer agua del mar o devolverla a él en el volumen que deseemos—. La idea de que el agua pudiera tomarse sólo en cantidades discretas, por ejemplo en barriles de cierto tamaño, contradice la experiencia diaria y, sin embargo, ése es el modo en que parece comportarse la energía a nivel molecular. ¿Podría ser que, al igual que el agua en última instancia está compuesta por moléculas, la energía se diera en cuantos discretos o paquetes?

Planck presentó su ecuación por primera vez en una conferencia. El viernes 14 de diciembre, poco después de las cinco de la tarde, abordó la lectura de un breve artículo sobre su obtención de la ley de la cavidad ante un auditorio de físicos berlineses y en el marco de una de las reuniones quincenales de la Sociedad Alemana de Física. La primera vez que Planck mencionó la ecuación $E = hf$ durante su discurso lo hizo sin mayor énfasis y sus colegas, al parecer, se mostraron interesados pero no especialmente impresionados.

Según la opinión más extendida, esta presentación supuso la revelación al mundo de la idea del cuanto por parte de Planck.²⁰ Sin embargo, muchos historiadores de la teoría cuántica, entre ellos Thomas Kuhn, consideran que esta forma de verlo es simplista. Planck escribió que consideraba la cuantización de la energía «una premisa puramente formal, a la que no hubiera prestado demasiada atención de no ser porque debía obtener un resultado positivo bajo cualquier circunstancia y a cualquier precio». Afirmaciones como ésta convencieron a Kuhn de que, en 1900, Planck no comprendía realmente el significado de «cuanto de energía» y no creía que la energía estuviera cuantizada. Antes bien, según Kuhn, Planck creía como los demás que los átomos podían tener cualquier energía, pero

²⁰ Éste es el punto de vista de Martin Klein, quien lo desarrolla en su artículo «Thermodynamics and quanta in Planck's work», *Physics Today*, 19 (11), págs. 23-32. En *Planck's Original Papers in Quantum Physics*, Londres, Taylor and Francis, 1972, hallamos la traducción al inglés de los artículos originales de Planck, anotada por Hans Kangro.

él dividía esa energía en cuantos, simplemente porque dicho artificio matemático hacía que sus cálculos resultaran correctos.²¹

En lo que todos los expertos están de acuerdo es en que Planck apreció en toda su magnitud la trascendencia de su nueva constante, h .²² Durante varios años intentó justificarla en términos de física clásica y más tarde escribió que muchos de sus colegas consideraron erróneamente que su fracaso en este sentido había sido una tragedia. Al final tuvo que aceptar haber descubierto la última de un reducido grupo de auténticas constantes fundamentales —entre las que se encuentran la velocidad de la luz y la constante de la gravitación de Newton— que aparecen en las ecuaciones de física y cuyos valores no pueden ser deducidos. Un descubrimiento así es extremadamente infrecuente en la historia de la ciencia: desde que Planck encontró la constante h , no se ha identificado ninguna otra constante fundamental.

La teoría de Planck implicaba una segunda constante, k , relacionada con la teoría estadística de Boltzmann, por lo que Planck le dio su nombre con una generosidad que llegaría a lamentar, ya que ni Boltzmann la introdujo ni pensó nunca en averiguar su valor. Comparando las predicciones de su ley matemática con los datos de radiación de la cavidad del

²¹ Thomas Kuhn expone más claramente sus puntos de vista en «Revisiting Planck», *Historical Studies in the Physical Sciences*, 1984, 14 (2), págs. 231-252. Había realizado una presentación más detallada, pero más opaca, en *Blackbody Theory and the Quantum Discontinuity 1894-1912*, Clarendon Press, 1978.

²² Peter Galison hace un equilibrado análisis de la controversia generada por Kuhn en torno a la obra de Planck en «Kuhn and the quantum controversy», *British Journal for the Philosophy of Science*, 1981, 32, parte 1, págs. 71-85.

Reichsanstalt, Planck halló el valor de las dos constantes. La obtención de la constante de Boltzmann fue especialmente grata para Planck, pues le permitió efectuar la medición más exacta de la masa del átomo realizada hasta la fecha.²³ El valor de la constante permitió posteriormente a los científicos calcular la energía media de un átomo de cualquier sustancia en cualquier lugar del universo y a cualquier temperatura.²⁴

Aunque hoy nos parezca extraño, Planck se entusiasmó menos con su teoría cuántica y con la ecuación $E = hf$ que con la posibilidad de que la teoría diera lugar a un conjunto de unidades de medida de longitud, masa y tiempo de aplicación natural en cualquier punto del universo. En la Tierra medimos la longitud en metros o pies, el tiempo en segundos y la masa en kilogramos o libras por tradición o comodidad, pero no existe razón alguna de peso que haga que estas unidades sean mejores que otras. Si la historia hubiera seguido otros derroteros, tal vez ahora mediríamos la longitud respecto al tamaño del dedo meñique de Julio César y la masa y el tiempo, en relación con el peso de su corona y la frecuencia con que latía su corazón.

²³ Los valores encontrados por Planck para las constantes h y k varían muy poco respecto a los aceptados actualmente. En unidades modernas, $h = 6,63 \times 10^{-34} \text{ J}\cdot\text{s}$ y $k = 1,38 \times 10^{-16} \text{ JK}^{-1}$, donde J es el símbolo del julio, la unidad moderna de energía, s es el símbolo del segundo y K, el símbolo del grado Kelvin, la unidad de temperatura en la escala absoluta. En esta escala, la temperatura más baja posible es cero grados y el punto de congelación del agua en condiciones normales es 273,16 grados.

²⁴ La energía del átomo para cualquier sustancia —sólido, líquido o gas— es aproximadamente igual a la constante de Boltzmann multiplicada por la temperatura de dicha sustancia expresada en la escala absoluta.

Planck se dio cuenta de que la nueva constante h le permitía establecer unidades que, en vez de ser arbitrarias, emergían de las leyes de la naturaleza. Observó que podía calcular valores únicos de longitud, masa y tiempo empleando combinaciones especiales de su nueva constante universal con otras dos, la velocidad de la luz y la constante de la gravitación de Newton.²⁵ Planck razonó que si esas tres constantes habían sido siempre las mismas en todo lugar, los valores calculados de masa, longitud y tiempo debían ser también válidos en cualquier punto del universo y, por lo tanto, más naturales que los establecidos por una autoridad terrestre, por muy augusta que ésta fuera. Encontró que el valor único resultante para la masa corresponde más o menos a la de una ameba gigante (10^{-8} kilogramos), que el de la longitud es alrededor de una billonésima de una billonésima del tamaño de un átomo (10^{-35} metros) y que el relativo al tiempo resulta ser 10^{-43} segundos, en torno a una millonésima de una billonésima de una billonésima de una billonésima de lo que se tarda en pestañear. Ninguno de los tres resulta adecuado para la vida diaria, desde luego, pero lo importante es que Planck se encontró ante un nuevo hecho esencial: que no sólo existían algunas leyes que tenían validez universal y absoluta, sino que había un conjunto de unidades que la tenían también.

²⁵ Esos valores únicos se conocen hoy como masa, longitud y tiempo de Planck. Las fórmulas mediante las que se obtienen a partir de la constante h , la constante de la gravitación de Newton, G , y la velocidad de la luz en el vacío, c , son las siguientes: masa de Planck = $\sqrt{hc/G}$; longitud de Planck = $\sqrt{hG/c^3}$; tiempo de Planck = $\sqrt{hG/c^5}$. Sus magnitudes son importantes para los astrofísicos que estudian el origen del universo, tal como explica Joseph Silk en *A Short History of the Universe*, W. H. Freeman and Co, 1997, págs. 74-76.

En general, los colegas de Planck vieron su ley de la radiación de la cavidad como poco más que una fórmula matemática que, casualmente, se ajustaba a los datos. Ninguno de los gurús de la física afincados en Berlín vio claramente las implicaciones del trabajo de Planck y, en particular, de su nueva ecuación $E = hf$. Ese honor le correspondería a un joven licenciado que trabajaba casi en solitario en Suiza.

§. III

El poeta Paul Valéry tenía la costumbre de llevar un cuaderno en el bolsillo para anotar sus ideas. Cuando le preguntó a Einstein si él hacía lo mismo, Einstein replicó: «Oh, no es necesario», y añadió con melancolía: «se me ocurren tan pocas...». El comentario es de la década de 1920, cuando la parte más creativa de su carrera se aproximaba a su fin. Sólo Dios sabe cuántos cuadernos llenaría cuando aquélla estaba en sus comienzos.

En el otoño de 1900, cuando Planck escribió por primera vez su ecuación $E = hf$, Einstein se encontraba en Zúrich, dando clases particulares para ganarse la vida.²⁶ Algunos años antes había leído acerca del problema de la radiación de la cavidad, y en la primavera de 1901 estaba al tanto de los trabajos de Planck. Su reacción frente a ellos prueba que era poseedor de un talento especial. Nadie había visto más allá del virtuosismo matemático de Planck y del

²⁶ Albrecht Fölsing es autor de una hermosa biografía del gran físico: *Albert Einstein*, Penguin Books, 1997. Para un relato exhaustivo y documentado de las contribuciones de Einstein a la física, véase: Abraham Pais, *Subtle is the Lord...*, Oxford University Press, 1982. [Trad. esp.: *El Señor es sutil...*, Barcelona, Ariel, 1984].

éxito superficial de su fórmula a la hora de explicar los datos del Reichsanstalt. Sin embargo, Einstein se dio cuenta enseguida de que era el principio de una revolución. Más tarde escribiría que, tras la aparición de los trabajos de Planck, «fue como si nos hubieran retirado la tierra bajo los pies y no existiera cimiento alguno sobre el que empezar a construir».

Einstein reflexionó durante casi cuatro años sobre las implicaciones de la obra de Planck antes de publicar sus revolucionarias ideas sobre la luz —o, de forma más general, la radiación— y lo que sucede cuando ésta interacciona con la materia. Por aquel entonces, el joven y alegre físico había logrado empleo estable como examinador de patentes de «categoría III» en la oficina de patentes de Berna y se había casado con una antigua compañera de estudios. A comienzos del otoño de 1904 había nacido su hijo Hans, la familia se había mudado a un apartamento de dos habitaciones y Einstein disponía de un contrato fijo. De algún modo, en los intersticios de su vida en el hogar y en la oficina, desarrollaba sus ideas sobre la luz, la relatividad y la estructura molecular de la materia, asunto que adoptó como tema de su tesis doctoral. En lo que hoy se considera como una de las más espectaculares explosiones de talento en la historia de la ciencia, las tres líneas de investigación einstenianas dieron simultáneamente su fruto en 1905.²⁷ El primer artículo que publicó ese año, reconocido hoy como

²⁷ *Einstein's Miraculous Year*, editado por John Stachel (Princeton University Press, 1998), proporciona una introducción compacta y elegante a los grandes artículos que Einstein escribió en 1905, junto a una traducción al inglés anotada de cada uno de ellos.

su primera gran contribución a la ciencia, se refería a los cuantos de luz, una idea que él mismo calificaba de «muy revolucionaria» en una carta a un amigo.

El artículo es una joya. Aunque su lenguaje es moderado, buscando la comprensión, los razonamientos tienen la audacia de la que sólo las mentes más brillantes y libres son capaces. Einstein iba directamente al grano al afirmar con timidez que, contrariamente a lo establecido por la teoría ondulatoria de Maxwell sobre la luz, «la energía (de un rayo de luz emitido por una fuente puntual) no está distribuida de forma continua a lo largo de volúmenes de espacio siempre crecientes, sino que consiste en un número finito de cuantos de energía localizados en puntos del espacio, los cuales se mueven sin dividirse y sólo pueden ser absorbidos o generados como unidades completas». Para los seguidores de Maxwell —lo que equivale a decir para todos los grandes físicos de la época— se trataba de una auténtica herejía.

Einstein continuaba después enmendándole la plana a Planck, demostrando que el razonamiento que éste había utilizado para estudiar la radiación de la cavidad presentaba un fallo grave. Mediante matemáticas simples, Einstein demostraba que la energía total de dicha radiación era infinita según la física clásica. Si Planck se hubiera dado cuenta de ello, probablemente habría descartado su teoría y obviado el gran descubrimiento de los cuantos de energía, como el mismo Einstein comentaría más tarde. Desconfiando de la fórmula de Planck para la intensidad de la radiación de la cavidad,

Einstein empleó la más antigua de Wien, que se ajustaba mejor a los datos obtenidos para longitudes de onda pequeñas o, lo que es lo mismo, para frecuencias altas. Einstein observó que la ecuación para la densidad de esta radiación era exactamente la misma que la correspondiente a un gas de cuantos, cuando todos ellos rebotan independientemente unos de otros. Así pues, y aquí estaba el audaz paso, Einstein sugería que la radiación descrita por la ley de Wien se comportaba *como si fuese* un gas, ya estuviera *dentro o fuera* de la cavidad. La comparación le proporcionaba también una ecuación simple para la energía de cada «cuanto» en el gas de radiación: se trataba de $E = hf$ donde h es la constante de Planck y f la frecuencia de la radiación.

Aunque la fórmula de Einstein parece idéntica a la de Planck, ambas hablan de cosas totalmente diferentes: la del primero se refiere a la energía de cualquier cuanto de luz, mientras que la del segundo se ocupa del caso especial de la energía de los átomos en una cavidad cuando éstos interaccionan con la luz. Además, Einstein decía algo más sobre esa interacción: proponía que la materia absorbía o emitía radiación no mediante un flujo continuo —como afirmaba la teoría clásica—, sino *como si* (en palabras de Einstein) la radiación estuviera compuesta por cuantos. De este modo, la materia podía absorber o emitir uno, treinta y siete o cualquier otro número entero de cuantos de radiación, pero no dos y medio o, en general, un número fraccionario de ellos.

Si la energía es liberada en cuantos, ¿por qué no somos conscientes de que a nuestros ojos llegan paquetes individuales de energía? Aunque Einstein no se detuvo en esta cuestión explícitamente, sabemos que conocía la respuesta: la energía de un cuanto es tan diminuta y el número de cuantos que inciden en nuestros ojos tan inmensamente grande, que nuestro cerebro no es capaz de diferenciar la llegada individual de cada uno, con lo que los percibe como un flujo continuo. Cada cuanto de luz visible tiene sólo una billonésima parte de la energía del batido del ala de una mosca, lo que equivale a decir que una simple vela emite alrededor de mil trillones de cuantos por segundo —demasiados para que nuestros ojos puedan distinguir uno de otro.

Einstein concluía su artículo indicando cómo se podría verificar experimentalmente su idea de los cuantos de luz. Su propuesta más notable estaba relacionada con otro problema aparentemente oscuro: qué sucede cuando la radiación incide sobre un metal (el denominado efecto fotoeléctrico). El metal refleja y absorbe la radiación, pero los experimentadores habían constatado que ésta hacía que del metal se desprendieran algunos electrones. Si el concepto cuántico de la radiación es correcto, argumentaba Einstein, parece razonable suponer que cada electrón resulta desplazado por un único cuanto de radiación, de energía $E = hf$. Asimismo, si el cuanto cede toda su energía al electrón, la energía del electrón emergente es simplemente igual a la energía de la radiación —el cuanto— menos la energía necesaria para extraer el

electrón del metal. Einstein expresaba matemáticamente esta idea mediante lo que se conocería como su ley fotoeléctrica.

Einstein calificaba de «heurístico» —algo que sirve de ayuda para el estudio— su punto de vista sobre la luz, con lo que parecía evadir la cuestión de si los cuantos de luz eran reales. Es comprensible su postura; si su razonamiento era correcto, la venerada teoría de Maxwell estaba equivocada y la radiación se comportaba de forma corpuscular y no como una onda. Y si había algo sobre lo que todos los físicos estaban de acuerdo acerca de la radiación era que ésta respondía *realmente* al comportamiento ondulatorio. Disponían, además, de una prueba irrefutable: si se hace pasar un haz de luz por una rendija lo suficientemente estrecha, la luz resulta difractada —en términos menos técnicos, se difunde—. La luz difractada suele presentar patrones característicos de máximos y mínimos de luminosidad que sólo son explicables si se admite que la luz es una onda.

Einstein era consciente de ello y comprendía la fuerza del argumento, pero no le iban a intimidar unos hechos experimentales incómodos. Completó su artículo a mediados de marzo y lo envió a la revista de investigación física más importante del mundo, *Annalen der Physik*, editada por Planck y Wien. Planck era un editor particularmente eficaz: rechazaba sistemáticamente las propuestas mediocres, pero publicaba con gusto todo tipo de artículos, ortodoxos o heréticos, con tal de que estuvieran bien argumentados

y su lógica fuese consistente. El revolucionario artículo de Einstein fue publicado algunos meses después.

Es muy habitual rechazar una idea verdaderamente revolucionaria porque, simplemente, resulta incomprensible en el marco de los conceptos que trata de reemplazar. Así sucedió con el artículo de Einstein, cuya publicación pasó totalmente inadvertida en el mundillo de los físicos profesionales —algunos miles— y en el del resto de los científicos. No sería justo condenarlos: las ideas einstenianas no tenían sentido alguno en el marco de la teoría de Maxwell, la cual llevaba cuarenta años ajustándose a todos los experimentos. La propia ecuación $E = hf$ parecía un contrasentido, pues vinculaba la energía de un *cuanto* con la frecuencia de la radiación, un concepto que sólo tenía sentido si la radiación era una *onda*. Y además, ¿quién diablos era ese Einstein, después de todo?

Einstein no se desanimó. En enero de 1906, los artículos que había escrito el año anterior habían sido publicados sin que Planck o Wien plantearan objeción alguna y él se había convertido en «*Herr Doktor Einstein*». Aunque era feliz trabajando sus ocho horas en la oficina de patentes, empezaba a acariciar la idea de dedicarse por completo a la enseñanza. Mientras tanto, seguía pensando en la relación entre sus cuantos de luz y los trabajos de Planck. Inicialmente, Einstein había pensado que las dos teorías eran complementarias, pero ahora estaba convencido de que Planck había empleado la idea de los cuantos de luz, aunque fuera de forma implícita. Einstein se dio cuenta asimismo de que, en la teoría de Planck, la energía de

cada uno de los átomos de las paredes de la cavidad estaba también cuantizada —algo en lo que, desde luego, Planck no había caído—. Si todo átomo vibraba un número fijo de veces por segundo, es decir, con una frecuencia fija, su energía sólo podía ser un múltiplo entero del producto de la constante de Planck por la frecuencia. Así pues, la energía mínima de un átomo en vibración era $E = hf$ y los valores que esa energía podía adoptar eran $2hf$, $3hf$, $4hf$, etc. Einstein pretendía decir con ello que la ecuación $E = hf$ era aplicable a *cada uno* de los átomos de un sólido. No se trataba, pues, de una subdivisión matemática de la energía total del *conjunto de los átomos*, como Planck había propuesto.

En noviembre de 1906, Einstein mostró cómo comprobar su teoría mediante el proceso por el que los sólidos absorben calor, algo que la física clásica no sabía explicar. Imaginó un sólido cristalino ideal, con todos sus átomos, uniformemente espaciados en una matriz tridimensional, vibrando a la misma frecuencia e independientemente unos de otros. Einstein sabía que ese escenario no era del todo realista, ya que los átomos no se mueven de manera independiente, pero supuso que se trataba de una simplificación aceptable. Asumiendo también que la energía de vibración de cada átomo estaba cuantizada, predijo que la energía media de los átomos del sólido caía lentamente con la temperatura, hasta convertirse en cero. Sus predicciones concordaban con las enigmáticas medidas realizadas veinticinco años atrás. Su exitosa comparativa, la primera de las tres únicas ocasiones en las que

Einstein publicó una gráfica que contrastaba sus predicciones teóricas con los experimentos, inyectó credibilidad en la recién nacida teoría cuántica. Como no se hablaba de radiación, los físicos quedaron impresionados sin que tuvieran que enfrentarse a las molestas contradicciones con la teoría de Maxwell.

Debido sobre todo a su teoría de la relatividad, la fama de Einstein se propagó rápidamente entre los físicos, muchos de los cuales se sorprendían al descubrir que el autor de esa gran teoría trabajaba ocho horas diarias en una oficina de patentes. El primero en reconocer el talento de Einstein fue uno de los teóricos que él más admiraba, Max Planck. Ambos se encontraron por primera vez en septiembre de 1909 en un congreso celebrado en Salzburgo, en el que Planck le había invitado a dar la conferencia inaugural ante las luminarias de la física teórica. Para sorpresa de su anfitrión, Einstein no eligió como tema su aclamada y elegante teoría de la relatividad, sino «La naturaleza y constitución de la radiación».

Su conferencia fue toda una hazaña. Siguiendo la tradición de Clausius, obvió cualquier tipo de material experimental y toda clase de detalles matemáticos para concentrarse solamente en los principios. Su audiencia debió quedar sobrecogida al oír a ese nuevo miembro de su comunidad; en lugar de a una tímida disertación de principiante, asistían a un revolucionario manifiesto sobre la naturaleza de la luz. Einstein llevó más allá sus ya polémicas ideas sobre los cuantos de radiación al afirmar que, al igual que un electrón, todo cuanto viaja en una dirección específica —en lenguaje

técnico— posee un momento. Por primera vez, Einstein sugería en público que la radiación se componía de partículas. Por otra parte, y dado que la teoría de la relatividad había convertido el éter en algo superfluo, ya no era necesario pensar que la radiación necesitaba un «soporte» para existir, sino que era «algo que existía de manera independiente, al igual que la materia». En opinión de Einstein, el problema de comprender la radiación era tan importante que «todos deberían trabajar en él».

Los asistentes debieron de quedarse tan perplejos como quienes presenciaron el estreno del segundo cuarteto de Schönberg en la vecina Viena, nueve meses atrás. Planck, hablando en nombre de casi todos sus colegas, indicó respetuosamente que le parecía prematuro abandonar las ecuaciones de Maxwell y que no era proclive «a asumir que las ondas de luz estaban constituidas por átomos». Planck, sin embargo, había suavizado su oposición a la idea de los cuantos y llegado a aceptar a regañadientes que cuando la radiación interacciona con la materia, la energía se transfiere a los átomos constituyentes en cuantos discretos. Pero, al igual que la práctica totalidad de los físicos, no aceptaba que la propia energía de la radiación estuviese cuantizada.

En julio de 1909, Einstein presentó su renuncia en la oficina de patentes con objeto de asumir su primer cargo académico como un extraordinario profesor adjunto de física teórica en la Universidad de Zúrich. Empleaba la mayor parte de su tiempo en reflexionar sobre su problema favorito: la comprensión de los cuantos de luz.

Probó muchas aproximaciones diferentes e incluso trató de enmendar, sin éxito, las ecuaciones de Maxwell. El lenguaje que empleaba para describir los cuantos de radiación era invariablemente cauto, evitando afirmar de manera rotunda que los cuantos existían; en lugar de ello, hablaba de que la radiación se comportaba *como si* su energía estuviera cuantizada. No es raro que muchos de sus colegas pensaran que su compromiso con los cuantos era poco entusiasta.

Uno de los que no habían pasado por alto la ambigüedad de Einstein frente a los cuantos de luz era el norteamericano Robert Millikan, un excelente experimentador energético con una proverbial habilidad para incidir en las cuestiones más candentes del momento. En 1912, durante un semestre sabático de la Universidad de Chicago, visitó a sus colegas de Berlín y pasó gran parte del tiempo con Planck, con quien compartía su interés por la medición de las constantes fundamentales. Ambos discutieron sobre los cuantos de radiación y Planck le comentó lo mucho que discrepaba de las ideas de Einstein. Como otros visitantes, Millikan fue invitado a asistir a una velada musical en el hogar de los Planck. El físico acompañaba a su mujer en un recital de canciones alemanas y, según recordaba el propio Millikan cuarenta años más tarde, improvisaba hábilmente al piano.

Millikan era consciente de la importancia de realizar experimentos que aclarasen el modo en el que la radiación hace que se desprendan electrones de la materia, el efecto fotoeléctrico. Aunque

en absoluto creía que la teoría de Einstein fuese correcta, valoraba el hecho de que efectuara predicciones comprobables que contrastaban abruptamente con las de las teorías competidoras. Se trataba de una excelente oportunidad, si lograba superar las dificultades que acarreaban los experimentos, para arrojar luz sobre lo que los mejores físicos consideraban el problema más trascendente de aquel tiempo: comprender la radiación. Apenas regresó a Chicago, Millikan inició sus experimentos fotoeléctricos, que tardaría tres años en completar.

Mientras tanto, los cuantos de luz de Einstein habían despertado escaso interés y cosechado muy pocos partidarios. En junio de 1913, Planck puso sus reservas por escrito de modo confidencial cuando él y tres de sus colegas berlineses propusieron a Einstein como miembro de la prestigiosa Academia Prusiana de las Ciencias. En lo que, por otra parte, constituía un empalagoso reconocimiento de los logros de Einstein, Planck pedía disculpas en nombre de su joven colega por «haberse entusiasmado en exceso con sus especulaciones» y rogaba que esto «no se le tuviera en cuenta en su contra, ya que, incluso en las ciencias más exactas, no es posible alcanzar la verdadera innovación sin asumir algún riesgo». Einstein, en cambio, no era tan lisonjero con Planck, a quien describía como «tercamente aferrado a opiniones preconcebidas que son, sin duda alguna, falsas».

Ambos, sin embargo, estaban intrigados por las noticias procedentes de Copenhague acerca de la aplicación de las ideas

cuánticas a la estructura atómica por parte del más brillante de los físicos daneses. Niels Bohr, un joven en apuros por aquel entonces, había reelaborado la conocida imagen del átomo como un diminuto núcleo rodeado de electrones. En tres artículos publicados en 1913, Bohr afirmaba que, en cada tipo de átomo, los electrones sólo podían adoptar ciertas órbitas permitidas, las cuales correspondían a los valores característicos de energía de ese átomo.²⁸ Aunque no tuviera sentido en el marco de la física clásica, la teoría justificaba de un plumazo por qué cada átomo emite o absorbe luz de ciertas longitudes de onda —éstas corresponderían a los saltos cuánticos entre los valores característicos de energía—. Y lo que era aún mejor: hacía predicciones que concordaban con los experimentos. El modelo del átomo de Bohr se ajustaba perfectamente a las longitudes de onda de la luz emitida o absorbida por el átomo más ligero de todos, el de hidrógeno. Esta circunstancia convenció rápidamente a los teóricos cuánticos de que debían concentrar sus esfuerzos en comprender el átomo. Einstein reflexionó largamente sobre el tema y más tarde mostró cómo el modelo de Bohr explicaba la ley de Planck para la radiación de la cavidad, mediante una imagen en la que los cuantos de radiación, dotado cada uno de una energía $E = hf$, eran emitidos o absorbidos por los átomos de la cavidad.

²⁸ John L. Heilbron y Thomas S. Kuhn, «The genesis of the Bohr atom», *Historical Studies in the Physical Sciences*, 1969, vol. 1, págs. 211-290. En la biografía *Niels Bohr's Times: in Physics, Philosophy and Polity*, de Abraham Pais (Clarendon Press, 1991, págs. 132-159) hallamos un relato pormenorizado de los trabajos de Bohr sobre el átomo.

Cuatro meses antes de estallar la primera guerra mundial, Planck atrajo a Einstein a Berlín desde Praga, donde había estado trabajando durante un año. Mientras Planck participaba en la ola de patriotismo que precedió a la guerra, llegando a apoyar públicamente la causa alemana, el pacifista Einstein deploraba el conflicto, que constituía para él un alarde de insensatez colectiva. Ambos hombres, sin embargo, dejaron a un lado sus diferencias políticas y congeniaron bastante. Pocas semanas después de su llegada, Einstein fue invitado por Planck a una de sus veladas musicales. Interpretaron el Trío con piano en *re* mayor de Beethoven en compañía de un violonchelista profesional. Einstein ejecutó la brillante parte del violín de un modo un tanto irregular, mientras el maestro Planck se lucía en el segundo movimiento. Los dos grandes físicos disfrutaban haciendo música juntos; durante los dieciocho años siguientes, mientras socavaban irremediablemente los cimientos de la física clásica por el día, se limitaban a explorar las formas musicales más ortodoxas por la noche. El revisionismo musical del *Wozzeck* de Alban Berg, del *Oedipus Rex* de Stravinski o de la *Salomé* de Richard Strauss no estaba hecho para ellos.

En la primavera de 1915, Millikan había completado sus experimentos fotoeléctricos, constatando con sorpresa que la ley de Einstein era correcta. Pero ello no quería decir que su modelo cuántico de la radiación fuese acertado, como el propio Millikan se

encargó de puntualizar.²⁹ En la primera frase del artículo en el que publicó sus resultados escribió que la ley de Einstein «no puede, en mi opinión, ser considerada hoy en día como basada en ningún fundamento teórico satisfactorio». A pesar de ello, los resultados de Millikan añadieron credibilidad a la ecuación $E = hf$, como Einstein indicaría después al escribir sobre su ley fotoeléctrica.

A la por naturaleza escéptica comunidad científica no le iba a convencer una única evidencia experimental a favor de una ecuación que parecía estar reñida con decenas de otros experimentos. Algunos físicos cuestionaron los resultados de Millikan y varios grupos se dedicaron durante años a contrastar y ampliar dichos resultados, pero ninguno encontró discrepancia importante alguna respecto a las predicciones de Einstein. En cualquier caso, esto no demostraba que la radiación estuviera compuesta por partículas: la teoría del efecto fotoeléctrico trataba de la energía de la radiación, pero no decía nada sobre las direcciones en las que viajaban los cuantos. Para comprobar la validez del modelo de partículas, los experimentadores necesitarían demostrar que cada cuanto de radiación se desplaza a través del espacio en una dirección concreta. Los resultados de Millikan, junto con la credibilidad cada vez mayor de la teoría que los soportaba, convencieron a Einstein hacia 1917 de que las partículas de radiación existían realmente. «Ya no albergo duda alguna acerca de que los cuantos de radiación sean *reales*», escribía a un amigo

²⁹ Gerald Holton, «R. A. Millikan's struggle with the meaning of Planck's constant», *Physics in Perspective*, 1999, vol. 1, págs. 231-237.

algunos meses después, «aunque soy la única persona que opina así»; sin embargo, Einstein olvidaba a un buen puñado de científicos de inferior nivel que estaban claramente de acuerdo con su teoría.

Así estaban las cosas a principios de noviembre de 1919, cuando Einstein se convirtió en una celebridad internacional de primera categoría al anunciar a un grupo de astrónomos británicos que los resultados obtenidos parecían confirmar la superioridad de la teoría einsteniana de la gravitación frente a la teoría de Newton. Exhausto y desmoralizado tras la guerra, el mundo parecía dispuesto a venerar a un héroe.

Haydn confesaba en cierta ocasión que, aunque sus amigos alabasen a menudo su talento, siempre había tenido claro que su joven colega Mozart era mucho mejor que él. Planck pensaba lo mismo de Einstein, a quien denominaba «el nuevo Copérnico», por lo que no le sorprendió demasiado que éste se convirtiera en algo tan singular como una celebridad científica. Su teoría de la relatividad —y no su trabajo sobre los cuantos— dio pie a todo tipo de chistes en media Europa y se convirtió en tema de conversación en las reuniones de sociedad. Mucha gente ha creído que dicha teoría fue la principal preocupación de Einstein, pero en realidad fue al revés; como escribiría más tarde: «He reflexionado cien veces más sobre la teoría cuántica que sobre la teoría de la relatividad general».

La «fiebre Einstein» alcanzó su apoteosis en Nueva York en abril de 1920, cuando el famoso científico atravesó la gran manzana vitoreado por miles de entusiastas admiradores alineados a lo largo

de sus calles. Al mismo tiempo, en San Luis, sus argumentos sobre la naturaleza corpuscular de la luz estaban a punto de ser verificados en una serie de experimentos por Arthur Compton, uno de los físicos jóvenes más dotados de Estados Unidos.³⁰ Compton había reflexionado sobre la dispersión de las ondas de radiación por los electrones e imaginaba el fenómeno como si se tratase de olas que hacen oscilar las boyas de un puerto. Si la idea era correcta, la frecuencia de la onda tras la dispersión debía seguir siendo la misma que antes. Sin embargo, Compton halló que la frecuencia variaba: la radiación tenía una frecuencia más alta antes de ser dispersada. Era como si el simple hecho de dispersar la luz convirtiera el azul en rojo. Tras varios años de ensayos, Compton había sido incapaz de explicar su extraño descubrimiento, ni siquiera revisando los supuestos ampliamente aceptados acerca del tamaño y forma del electrón o adaptando el modelo cuántico de la radiación. En noviembre de 1922, las nubes se disiparon.

La clave del tema, según concluyó Compton sin conocer la propuesta que había hecho Einstein años atrás, era que los cuantos de radiación tenían un momento. De este modo, tanto la radiación como los electrones podían ser imaginados como si fueran partículas, y sus colisiones, como la versión microscópica de un billar. Compton elevó a teoría las consecuencias de esta idea y se

³⁰ Roger Stuewer, *The Compton Effect*, Nueva York, Science History Publications, 1975. Para un relato asequible sobre la obra de Compton y la de otros experimentadores estadounidenses cuyos trabajos forman parte de la historia oculta de $E = hf$ véase: Daniel J. Kevles, *The Physicists: The History of the Scientific Community in the Modern America*, Harvard University Press, 1997.

dio cuenta de que explicaba perfectamente los resultados de sus experimentos. En el fondo estaba la ecuación $E = hf$: la energía del fotón de rayos X dispersado es más baja que la del fotón original debido a que parte de ella se transfiere al electrón. De esto se deduce que la frecuencia de la radiación dispersada debe ser también menor. Compton se apresuró a anunciar sus conclusiones tres semanas antes de Navidad en una reunión de físicos celebrada en un gélido Chicago. La sensacional noticia de que los cuantos de radiación se comportaban realmente como si fueran partículas — que poseían a la vez energía y momento— pasó como un maremoto sobre las tranquilas aguas de la comunidad física internacional.

Como toda observación verdaderamente importante, la de Compton fue puesta en duda por los cautos y los escépticos. Tras dos años de comprobaciones, a finales de 1924 se había extendido el convencimiento de que los resultados de Compton eran correctos y de que cabía afirmar que su descubrimiento suponía «un cambio revolucionario en nuestras ideas acerca del proceso de dispersión de las ondas electromagnéticas». La hegemonía del modelo de Maxwell sobre la naturaleza ondulatoria de la luz había terminado. Ignorante de las ideas einstenianas sobre la naturaleza corpuscular de la radiación, Compton no mencionaba a Einstein en el artículo en el que relataba su descubrimiento. Pero Einstein quedó tan encantado como si lo hubiese hecho: se convertía en la primera persona en predecir correctamente la existencia de una partícula fundamental.

Tal vez «einstenón» hubiera sido un nombre plausible para esa nueva partícula pero, afortunadamente, nadie lo sugirió. En 1926, al químico norteamericano Gilbert Lewis se le ocurrió una denominación que se hizo popular inmediatamente. La acuñó en un artículo, hoy desacreditado, en el que proponía la idea de que la radiación estaba compuesta por átomos. Lewis usaba para esos «átomos» un término que se ha convertido actualmente en el sinónimo universal de partícula de radiación: el fotón.

§. IV

Mientras tanto, algunos científicos ajenos al entorno de Planck y Einstein habían utilizado la ecuación $E = hf$ para abrir otra rica veta en el mundo subatómico. Mientras la flor y nata de los físicos se dedicaban a averiguar si era aplicable a la radiación y, en caso afirmativo, de qué modo, en 1923 un oscuro físico francés hizo la asombrosa sugerencia de que la ecuación no sólo describía la radiación, sino la propia materia. Este crucial punto de vista condujo rápidamente a la moderna mecánica cuántica.³¹

El desconocido físico era Louis de Broglie, un estudiante universitario de treinta años que había comenzado su carrera científica formal sólo dos años antes, tras ser desmovilizado en la primera guerra mundial. De Broglie, un hombre taciturno y reflexivo, no era un estudiante cualquiera. Era un verdadero príncipe, miembro de una de las familias más ilustres de Francia

³¹ Esta parte de la historia de $E = hf$ está cubierta de manera exhaustiva por Max Jammer en *The Conceptual Development of Quantum Mechanics*, McGraw-Hill, 1966.

(su tatarabuelo Charles, de tendencia liberal, había celebrado en su día el advenimiento de la Revolución francesa, lo que no le evitó ser condenado a la guillotina un mes antes de la caída de Robespierre). De Broglie se graduó en historia y leyes en la Sorbona en 1913, pero decidió emprender la carrera de física persuadido por su hermano el duque Maurice de Broglie, un distinguido experimentador lo suficientemente acaudalado como para poseer un laboratorio en su propia casa. Los planes de Louis se vieron interrumpidos algunos meses después, al ser llamado a filas. Pasaría la mayor parte de la guerra trabajando, sin especiales privilegios, como operador de radio en la Torre Eiffel, que había sido habilitada como puesto de telegrafía sin hilos.

De Broglie comenzó su doctorado en teoría cuántica trabajando en el laboratorio privado que su hermano tenía en su lujosa mansión, un edificio de varias plantas a dos minutos del Arco de Triunfo y a un tiro de piedra de los Campos Elíseos. Mientras Compton trataba de interpretar sus datos en San Luis, Maurice de Broglie iniciaba a su hermano en las últimas técnicas experimentales y dirigía su atención hacia la necesidad de conjugar el modelo de ondas de la radiación con el modelo corpuscular. Como el propio Louis escribiría después, «las largas conversaciones con mi hermano sobre las propiedades de los rayos X [...] me llevaron a reflexionar en profundidad sobre la necesidad de asociar siempre los aspectos relativos a las ondas con los de las partículas». Fue esta línea argumental la que le condujo a plantearse una profunda cuestión

cuya respuesta iba a cambiar el curso de la ciencia: si las ondas electromagnéticas se pueden comportar como partículas, ¿las partículas como los electrones pueden comportarse como ondas?

Einstein había tocado el tema un año antes, pero el planteamiento de Louis de Broglie era más profundo y enérgico. De Broglie recordaba después haber tenido una especie de revelación, uno de esos *eureka* que son mucho más infrecuentes en la ciencia real de lo que cuentan las leyendas: «Tras una larga meditación en soledad, en agosto del 1923 vi que el descubrimiento hecho por Einstein en 1905 debía ser generalizado a todas las partículas materiales y, en particular, a los electrones». De Broglie había caído en la cuenta de que la ecuación $E = hf$ era aplicable tanto a la materia como a la radiación. Cuando Einstein utilizaba la ecuación para describir la radiación, tenía que explicar cómo estaba relacionada la energía de un cuanto con cierta propiedad de una onda, la frecuencia. De Broglie tenía el problema opuesto: todos estaban familiarizados con los electrones y con la existencia de otras partículas de materia, pero ¿qué diablos era esa onda asociada a ellas? Según De Broglie, toda partícula tiene asociada cierta onda de materia. Usando una analogía con la ecuación $E = hf$ escribió una sencilla fórmula para la longitud de onda de la onda asociada a una partícula libre —aquella en la que la resultante neta de las fuerzas que actúan sobre ella es nula—. ³² A medida que se incrementa el momento de la partícula,

³² La fórmula de De Broglie dice que la longitud de onda de un cuanto, ya sea de radiación o de materia libre, viene dada por la constante de Planck dividida por el valor del momento del cuanto.

argumentaba De Broglie, su longitud de onda disminuye y el valor de esa longitud de onda depende del de la constante de Planck, h .

«Parece un disparate, pero es una idea muy sólida», le dijo Einstein a un colega tras leer, en diciembre de aquel mismo año, la tesis doctoral de De Broglie. Era estupendo contar con la aprobación del maestro, pero ¿qué opinaba la naturaleza? De Broglie se había dado cuenta enseguida de que si la materia tenía propiedades ondulatorias, un haz de electrones debía poder ser dispersado — difractado— como si fuera un rayo de luz. Fue, sin embargo, un estudiante a quien se le ocurrió la mejor manera de comprobar la teoría. Walter Elsasser, un joven físico de la universidad alemana de Gotinga, hizo observar que si De Broglie cataba en lo cierto, un simple cristal debería difractar un haz de electrones que se hiciera pasar a su través. Elsasser calculó que, si los electrones eran acelerados bajo una diferencia de potencial de 150 voltios, debían tener una longitud de onda de una diezmilésima de micra, algo menos que la distancia entre átomos en un metal típico. Eran las condiciones ideales para la difracción, por lo que, si De Broglie tenía razón, los experimentadores debían detectar picos y valles en el número de electrones dispersados por un cristal a diferentes ángulos. Al igual que las ondas de agua o las de luz, los electrones debían resultar también difractados.

«Joven, está usted sentado sobre una mina de oro», le dijo Einstein a Elsasser. Sin embargo, el estudiante se convertiría en un mero «espectador catalítico» cuando dos experimentadores llevaron su

idea a la práctica. En agosto de 1926 habían tenido la suerte de asistir a una reunión en Oxford en la que las ideas teóricas de De Broglie y Elsasser flotaban en el aire. La reunión era la conferencia anual de la Asociación Británica para el Progreso de la Ciencia, en la que los científicos tenían la oportunidad de contactar unos con otros y con el público. Aunque los dos experimentadores no llegaron a relacionarse durante la conferencia, salieron de ella dándole vueltas a cómo podrían confirmar las ideas de De Broglie. Poco después, escribirían el capítulo final de la historia de $E = hf$.

El primero de ellos era Clinton Davisson, un enjuto y frágil experimentador que había ido adquiriendo una buena reputación como experto en la dispersión de electrones gracias a sus trabajos en los laboratorios de la Bell Telephone ubicados al sur de Manhattan, a pocas manzanas del mercado de carne.³³ Durante cinco años había llevado a cabo un programa de experimentos rutinarios, aunque cada vez más complejos, para ver lo que sucedía cuando un haz de electrones incidía sobre un blanco metálico. Hacer colisionar una cosa con otra puede parecer una forma un tanto pedestre de hacer ciencia, pero los experimentos de este tipo han demostrado ser especialmente fructíferos (recordemos los éxitos obtenidos por los aceleradores de partículas tras la segunda guerra mundial). Los resultados de Davisson no arrojaron nada especial hasta abril de 1925, fecha en la que estalló una botella de aire líquido en sus instalaciones: uno de los accidentes de laboratorio

³³ K. K. Darrow, «The scientific work of C. J. Davisson», *The Bell System Technical Journal*, 1951, vol. 30, págs. 786-797.

más afortunados en la historia de la ciencia. Antes del incidente, el blanco de níquel empleado por Davisson consistía en un conglomerado aleatorio de diminutos cristales pero, al reparar los daños causados por la explosión, convirtió sin darse cuenta la muestra en un único cristal de níquel. Cuando empleó este nuevo blanco para dispersar el haz de electrones, los resultados habían cambiado por completo, aunque la difracción de los electrones aún no fuera evidente.

Davisson había llevado sus resultados a la conferencia de Oxford y, al regresar, volvía con la sospecha de que la teoría cuántica podría explicarlos. De nuevo en Nueva York, utilizó la teoría de De Broglie para predecir la posición de los picos de difracción, pero no fue capaz de encontrarlos. Sin desanimarse por ello, inició, en compañía de un colega, un detallado programa de investigación para determinar de una vez por todas si los electrones eran difractados realmente. Su recompensa llegó en enero de 1927, cuando detectó varios picos de difracción claramente definidos. Unos sencillos cálculos demostraron que su situación correspondía exactamente a lo que predecía la fórmula de De Broglie, inspirada a su vez en la ecuación $E = hf$. Era la prueba de algo que apenas cinco años antes nadie hubiera siquiera considerado: que las partículas de materia podían ser difractadas.

A 4.000 kilómetros de allí, en la austera ciudad de Aberdeen, el físico inglés George Paget Thomson realizaba también parecidos experimentos con haces de electrones, pero usando energías mucho

mayores que las de Davisson.³⁴ Tras su regreso de la conferencia de Oxford, había empezado a trabajar junto a uno de sus estudiantes en la detección de la difracción de electrones, aunque en vez de un único cristal como blanco empleaba películas delgadas preparadas a tal efecto. Los experimentos fueron enormemente exitosos; usando una película de celuloide, Thomson observó un patrón de difracción cuya forma era exactamente la que predecía la fórmula de De Broglie. La observación cerraba un círculo singular para la familia Thomson: George Paget Thomson descubría que el electrón era una onda, apenas veintiocho años después de que su padre, «J. J.», descubriese que era una partícula. Y, como en tantas discrepancias entre padres e hijos, ambos tenían razón: el electrón se comporta como una partícula cuando los experimentadores comprueban sus interacciones, mientras que lo hace como una onda cuando estudian su propagación.

La consecuencia de todo lo anterior es que tanto la luz como la materia se pueden comportar a la vez como ondas y como partículas. Ésta es la historia de $E = hf$. Los científicos actuales usan rutinariamente la ecuación sin apenas recordar los veintiséis años de trabajo que costó desentrañar su significado. La ecuación es más conocida por predecir la energía de un fotón que por su papel clave a la hora de estimular el debate sobre la dualidad de la materia. De Broglie fue el primero en ver que la radiación y la materia tienen dos caras: muestran la faceta corpuscular en sus

³⁴ George P. Thomson, «Early work on electron diffraction», *American Journal of Physics*, 1961, vol. 29, págs. 821-825.

interacciones y la faceta ondulatoria en su propagación. Todo estudiante de física se queda un tanto perplejo cuando se enfrenta a esta dualidad, lo cual es comprensible: mantuvo en jaque a los mejores científicos durante años. La solución al problema llegó a finales de la década de 1920, cuando los físicos desarrollaron la teoría cuántica de campos,³⁵ que hacía posible una descripción unificada de materia y radiación. En medio del formidable aparato matemático de la teoría cuántica de campos, la sencilla ecuación $E = hf$ sobrevive como un vestigio original de la historia cuántica.

§. V

«Hay revoluciones científicas, pero no existen los revolucionarios», según el historiador científico Simón Schaffer. Para él es indudable que, de vez en cuando, se dan cambios importantes en la forma en la que los científicos contemplan y estudian la naturaleza, pero no cree en el tópico de científico innovador como un héroe solitario que transforma el modo de hacer ciencia tras un raptó de inspiración. «Nadie encabeza una revolución científica por sí mismo. Ni siquiera Planck o el propio Einstein».

¿La historia de los cuantos de energía confirma el punto de vista de Schaffer? No hay duda alguna de que se trató de una revolución — la aceptación de la teoría cuántica supuso una ruptura total con la tradición clásica—. Planck y Einstein merecen, por supuesto, su

³⁵ Steven Weinberg explica el concepto de teoría cuántica de campos en *Dreams of a Final Theory*, Londres, Hutchinson, 1993, págs. 18-19. [Trad, esp.: *Sueños de una teoría final*, Barcelona, Editorial Crítica, 1996].

aureola de grandes científicos, pero ¿podemos atribuir el origen de la revolución cuántica a uno de ellos dos en solitario? Tradicionalmente, los científicos consideran a Planck el padre de la teoría cuántica, pero no está nada claro que el gran físico comprendiera inicialmente el impacto que su teoría cuántica podía tener en nuestra forma de pensar sobre la energía, tal como Thomas Kuhn ha subrayado. Para Kuhn, la revolución cuántica empezó con Einstein. Pero ¿es justo considerarlo un revolucionario solitario? Es bien conocido que se basó ampliamente en la obra de Planck y que durante catorce años su defensa de los cuantos de radiación fue un tanto ambigua. Ni siquiera en 1924, entusiasmado ante la importancia de los resultados obtenidos por Compton, que avalaban el modelo corpuscular de la radiación, Einstein quemó las naves, afirmando que las partículas fueran reales. En lugar de ello, escribió en un artículo que «[El experimento de Compton] demuestra que la radiación se comporta como si estuviese compuesta de proyectiles de energía discretos...». Obsérvese el *como si*.

En mi opinión, la revolución cuántica no fue obra de un único revolucionario, sino de una «junta» desorganizada. No hubo divisa ni manifiesto alguno y muchos de los que pertenecieron a ella se hubieran sentido mucho más a gusto bajo el *antiguo régimen*. Tampoco está claro cuándo se constituyó la hermandad y quiénes fueron sus verdaderos miembros. Además de Einstein y Planck, habría que incluir a De Broglie y a una legión de experimentadores: los expertos en radiación de la cavidad del Reichsanstalt, que

proporcionaron a Planck datos cruciales; Robert Millikan, que verificó la ley fotoeléctrica de Einstein; Arthur Compton, que fue el primero en demostrar la existencia de los fotones, y, por supuesto, Clinton Davisson y George Thomson, quienes demostraron que los electrones se pueden comportar como ondas.

La junta tenía también sus disidentes. Millikan y Compton, por ejemplo, realizaron gran parte de sus trabajos experimentales al tiempo que rechazaban las explicaciones cuánticas a las que se adherirían más tarde. Nadie se esforzó más que Planck por preservar la herencia de la física clásica y siempre nos quedará la duda de si realmente llegó a asumir todas las implicaciones de la revolución cuántica. En 1927, dos años después de que la práctica totalidad de la comunidad científica hubiera aceptado que los fotones existen, Planck escribió un breve artículo para el Instituto Franklin de Filadelfia en el que explicaba por qué aún no estaba preparado para creer en ellos. No consta que posteriormente cambiara de opinión.

La amistad entre Einstein y Planck fue durante años muy estrecha, pero cesó abruptamente cuando este último llegó a un acuerdo con Hitler. Se habían reunido por primera vez en 1933, poco después de que Hitler se convirtiera en canciller, y Planck hizo muchas concesiones al régimen nazi, pues hacía uso habitualmente del saludo hitleriano, daba las clases en aulas decoradas con esvásticas y en sus discursos omitía sistemáticamente los nombres de importantes físicos judíos, incluido Einstein. Horrorizado ante la

corrupción del aparato estatal propiciada por el Führer, la devoción de Planck por las instituciones alemanas de física, su patriotismo y, probablemente, su ingenuidad le llevaron, a pesar de todo, a realizar concesiones cada vez más humillantes. En muchos aspectos, se comportaba como el protagonista de cierta novela de Kazuo Ishiguro: alguien cuya manera de contemplar el mundo que le rodea está cada vez más lejos de los acontecimientos.

Hacia el fin de la guerra, en la noche del 15 de febrero de 1944, un ataque aéreo destruyó el suburbio de Grünewald en el que Planck había vivido durante casi cincuenta años. Su biblioteca, en la que probablemente había escrito por primera vez la ecuación $E = hf$, así como las habitaciones en las que había hecho música junto a Einstein, quedaron destruidas. El estoicismo de Planck fue puesto a prueba hasta unos límites insoportables: había perdido a su hijo mayor en las trincheras de la primera guerra mundial y a sus dos hijas en el parto, y su hijo pequeño sería asesinado por la Gestapo en febrero de 1945.

La muerte le llegó a Planck, casi como una liberación, en octubre de 1947, pocos meses antes de cumplir los noventa años. En la ceremonia realizada en su memoria en Gotinga, una brillante mañana de primavera, la música de Bach, Brahms y Beethoven se entremezcló con los elogios. Seguramente, todos los físicos allí reunidos consideraban ya que el fotón era una partícula más, algo en lo que Planck no había logrado ponerlos de acuerdo en vida. La

razón nos la había dado él mismo en las notas autobiográficas escritas pocos meses atrás:

«Un nuevo paradigma científico no triunfa convenciendo a los escépticos y haciéndoles ver la luz, sino gracias a que esos escépticos acaban muriendo y crece una nueva generación acostumbrada a él».

Einstein fue uno de los dignatarios que enviaron un mensaje de condolencia. Aunque nunca perdonó a Planck el haber colaborado con los nazis, escribió un sentido homenaje a su antiguo amigo, en el que afirmaba que

«Un hombre al que le ha sido dado bendecir al mundo con una grande y creativa idea no necesita alabanzas de la posteridad».

Einstein escribió estas palabras en el Instituto de Estudios Avanzados de Princeton, lugar donde residía desde 1933. Había abandonado Alemania en diciembre del año anterior, siete semanas antes de la llegada de los nazis al poder, y ya nunca regresaría.

A los ojos de mucha gente, Einstein era una autoridad capaz de darle una oportunidad a Dios con sólo mencionarlo, como escribiría Saúl Bellow. Pero para la mayoría de sus colegas científicos, Einstein era por aquel entonces una desconcertante figura que se obstinaba en no aceptar la teoría cuántica que los físicos habían terminado de formular en la década de 1920. El gran físico era feliz,

como siempre, trabajando en solitario, siendo un individualista intelectual. La música continuaba apasionándole. Aunque había dejado de tocar el violín poco después de la segunda guerra mundial, el Cuarteto Juilliard logró persuadirle para que improvisara una interpretación con ellos con motivo de una visita a su casa en el otoño de 1952. Cuando le propusieron elegir la pieza, Einstein escogió de inmediato el conmovedor Quinteto en *sol* menor de Mozart, del que interpretó el segundo violín, aunque sin ser demasiado fiel a la partitura.

Si Planck había sido el Moisés de la ciencia cuántica, Einstein fue su Josué. Desde su Monte Nebo en Berlín, Planck vio la Tierra Prometida y envió a sus seguidores a ella, pero él nunca realizaría el viaje. Fue Einstein quien los condujo hasta allí, aunque para él el recorrido aún no había terminado, como se deduce de sus últimas palabras sobre los cuantos de radiación, escritas en diciembre de 1951: «Estos cincuenta años de reflexión no me han permitido aún responder a la cuestión de "¿qué son los cuantos de luz?". Hoy en día hay muchos que creen conocer la respuesta, pero se equivocan». Para entender de verdad el fotón, según Einstein, la teoría cuántica no era suficiente.

Cincuenta años después, la teoría no ha sido cuestionada por los experimentos y, menos aún, sustituida por otra. En todo caso, si sobreviniera una contrarrevolución, Einstein se convertiría en el miembro fundador de su junta, a título póstumo. Y su fama

recibiría un impulso adicional que subrayaría su más grande contribución a la ciencia.

Agradecimientos

Son muchos los colegas y amigos que me han proporcionado un valioso asesoramiento durante la preparación de este ensayo. Me complace dar especialmente las gracias a David Cahan, Stuart y Corinne Freake, John Heilbron, Dieter Hoffmann, Russell McCormmach, Simón Schaffer, Chuck Schwager y Andrew Warwick.

Parte 2

La ecuación del sextante

$$E = mc^2$$

Peter Galison

El 17 de noviembre de 1945, John Wheeler, físico de Princeton, veterano del Proyecto Manhattan y pionero de una nueva era en la física, daba un repaso al estado de esta ciencia ante los asistentes a un simposio. Empezó recordando aquel primer momento de la era nuclear, a comienzos de la guerra, en la Universidad de Chicago. Una figura clave en el escenario bélico había telefoneado a Washington para relatarle al presidente de Harvard y jefe del Comité Nacional de Investigaciones Científicas, James Conant, los acontecimientos de los que el físico refugiado Enrico Fermi acababa de ser protagonista: «El navegante italiano ha descubierto América». «Espléndido», replicó Conant, «¿y el nuevo continente es seguro?». La respuesta fue: «Sí, y Colón cree que los nativos son amigables». Era el 2 de diciembre de 1942 y la conversación en clave hizo saber a Conant y a todos los responsables del equipo científico norteamericano desplegado para la guerra que el primer reactor nuclear del mundo había iniciado de forma segura una reacción en cadena automantenida. Los físicos habían desembarcado en el continente de la fisión nuclear aplicada y empezaban a vislumbrar la energía y también el poder destructivo que anidaban en el corazón del átomo de uranio. A lo largo de los siguientes treinta y

dos meses, los científicos del proyecto de la bomba atómica caminaron inexorablemente hacia la creación de un arma nuclear, hasta culminar —o, más bien, hacer una pausa— en el cataclismo que asoló Hiroshima y Nagasaki en agosto de 1945.

Ahora, cuando Wheeler hablaba, apenas habían transcurrido tres meses desde el final de la guerra. La física, hasta hacía no mucho un oscuro reducto académico, era actualmente el centro de la atención del país. Contemplando la física y a la sociedad en torno a ella, Wheeler veía la «formación del nuevo mundo» augurado por la física nuclear, «el gran continente que se extiende más allá [de la fisión] y que representa el último territorio inexplorado para el conocimiento del universo físico». Los matemáticos de la época de Colón, comentaba Wheeler, podían engañarse a ellos mismos sobre cuán lejos había llegado el explorador en su intento de circunnavegar el globo. Los físicos de la década de 1940, por el contrario, no podían hacer lo mismo respecto a lo que aún aguardaba ser descubierto. Los científicos tenían ahora en sus manos un sextante de una simplicidad tal que no dejaba margen para el engaño. Ese instrumento teórico, esa medida del progreso científico le diría en todo momento a la raza humana cuán lejos había progresado hacia la aniquilación total de la materia en energía. Aun siendo poderosa, la fisión del uranio había hecho avanzar a la humanidad apenas una milésima parte del camino hacia la conversión total en energía, pues sólo una milésima parte de la masa del átomo de uranio se transforma en pura energía

cuando el núcleo se divide. Una transformación total de materia en energía supondría, por el contrario, el límite último en la generación de ésta, la máxima eficiencia alcanzable en la producción de energía para construir un nuevo mundo industrial. O para crear un arma de inimaginable potencia. Y el sextante de la ciencia moderna que mediría el progreso alcanzado, mostrando a la humanidad su lugar exacto en la escala de la conversión total, era $E = mc^2$, de Albert Einstein, la ecuación más famosa de la historia de la ciencia.

Una ecuación que significa que, si al escindir un átomo de uranio se pierde una masa de m gramos (las partes pesan menos que el todo), la cantidad de energía liberada en el proceso de fisión será E (en ergios), donde E resulta de multiplicar la masa m por el cuadrado de la velocidad de la luz en el vacío (treinta mil millones de centímetros por segundo). Curiosamente, en su primer artículo, Einstein no utilizó la letra E para la *Energie* o *Energía* (en alemán y griego, respectivamente) ni la c para la *celeritas* (velocidad en latín), sino L (simbolizando tal vez *lebendige Kraft*, energía cinética) y V (para la velocidad de la luz). Aunque hoy los símbolos que componen $E = mc^2$ nos parecen inevitables a quienes hemos crecido con ellos, Einstein utilizó la E y la c sólo a partir de 1912. La energía puede ser liberada de varias formas: en la versión más simple de fisión nuclear, un átomo de uranio se escinde en dos núcleos más pequeños, que se separan el uno del otro a enorme velocidad. La energía liberada por la fisión de un único átomo de uranio sería suficiente para hacer saltar de manera visible un grano de arena

encima de una mesa; la energía de fisión contenida en el cuatrillón de átomos de que consta un kilogramo de uranio podría destruir —y destruyó, de hecho— varios kilómetros cuadrados de una ciudad.

A finales de 1945, la fisión, el fenómeno físico que subyace en los reactores nucleares y en las bombas atómicas, todavía presentaba interrogantes, pero era, en gran medida, una técnica dominada. Más allá de la cascada de neutrones de la reacción en cadena (neutrones que escinden núcleos de un modo tal que emergen más neutrones para romper otros núcleos, lo que a su vez genera nuevos neutrones...), existe aún toda una panoplia de fenómenos, por completo fuera del control de los físicos. ¿Qué hace que las colisiones de protones y neutrones produzcan nuevas partículas? Cada mes llegaban nuevos datos sorprendentes sobre esos intrigantes procesos, obtenidos de las observaciones de los rayos cósmicos —compuestos básicamente por protones— que inciden en las capas altas de la atmósfera terrestre provenientes del espacio profundo. Según Wheeler, «la posibilidad de una conversión completa de materia en energía es sugerida por la hoy día incompleta información sobre la producción de partículas de baja masa a partir de los protones en la alta atmósfera de la Tierra». Wheeler soñaba con un proceso que convirtiera la *totalidad* de una porción de materia en energía.

Comprender la naturaleza de esas transformaciones de las partículas era algo que fascinaba a Wheeler y a sus contemporáneos. Pronto embarcarían equipos de físicos en

bombarderos recién llegados del frente a investigar la alta atmósfera; en los polígonos de ensayos de White Sands y junto a científicos alemanes capturados, Wheeler lanzaría cohetes V-2 no tripulados cargados de instrumentos, que alcanzaban más de cien kilómetros de altura. Allí les aguardaban ráfagas de partículas de alta energía procedentes del espacio profundo —especímenes demasiado escasos para justificar una campaña de investigación a gran escala—. Lo que hacía falta era una fuente de partículas energéticas continua y abundante; en este sentido, el espacio no podía competir con el proyecto de construir aceleradores de partículas más grandes y más potentes. También se necesitaban observaciones que registrasen los cambios inducidos en los fragmentos de materia al ser golpeados con partículas de alta energía. Y, por último, los físicos tendrían que formular una teoría nueva y consistente que plasmara las relaciones entre las partículas elementales y las fuerzas que gobernaban sus interacciones.

Según Wheeler, la ecuación del sextante, $E = mc^2$, guiaría a los físicos en el manejo de aceleradores, rayos cósmicos y teorías hacia la creación de un nuevo campo en la ciencia: la física de las partículas elementales. Y así ha sido. En las décadas posteriores, los aceleradores de partículas golpearon blancos estacionarios con proyectiles cada vez más rápidos y luego pasaron a hacer chocar partículas con sus correspondientes antipartículas. En los nuevos aceleradores, los electrones colisionaron contra los positrones y los protones, contra los antiprotones, incrementando progresivamente

la cantidad de energía y penetrando cada vez más en la física de lo muy pequeño. Desde finales de la década de 1940 hasta los primeros años del siglo XXI, ese floreciente campo de la física ha venido utilizando la conversión de energía en masa para traer al mundo observable los componentes básicos de la materia. Comenzando con el protón, el neutrón, el electrón y el positrón, la población del zoo de las partículas creció a medida que los físicos utilizaron la energía producida en las colisiones para crear nuevos especímenes. Ya en 1932, el positrón —la antipartícula del electrón— había aparecido en un recipiente de ensayo, mostrando de un modo espectacular que la materia y la antimateria se aniquilan entre ellas, dando lugar a pura energía, y que la simple energía puede producir una partícula y su antipartícula gemela.

En las décadas que siguieron a la segunda guerra mundial se consiguió generar, y luego manipular, partículas como el pión, con una masa intermedia entre las del protón y el electrón. De las colisiones de protones y mesones contra núcleos surgieron variantes más pesadas de neutrones y protones, y la colección siguió creciendo. Cuando electrones y positrones, piones y antipiones o protones y antiprotones pudieron ser dirigidos con precisión los unos contra los otros, su aniquilación fue completa y la totalidad de la energía equivalente a ambas partículas se aplicó a la generación de nuevos entes subatómicos. En los años sesenta y setenta, a esas parejas de partículas se les unieron los quarks y antiquarks en sus distintas variedades, junto con versiones más pesadas del electrón y

nuevas partículas portadoras de fuerzas, dando lugar al *Modelo Estándar* de física de partículas. Tras los enormes aceleradores que durante más de treinta años han hecho colisionar partículas contra sus antipartículas gemelas siempre se ha hallado la ecuación $E = mc^2$. Fruto de las observaciones obtenidas en esas grandes instalaciones, la formulación canónica de la física de partículas, realizada en la década de 1970, permanece intacta desde entonces. En aquellos meses finales de la segunda guerra mundial, la equivalencia entre masa y energía constituía a la vez una promesa sin límites y una tremenda amenaza. En junio de 1945, Wheeler reflexionaba: «Descubrir el modo de liberar esa energía latente a una escala razonable podría alterar por completo nuestra economía y las bases de nuestra capacidad militar. Por este motivo, debemos prestar especial atención a la rama de la ultra-nucleónica [la física que va más allá de la entonces relativamente bien conocida física de los nucleones, es decir, de los protones y los neutrones]». Ese campo más lejano traería consigo una física nunca vista en los laboratorios del periodo de la guerra: fenómenos de rayos cósmicos, teoría de campo en la física de mesones, producción de energía en las supernovas y física de la transformación de partículas. La investigación abstracta a una escala inferior a la nuclear, según Wheeler, tenía una conexión clara con la «potencia bélica del país». Sabía perfectamente que entre los tópicos de la *ultra-nucleónica* estaba la posibilidad de un aprovechamiento más eficiente de la

energía prometida por $E = mc^2$ que esa minúscula milésima parte liberada por la fisión nuclear.

La liberación parcial de energía por parte de la fisión significaba que Hiroshima había sido destruida por la transformación de una masa considerablemente menor que la de una aspirina. Dichas reflexiones habían llevado a Wheeler —y a otros muchos físicos— a preguntarse si la ecuación del sextante marcaría el rumbo hacia una liberación mucho más eficiente de energía.

Antes de que en el lugar donde sólo había una escuela comarcal se levantara el laboratorio de armamento de Los Alamos, un pequeño grupo de expertos atómicos se reunía en Berkeley para hablar de armas nucleares. J. Robert Oppenheimer estaba allí en calidad del más importante teórico cuántico de Estados Unidos. También estaba Hans Bethe, el físico que, antes de huir de Alemania en 1935, había descubierto la física nuclear que hace brillar al Sol. Les acompañaba un grupo estelar, del que formaba parte el refugiado húngaro Edward Teller, conocido después por ser el «padre de la bomba H». En las toscas instalaciones de aquella primera época, las armas de fisión les parecían triviales: bastaba con reunir la suficiente cantidad de uranio fisible para que éste detonara. Asignaron el proyecto a un joven físico de Berkeley, Robert Serber, y ellos se reservaron un problema infinitamente más sutil y provocativo: la bomba de hidrógeno, o bomba H. La bomba H funcionaría juntando núcleos de baja masa, como los de hidrógeno, en vez de núcleos pesados, como los de uranio. Pero, a medida que

el laboratorio de Los Alamos tomaba cuerpo, se hacía cada vez más evidente que construir una bomba atómica distaba de ser trivial. Algunos jefes de proyecto, incluyendo a Oppenheimer y Bethe, dejaron aparcada la bomba H, dando prioridad al objetivo de producir una bomba utilizable antes del fin de la guerra. Edward Teller, sin embargo, se aferró tenazmente a su idea y, abstrayéndose de la guerra, se apartó de la línea principal basada en la fisión y prosiguió su lucha por obtener el arma que cautivaba sus sueños.

El 12 de agosto de 1945, desde la isla de Tinian en el Pacífico, el escenario de sus ensayos nucleares, Wheeler escribía a Teller: «Querido Edward. Con el final de la guerra hoy mi trabajo aquí terminará en breve. [...] Lo más provechoso a lo que podría dedicarme a partir de ahora es, según creo, la investigación fundamental. Pero me temo que no me sentiría demasiado cómodo si lo hiciera en los próximos cinco años». Recordaba la antigua invitación de Teller a trabajar en la bomba de fusión y su propio convencimiento de que la bomba H era un arma destinada a la siguiente guerra y no a la actual contra el Eje. Con la capitulación de Japón, ese nuevo conflicto resultaba inevitable a los ojos de Wheeler; estaba convencido de que en breve comenzaría una guerra contra la Unión Soviética y, en ella, la fusión, la conversión de una proporción mayor de materia en energía (según $E = mc^2$), resultaría vital. Por motivos de seguridad, Wheeler continuaba con metáforas: «Hay un grupo de hombres completamente aislados en una isla. Ha comenzado una pelea. Dos grupos de hombres con formas muy

distintas de hacer las cosas se han unido y han intentado apaciguar a los alborotadores. Nuestro grupo ha aprendido a utilizar a la vez un arco y una flecha. Por medio de ellos hemos logrado poner fin a la lucha. Nuestro aliado observaba. Ahora que el conflicto ha terminado ha regresado a su territorio, tras el muro. Sabemos que a muchos de sus hombres les gustaría fabricar un arco y una flecha por ellos mismos. Sospechamos que algunos de ellos no dudarían en usar ese arco y esa flecha contra nosotros si algún día no nos ponemos de acuerdo sobre quién puede coger la fruta de un árbol que hay por aquí. Ya sea por una razón u otra, los antiguos aliados no parecen ser capaces de poner el arco y la flecha en manos de alguien que los custodie y en quien ambos confíen. [...] Algunos de nuestro grupo se encogen de hombros y hacen planes para irse a pescar. Yo soy de los que opinan que si va a haber una carrera de armamentos, mejor empezar ahora y tratar de fabricar la mejor arma posible: una ametralladora que deje en ridículo a ese arco y a esa flecha».

Wheeler concluía diciendo que, en su opinión, más valía empezar a pensar ya en la «ametralladora» si el conflicto podía estallar en los próximos cinco a diez años. Consecuente con sus ideas, puso en marcha un proyecto en Princeton para diseñar la bomba H —el proyecto se denominó «Matterhorn B»—, junto a otras investigaciones más pacíficas sobre la transformación de materia en energía. De hecho, puerta con puerta del despacho de Wheeler en el proyecto Matterhorn estaba la sede de otro proyecto, cuyo objetivo

era la producción de energía para uso civil mediante la fusión nuclear. Con la fusión, en cada choque de núcleos se liberaría mil veces más energía que en la fisión. Una bomba del tamaño de las empleadas en Hiroshima y Nagasaki podría producir una energía equivalente a la explosión de entre diez y veinte millones de toneladas de TNT, en lugar de las diez a veinte mil toneladas a que equivalían las bombas atómicas de la segunda guerra mundial. Y, en principio, se podía pensar en bombas de una capacidad destructiva sin límites; en pocos años, la gente comenzó a especular sobre la producción de bombas de hidrógeno de 1 gigatón que abrirían un agujero a todo lo largo de la atmósfera. En todo momento, Wheeler veía en la ecuación del sextante la brújula que iba a dar las coordenadas del mapa que conducía hacia la conversión total de materia en energía, tanto en la guerra como en la paz. Porque, a pesar de su tremenda potencia destructiva, hasta en la bomba de hidrógeno mucha de la masa original quedaba sin transformarse en energía.

Una línea de investigación pacífica condujo a Wheeler a imaginar un nuevo tipo de átomo: un electrón y un positrón orbitando el uno en torno al otro. Tras apenas una diez mil millonésima de segundo, el «positronium» se desintegraría debido a que sus dos componentes se precipitarían uno contra otro, aniquilándose mutuamente y liberando su energía en forma de dos fotones. Se trataba de un bello y típico ejemplo de $E = mc^2$: si tanto el electrón como el positrón tenían una masa m , la energía liberada sería $2mc^2$ y los fotones

tendrían una frecuencia f dada por la ecuación $hf = mc^2$ (ya que, como Einstein había demostrado en 1905, la energía de un fotón es $E = hf$). Cabía, pues, buscar esos dos fotones alejándose en direcciones opuestas. Los físicos del MIT los hallaron poco después en un experimento, en la frecuencia exacta anticipada por Wheeler, por medio de la ecuación de Einstein.

A partir de aquí, el horizonte parecía abrirse a miles de transformaciones. Los nucleones que chocaban entre ellos en las cámaras de niebla (recipientes llenos de vapor de agua, el cual hace visibles las trayectorias) producían toda una constelación de nuevas partículas. ¿En qué grado aumentaba la probabilidad de esas «explosiones nucleares» con la energía de las nuevas partículas? ¿Qué era lo que determinaba el tipo y el número de productos generados por las explosiones? En lo que a Wheeler y a muchos de sus colegas se refería, responder a preguntas como éstas era lo que llevaba a la física cada vez más cerca de aprehender en su totalidad el funcionamiento de su sextante mágico, $E = mc^2$.

Fisión, fusión, positronium, aceleradores, rayos cósmicos, dinámica de los agujeros negros... Buena parte de la física del siglo XX gravita sobre esa ecuación tan simple. Pero su origen se halla lejos de los grandes laboratorios como el Fermilab de Chicago o el CERN, ubicado en la frontera franco-suiza; lejos de los laboratorios de armamento de Los Alamos, Livermore o Arzamas-16. El joven Einstein nunca pudo imaginar todos esos desarrollos cuando escribió por primera vez la ecuación.

Debemos regresar al mundo de Einstein, al mundo que lo rodeaba cuando no era más que un empleado de una oficina de patentes en 1905. Era un mundo en el que la electrificación era la piedra angular de la modernización. Había legiones de obreros levantando las calles para instalar las vías de los tranvías eléctricos y los electricistas sustituían las lámparas de gas por bombillas en techos y paredes. Entrecruzándose a lo largo de Estados Unidos, Europa y Rusia, las compañías eléctricas tejían una inmensa red de líneas de alta tensión, generadores eléctricos y aparatos de medida con objeto de suministrar energía a las fábricas, ciudades y viviendas. La propia familia de Einstein —en concreto, su padre y un tío— poseía un pequeño taller electrotécnico en el que se fabricaban dispositivos de relojería para medir magnitudes eléctricas. Las ecuaciones de Maxwell, que hoy se enseñan en cualquier clase de física elemental, eran tan novedosas en la década de 1870 que no se estudiaban en su totalidad, ni siquiera en las escuelas más avanzadas, y a Einstein todas esas nuevas teorías y dispositivos le parecían fascinantes. La oficina de patentes de Berna había contratado a un Einstein de veintitrés años precisamente para examinar innovaciones electrotécnicas; su trabajo consistía en establecer su grado de novedad y aislar y determinar los principios que les servían de base.

Desde esa oficina de patentes y en su *annus mirabilis* de 1905, Einstein publicó cinco extraordinarios artículos. El primero de ellos fue recibido en la revista *Annalen der Physik* el 18 de marzo y

presentaba su teoría de los cuantos de luz; en cierto sentido, fue el artículo que dio el pistoletazo de salida a la física cuántica. Seis semanas más tarde, el joven físico presentaba su tesis doctoral, en la que mostraba cómo estimar el tamaño de una molécula mediante un argumento basado en la contribución de las grandes moléculas del tipo del azúcar a la viscosidad del agua azucarada. El 11 de mayo, Einstein presentaba su teoría del movimiento browniano, que demostraba la existencia de choques entre los átomos y moléculas reales y las pequeñas partículas en suspensión —pensemos en las partículas de polvo difundiéndose en el aire—. Se trataba de un poderoso argumento a favor de la realidad física de los átomos; en la teoría de Einstein, los átomos no eran meras ficciones *ad hoc* para explicar los procesos químicos, eran objetos físicos reales que impactaban estadísticamente contra la partícula en suspensión y era la resultante de esos impactos lo que hacía que aquella se moviera a través del fluido.

Los artículos más relevantes para el tema que nos ocupa eran el cuarto y el quinto, los que Einstein envió a finales de junio y en septiembre. Fue en esos artículos, sin duda los más famosos, donde Einstein presentaba la teoría especial de la relatividad y deducía como una consecuencia la célebre ecuación de la energía y la masa. El de la relatividad propiamente dicho, titulado «Sobre la electrodinámica de los cuerpos en movimiento», estaba construido en torno a dos simples principios de partida y efectuaba predicciones basándose en ellos. Evitando las premisas detalladas

sobre el modo en que los objetos concretos están hechos o interaccionan, la teoría de Einstein apenas se parecía a los trabajos de los grandes físicos de la época. Por el contrario, tenía el estilo propio de un profano, o quizá poseía una rara simplicidad.

Como en su teoría física ideal, la termodinámica, Einstein deseaba por encima de todo empezar desde unos *principios*. En la termodinámica, todo el edificio descansa sobre dos pilares: la conservación de la energía y la entropía, siempre creciente, del mundo. En la relatividad, Einstein tenía en mente otros dos principios fundacionales. En primer lugar, según Einstein, el viejo punto de partida de la física clásica debía ser válido también para la electricidad y el magnetismo. Desde Galileo, los físicos habían aceptado la proposición de que si alguien se halla dentro de un compartimiento cerrado en movimiento constante, no dispone de medio mecánico alguno que le permita determinar si su movimiento es «real» o no (Galileo imaginaba al observador en la bodega de un barco navegando suavemente por mar abierto; Einstein eligió un tren, desplazándose a lo largo de sus pulidos raíles, como escenario de sus experimentos mentales). El insistente mensaje de Einstein era que Galileo aún tenía algo que decirnos. En el interior del camarote sin ventanas de un barco que se mueve suavemente, ni viendo a un pez nadar en una pecera, ni dejando caer una pelota, ni realizando cualquier otro experimento mecánico seríamos capaces de detectar si nos movemos «realmente». Del mismo modo, añadía Einstein, si nos hallamos en un tren que se desplaza

uniformemente, ningún experimento con electricidad, magnetismo o luz nos puede revelar quién se encuentra «en verdadero reposo». Se trata del principio de relatividad.

El segundo punto de partida resultaba *a priori* —como el propio Einstein admitía— un tanto sorprendente: en el seno de un marco de referencia inercial (que no esté sometido a aceleración), la luz viaja siempre a la misma velocidad, independientemente de la velocidad de su fuente. Si en una estación de ferrocarril medimos la velocidad de la luz que emite el faro ubicado en lo alto de una máquina detenida en ella, obtenemos el valor de 300.000 kilómetros por segundo. Imaginemos ahora un tren que atraviese la estación a la mitad de la velocidad de la luz, es decir, a 150.000 kilómetros por segundo. En la física clásica ordinaria, una pelota lanzada hacia delante desde el tren en movimiento cruzaría la estación a la velocidad del tren *más* la que el lanzador le imprimiera a la pelota. Sorprendentemente, según Einstein, este razonamiento no es válido para la luz. Desde la estación, comprobaríamos que la luz procedente del faro situado en ese tren ultrarrápido nos llega a 300.000 kilómetros por segundo y no más deprisa. Más aún; aplicando el primer principio (de relatividad), si persiguiéramos a un rayo de luz que se alejase de nosotros, no lograríamos siquiera mantener constante su ventaja. Independientemente del marco inercial de referencia y de la velocidad de la fuente, la medición siempre arrojaría el mismo resultado para la velocidad de la luz: la

constante c . He aquí el segundo principio: la velocidad de la luz es absoluta.

Partiendo de estas dos premisas formuladas de forma tan sencilla, la de la equivalencia física de los marcos de referencia inerciales y la del carácter absoluto de la velocidad de la luz, Einstein transformó la ciencia para siempre. Durante el proceso, las nociones de espacio y tiempo que habían sido los cimientos de la física desde Newton resultaron trastocadas por completo. Tras redactar su artículo en mayo de 1905, Einstein comenzó a meditar de inmediato sobre algunas consecuencias de la nueva física. Cierta viernes del verano de 1905 escribía desde Berna a su amigo Conrad Habicht:

«Me habría encantado haberte tenido aquí. Hubieras sido otra vez el viejo bromista de siempre. Mi tiempo no tiene demasiado valor estos días; no siempre dispongo de asuntos que merezca la pena meditar. O, al menos, que sean lo suficientemente atractivos. [...] Ronda por mi cabeza una consecuencia del estudio de la electrodinámica. El principio de relatividad, unido a las ecuaciones fundamentales de Maxwell, exige que la masa sea una medida directa de la energía que contiene un cuerpo; la propia luz transporta masa con ella. En el caso del radio tendría lugar una reducción de masa perceptible. La idea es a la vez seductora y divertida; pero tal vez Dios todopoderoso me haya engatusado con ella y se esté ahora mismo riendo a mi costa».

Convencido, obviamente, de que no estaba haciéndole reír a Dios, en septiembre de 1905 Einstein redactó su artículo de tres páginas

sobre $E = mc^2$, titulado «¿Depende la inercia de un cuerpo de la energía que contiene?»; *Annalen der Physik* lo recibió el día 27 del citado mes.

Antes de Einstein ya existía una gran controversia acerca de la relación entre la masa y la energía electromagnética. De hecho, algunos de los principales físicos de la época trataban de explicar la masa inercial (la resistencia de la materia a ser puesta en movimiento) como la existencia de partículas cargadas que, al reaccionar frente a sus propios campos eléctricos y magnéticos, se oponían a la aceleración. Einstein nunca suscribió un programa reduccionista así, es decir, una línea de razonamiento que intentase demostrar que todo, incluso la inercia, era en última instancia sólo carga y campos eléctricos y magnéticos. También se hallaba bien establecido que un contenedor de energía electromagnética (p. ej., una caja formada por espejos, llena de luz) tenía una masa que crecía en proporción a la energía electromagnética almacenada.

Pero Einstein iba mucho más allá. No contento con analizar la luz, argumentaba que *toda* forma de energía tenía asociada una masa inercial. Como era de esperar, su artículo del $E = mc^2$ hizo estallar la polémica. Uno de los aliados de Einstein, Max Planck —uno de los físicos teóricos alemanes más importantes— se apresuró a indicar que una transferencia de calor también añadía masa. Según esto, una sartén caliente pesaba más que la misma sartén cuando estaba fría. Era una novedad: nada en la física de Newton llevaba a

sospechar que la masa pudiera variar por exclusivo efecto de la energía.

Cuando Johannes Stark, un físico muy conocido que más tarde se convertiría en un ardiente nazi, vio los resultados de Planck y Einstein, no dudó en atribuir a Planck el descubrimiento de la equivalencia. Era demasiado para el joven Einstein (que aún no había desarrollado su vena diplomática): «Me sorprende que usted no reconozca mi prioridad sobre la conexión entre masa inercial y energía». Stark rectificó enseguida: «Está muy equivocado, estimado colega, si cree que no le he hecho suficiente justicia a sus artículos. Le apoyo siempre que tengo ocasión y desearía tener la oportunidad de proponerle muy pronto para una cátedra teórica en Alemania». A lo cual un apaciguado Einstein replicó, arrepentido, que «un impulso mezquino me ha llevado a hacer ese comentario sobre prioridades. [...] Quien ha tenido el privilegio de contribuir al progreso de la ciencia no debería permitir que el placer por los frutos obtenidos en una labor conjunta lo enturbien semejantes asuntos».

En los años que siguieron a 1905, Einstein trabajó intensamente para generalizar el resultado, para demostrar que la equivalencia entre masa y energía era verdaderamente completa. Obligado una y otra vez a regresar a su famosa fórmula, llegaría a ofrecer tres maneras de deducirla. En la primera, la del artículo original de 1905, Einstein imaginaba un cuerpo que emitía un mismo destello de luz en dos direcciones opuestas. Seguidamente recordaba, según

la teoría especial de la relatividad, cómo cabía contemplar la misma situación desde un sistema de referencia no acelerado diferente. Combinando los resultados se podía deducir $E = mc^2$, pero para hacerlo de forma correcta era necesario precisamente observar cómo se transformaba la energía de un marco a otro. Veintinueve años después, en una conferencia dada en Pittsburgh, Einstein presentó un argumento distinto para $E = mc^2$, utilizando esta vez el hecho de que la energía y el momento debían conservarse en todos los marcos inerciales de referencia. Pero su explicación más simple fue la tercera, que ocupaba una sola página. En 1946, Einstein escribió una demostración de $E = mc^2$ para el *Technion Journal* que no hacía uso de la teoría de la relatividad, sino tan sólo de unas premisas básicas. Examinemos este último método y detengámonos a analizar el razonamiento de Einstein.

Supongamos, como Einstein sugiere, que aceptamos estos cuatro principios:

1. Que el principio de relatividad especial es correcto, es decir, que todos los marcos de referencia no acelerados son equivalentes. Ningún marco de referencia se halla en el «verdadero» reposo y sólo los movimientos *relativos* tienen significado físico.
2. Que el momento se conserva; al fin y al cabo, este principio es un artículo de fe incluso en la física clásica. Para la materia ordinaria, el momento es igual a la masa multiplicada por la velocidad. La conservación del momento significa que si, por

- ejemplo, sumamos los momentos de todas las bolas en una mesa de billar antes de que choquen unas contra otras y repetimos el cálculo tras la colisión, el resultado será el mismo.
3. Que la radiación posee un momento; se trata de un hecho verificado experimentalmente y aceptado desde hace mucho. (Es sabido, por ejemplo, que lo que empuja hacia el exterior la cola de los cometas es la luz del Sol).
 4. Que un observador en movimiento ve una fuente de luz como si ésta sufriera un cambio en su ángulo aparente («aberración estelar»). En otras palabras, se conocía desde hacía mucho que un observador en la Tierra, por ejemplo, ve la luz de las estrellas como si procediera de un punto desplazado un pequeño ángulo α respecto a la posición verdadera de la estrella en el cielo. Ese ángulo dependía de la velocidad de la Tierra, v , y se aceptaba generalmente que, para velocidades pequeñas comparadas con la de la luz, c , $\alpha = v/c$ aproximadamente. El efecto es fácil de comprender. Aunque la lluvia caiga perpendicular al suelo, si corremos a través de ella la percibimos como si cayera hacia nosotros con un cierto ángulo. Cuanto más rápido corramos, mayor será la «aberración» de la lluvia respecto a la vertical. Si, mientras lo hacemos, transportamos un «telescopio» consistente en un largo tubo de cartón, tendremos que inclinarlo un ángulo para que las gotas de lluvia atraviesen en línea recta el tubo. Del mismo modo, debido al movimiento de la Tierra, los telescopios

ópticos tienen que ser apuntados con un cierto ángulo respecto a la posición «verdadera» de la estrella para observar su luz.

Supongamos también, añadía Einstein, que tenemos un marco de referencia —el «marco en reposo»—, el cual podríamos imaginar como un transbordador espacial que flota en el espacio profundo con los motores apagados, lejos de cualquier objeto —como una estrella o un planeta— que pudiera ejercer una fuerza gravitatoria significativa sobre él (*figura 2.1*). En este marco, un libro flota inmóvil en medio del transbordador, antes de que dos lámparas de flash, ubicadas a idéntica distancia en los extremos opuestos, envíen un destello luminoso de energía $E/2$ directamente hacia el libro. La luz procedente de ambos flashes es absorbida entonces por el libro, que ve incrementarse su energía en E .

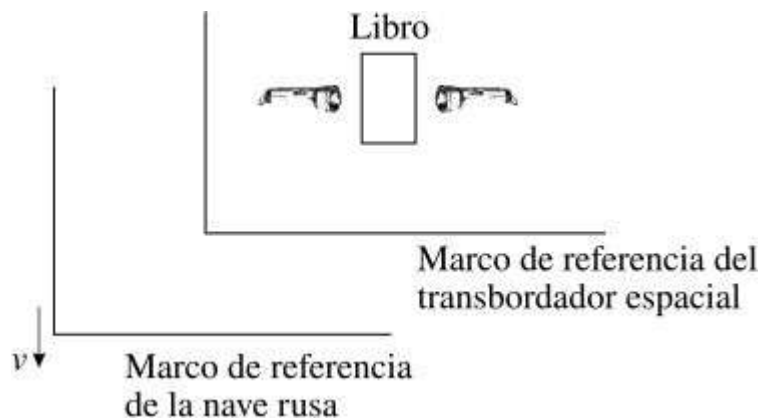


Figura 2.1

En el marco de referencia «en reposo», el libro no se mueve hacia ninguna parte, ya que ha recibido dos impactos luminosos de la misma magnitud procedentes de direcciones opuestas.

Ahora, continuaba Einstein, observemos el mismo proceso desde un marco de referencia «en movimiento» (p. ej., una nave espacial rusa) que se desplazara uniformemente hacia abajo con velocidad v . Vista desde este marco, la escena resulta ligeramente distinta. Contemplado desde la nave espacial rusa, antes de ser alcanzado por la luz de las dos lámparas, nuestro valioso libro se estaba moviendo hacia arriba con una velocidad v (*figura 2.2*). Esto quiere decir que, en el nuevo marco de referencia, antes de que los haces de luz alcancen el libro, de masa M , el momento del citado libro vale Mv . La teoría clásica de la luz nos dice que el momento de un destello de luz de energía $E/2$ vale exactamente $E/2c$. Por otra parte, en el marco de la nave rusa, la luz procedente de los flashes parece no viajar horizontalmente, sino (por efecto de la aberración) llegar con un pequeño ángulo, $\alpha = v/c$, respecto a la horizontal.

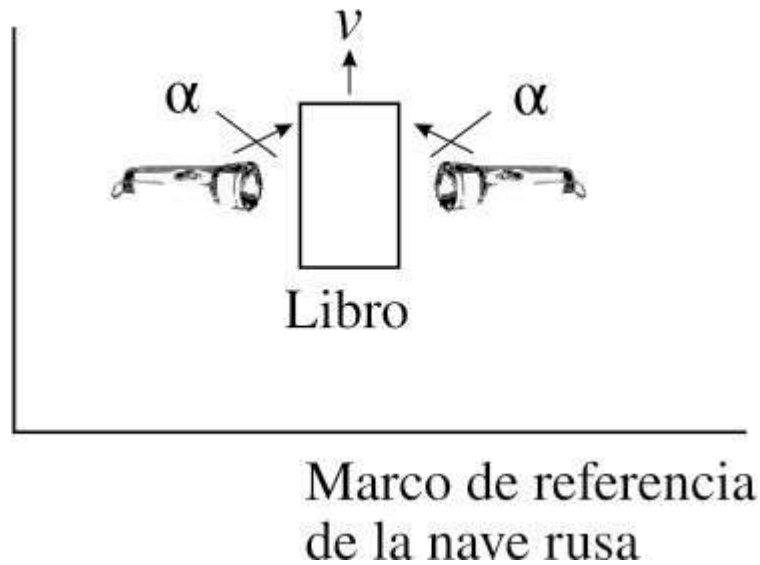


Figura 2.2.

En el marco de la nave espacial rusa, el momento del libro *tras* ser alcanzado por la luz de los flashes es igual a la suma del momento original ascendente (Mv) más el momento que el libro recibe del impacto de los dos haces de luz, que en este marco de referencia llegan con un «ángulo de aberración».³⁶ En consecuencia, los haces de luz contribuyen con un momento de magnitud Ev/c^2 al momento inicial del libro, que ya valía Mv . Tras la absorción de la luz, el momento total del libro en el marco de referencia de la nave rusa vale $Mv + Ev/c^2$.

³⁶ En el marco de la nave espacial rusa, los destellos parecen no desplazarse a lo largo de la horizontal sino, debido a la aberración, adoptar un pequeño ángulo de llegada, $\alpha = v/c$, respecto a esa horizontal. Debemos pensar, por lo tanto, que el momento de la luz tiene una componente horizontal y otra vertical. Las componentes horizontales de los dos destellos se cancelan mutuamente, ya que tienen la misma magnitud y sentidos opuestos. Pero ambos flashes poseen una pequeña componente vertical que vale $(E/2c) \sin \alpha$. Como α es pequeño, podemos hacer $\sin \alpha = \alpha$, con lo que la componente vertical del momento sería $(E/2c)\alpha$. Así pues, por efecto de la aberración, el momento vertical viene dado por:

$$\text{componente vertical del momento para cada haz de luz} = (E/2c)(v/c) = Ev/2c^2$$

Aunque el momento del libro se ha incrementado, su velocidad final hacia arriba sigue siendo v , la misma, en valor absoluto, que la de la nave rusa. (La velocidad del libro en el marco de referencia de la nave rusa ha de continuar siendo v necesariamente: en el marco del transbordador espacial, los haces de luz inciden en direcciones opuestas, con lo que el libro sigue estacionario; así pues, tras la absorción, el libro sigue moviéndose a velocidad v respecto al primero). Por lo tanto, según Einstein, la absorción de energía ha tenido que incrementar la masa del libro —ya que la velocidad del libro no aumenta, es la *única* manera de justificar que el momento sea mayor—. Si llamamos M' a la masa final del libro, en el marco de referencia de la nave rusa:

$$\text{Momento final del libro} = Mv + Ev/c^2 = M'v$$

Dividiendo la ecuación por v y restando M a ambos lados se obtiene: $M' - M = E/c^2$, que es lo mismo que decir que $E = (M' - M)c^2$. Si expresamos abreviadamente $M' - M$, la diferencia entre la masa del libro antes y después de la llegada de los destellos de luz, como m , masa adquirida, obtenemos el objeto de nuestros deseos:

$$E = mc^2$$

Ahora bien, dado que toda forma de energía siempre puede ser convertida en otra, el resultado no es solamente aplicable a los haces de luz. Por el contrario, significa que cualquier forma de energía se añade a la masa inercial: una bola de billar caliente tiene

más masa que una fría y un planeta en rotación es más masivo que uno que estuviera inmóvil. De hecho, si a la masa se le permite convertirse en energía, lo hará. ¿Qué es lo que pone límite a estos cambios? Las leyes de conservación: una ley de conservación es una afirmación de que ciertas magnitudes no cambian en un sistema cerrado; por ejemplo, no podemos crear una carga eléctrica de la nada. El momento —la tendencia de un cuerpo a moverse en línea recta una vez se ha puesto en marcha— permanece constante, salvo que apliquemos una fuerza. Debido a estas leyes de conservación, en la teoría de la relatividad un único electrón no puede desvanecerse, transformándose en pura energía, ya que esto alteraría la carga eléctrica del universo. Ahora bien, si un electrón choca con un *anti*-electrón (que tiene la carga opuesta), la historia es muy diferente. En este caso, la suma de las cargas es cero; para la masa conjunta del electrón y el positrón *sí* resulta posible la transformación total en energía. En sentido inverso y respetando siempre las leyes de conservación, la energía pura también puede convertirse en masa (p. ej., en un electrón más un positrón).

En las décadas que siguieron a 1905, $E = mc^2$ desembarcó en los laboratorios. En 1932, dos físicos del famoso Laboratorio Cavendish de Cambridge, los experimentalistas John Cockroft y Ernest Walton, lograron acelerar protones para desintegrar un núcleo de litio. Observaron que los fragmentos del núcleo resultantes pesaban menos que el núcleo de litio original. Al principio parecía que parte de la masa se hubiera desvanecido. Pero, midiendo la energía total

de los fragmentos desprendidos y empleando $E = mc^2$, los científicos de Cambridge hallaron que la masa «perdida» en la desintegración equivalía a la energía adquirida por los veloces fragmentos desprendidos del núcleo. La fórmula de Einstein acertaba de nuevo. Pero la aplicación de $E = mc^2$ que cambiaría la faz del mundo llegó con el descubrimiento de que los neutrones podían causar la fisión nuclear del uranio. Durante años, la física Lise Meitner había estado trabajando con el químico Otto Hahn en el Instituto Químico Káiser Guillermo.³⁷ En el frondoso suburbio berlinés de Dahlem, se dedicaban a bombardear núcleos con neutrones, utilizando la química para clasificar los productos resultantes. Tanto ellos como muchos otros —incluyendo el grupo de Enrico Fermi en Roma— pensaban que los productos de la reacción que observaban tras el bombardeo eran realmente nuevos elementos, situados más allá del uranio en la tabla periódica. Esos «transuránidos», como los denominaron, eran sensacionales, tal vez el mayor descubrimiento de la nueva *radioalquimia*. En la pareja de Berlín, las habilidades de sus miembros resultaban complementarias: Meitner era la físico del equipo y Hahn, el químico. Pero el buen entendimiento terminó abruptamente cuando los nazis cerraron el laboratorio y Meitner, que era judía, vio de pronto su destino pendiente de un hilo. Tras escapar de Alemania por ferrocarril el 13 de julio de 1938, Meitner inició una gris carrera científica en Suecia, desde donde

³⁷ La mejor fuente de información sobre Lise Meitner es la excelente biografía escrita por Ruth Lewin Sime, *Lise Meitner: A Life in Physics*, University of California Press, 1996; véanse especialmente las págs. 233-237.

ansiosamente aguardaba noticias de sus colaboradores mientras el mundo avanzaba hacia el abismo de la guerra.

En Berlín, los resultados de laboratorio sólo traían más confusión a Hahn, que seguía adelante con los experimentos. Desde hacía mucho tiempo, él y Meitner estaban acostumbrados a detectar productos de la colisión que, en algunas reacciones, se comportaban como elementos mucho más ligeros que el uranio. Pero eso —Hahn y todos los demás estaban plenamente convencidos— era una mera ilusión química, algo imposible: los elementos tenían que estar cerca del uranio en la tabla periódica. «Romper» un núcleo en partes mucho más pequeñas era, sencillamente, imposible. Cabía desprender, quizás, un protón o una partícula alfa (dos protones enlazados a dos neutrones); pero, tal como diría después cierto físico, romper netamente un núcleo en dos partes era como hacer estallar una casa lanzándole una pelota por la ventana. Si un producto de reacción parecía bario, por ejemplo, debía de ser radio, de características químicas similares. Los resultados eran cada vez más desconcertantes, hasta que una noche de diciembre de 1938 Hahn escribía a Meitner:

«19/12/38. Noche del lunes en el laboratorio. Querida Lise. [...] Son ahora las once en punto; a las 23:45 llegará Strassmann [su otro colaborador], con lo que finalmente podré volver a casa. Hemos descubierto algo sobre los "radioisótopos" tan notable que, por ahora, sólo te lo vamos a contar a ti. [...] Nuestros isótopos de Ra[dio] se comportan como si fueran Ba[rio]».

«Así que, por favor», rogaba Hahn, «piensa si hay alguna posibilidad de que exista una variedad de bario que sea mucho más pesada que la normal».

Hahn envió su artículo al editor tres días después de escribir a Meitner. En él se reflejaba el mar de dudas en el que se hallaban Hahn y su colega Strassmann. Detectaban lo que parecían ser elementos ligeros conocidos, pero no podían aceptarlo: «Como químicos [...] deberíamos sustituir los símbolos [de los elementos ligeros] por los [de los elementos pesados] que hemos citado. Como “químicos nucleares”, próximos también a la física, aún no podemos dar ese paso que contradice toda la experiencia anterior en física nuclear».

Cuando le llegó la carta del 19 de diciembre, Meitner y su sobrino, el físico Otto Robert Frisch, que también se había exiliado, salieron a dar un paseo en la nieve y comenzaron a analizar la misteriosa carta. ¿Qué sucedería —se preguntaron— si el núcleo de uranio, al ser alcanzado por un neutrón, comenzara a oscilar como una gruesa gota de agua? Esta imagen del núcleo había sido relativamente común en aquella época. Supongamos, continuaron, que la gota se hallara normalmente en delicado equilibrio, con los noventa y dos protones repeliéndose furiosamente unos a otros, pero mantenidos juntos por ciertas fuerzas de atracción, potentes aunque de corto alcance, entre los doscientos treinta y ocho protones y neutrones. Podría ocurrir que la gota se dilatara al oscilar, tal vez hasta un punto en el que pareciera una haltera

viscosa, un par de esferas unidas por un delgado vástago nuclear. En ese estado, el efecto de la repulsión mutua de los protones situados en las dos esferas podría ser mayor que la atracción debida a las uniones de corto alcance. En un momento dado, la repulsión eléctrica podría hacer que el núcleo se dividiera en dos, con lo que las esferas se separarían una de otra bajo la fuerza debida a los dos grupos de unos 46 protones que contienen. Meitner calculó: dos núcleos ligeros pesarían menos que el núcleo pesado original. Y la energía debida a esa diferencia de masa, de acuerdo con $E = mc^2$, sería enorme. Ella y su sobrino acababan de saber algo que nadie en el mundo sospechaba: que en Dahlem se producía la fisión nuclear.

Los acontecimientos se sucedieron deprisa. Al escuchar la interpretación de Meitner y Frisch, el físico danés Niels Bohr, considerado el padre de la teoría cuántica por muchos de sus colegas, comprendió de inmediato dónde estaba el error en sus anteriores razonamientos. Wheeler, que zarpó hacia América con Bohr en 1939, le ayudó a componer un análisis teórico exhaustivo de la fisión en medio del Atlántico. No transcurriría demasiado tiempo hasta que la idea de la desintegración del átomo saltara del laboratorio a los titulares de los periódicos. Y la inmediata cuestión, crucial en un mundo tan inestable como el de aquellos años, no tardó en plantearse: ¿Podían los neutrones que salían despedidos en la división del núcleo causar fisiones adicionales? Si la respuesta era afirmativa, la gigantesca energía liberada por la fisión crecería

geométricamente. En pocos meses, varios físicos empezaron a sospechar que el mecanismo de la fisión podría llevar, en un futuro no muy lejano, a la fabricación de bombas nucleares. Algunos de ellos animaron a Einstein a escribir su famosa y decisiva carta del 2 de agosto de 1939 al presidente Roosevelt:

«En el transcurso de los cuatro últimos meses se ha convertido en probable —con los trabajos de Joliot en Francia y de Fermi y Szilárd en Estados Unidos— la posibilidad de desencadenar reacciones nucleares en cadena en una gran masa de uranio, mediante las cuales se generarían enormes cantidades de energía y nuevos elementos similares al radio. Hoy parece casi seguro que esto podría lograrse en un futuro inmediato. En cualquier caso, se trata de algo más que de una mera producción de energía. Este fenómeno nuevo podría también llevar a la construcción de bombas y es concebible —aunque mucho menos seguro— que puedan fabricarse bombas de un nuevo tipo, extremadamente potentes. Una sola bomba de esta clase, transportada por barco y explosionada en un puerto, destruiría por completo tanto ese puerto como una parte del territorio de alrededor».

Einstein insistía en la necesidad de establecer contactos entre la Administración y los físicos. A modo de siniestro augurio, Alemania había detenido la venta de uranio. El 1 de octubre de 1939, un

representante de los científicos mantuvo una entrevista con Roosevelt, en la que los partidarios del átomo avalaron sus propuestas con un memorándum más técnico del refugiado húngaro Leo Szilárd, el descubridor de la reacción nuclear en cadena. Por entonces, los nazis habían invadido Polonia y la bola de nieve comenzaba a rodar ladera abajo. Se temía que los alemanes poseyeran una bomba atómica; Pearl Harbor fue atacado y los británicos daban los primeros pasos hacia la construcción de un arma nuclear. En Estados Unidos, las comisiones se convirtieron en laboratorios y los laboratorios en las fábricas más grandes jamás vistas. Algunos años después, al recordar esos días, Einstein reflexionaba sobre los aspectos éticos de aquello a lo que él mismo había contribuido a poner en marcha, primero con las especulaciones de un joven empleado en la oficina de patentes y, más tarde, siendo el más famoso científico del mundo:

«Cometí un error cuando firmé aquella carta al presidente Roosevelt dando a entender que la bomba atómica debía ser construida. Pero tal vez se me pueda perdonar por ello, porque entonces todos pensábamos que había una alta probabilidad de que los alemanes estuvieran trabajando en el tema y de que llegaran a tener éxito y utilizaran la bomba atómica para convertirse en la raza dominante».

Cuando a Einstein le preguntaban por qué, en su opinión, había sido posible descubrir los átomos, pero no la forma de controlarlos,

respondía: «*Muy sencillo, amigo mío: porque la política es más difícil que la física*».

Al terminar la guerra, cuando John Wheeler realizaba la exposición con que iniciábamos este capítulo, $E = mc^2$ se había convertido para los físicos en el símbolo de la era atómica. Una era cuyo advenimiento se celebraba por haber forzado el fin de la guerra y, a la vez, se lamentaba por ser el punto de partida de una carrera armamentista de imprevisibles consecuencias. La ecuación era, a la vez, una guía hacia el futuro y el recordatorio de un gigantesco error.

Tras la segunda guerra mundial, $E = mc^2$ se hizo omnipresente, lejos del control de los físicos. Una pequeña firma incluso la adoptó como nombre comercial: «No consiste en trabajar más, sino en hacerlo de forma más inteligente», añadía como lema. «Con la imagen de Einstein repartida por toda la oficina como *mascota* del negocio, sería difícil hacerlo de otra manera».³⁸ $E = mc^2$ es también el nombre de un refresco, el de un campamento científico juvenil en Texas y el *logo* de un consorcio de escuelas de distrito en Nueva Jersey cuyo objetivo es mejorar la enseñanza de la ciencia. La ecuación es, asimismo, el título de un *best-seller* de Patrick Cauvin ($E = mc^2$, *mon amour*), una historia de amor entre dos precoces genios de once años que vuelan a Venecia. Y, por supuesto, hay toda clase de pósters con la foto de Einstein adornada con su emblemática ecuación.

³⁸ Véase: www.bizjournals.com/stlouis/stories/1998/11/09/smallbl.html

No parece que $E = mc^2$ sea el título más adecuado para una pieza musical, pero Big Audio Dynamite no opinó así y, hasta donde conozco, hay al menos otros diez grupos de rock que usaron la ecuación como título en sus canciones. Hay una película — disponible en vídeo— que también lo lleva; su argumento: «Un profesor de física de Oxford intenta llegar más lejos que Einstein mientras trata de satisfacer las demandas de su mujer y las de su amante. ¡Todo sobre la fisión nuclear!». Existe una empresa gráfica japonesa denominada $E = mc^2$ y también ostentan este nombre una compañía de sistemas de Internet francesa, grupos de estudios de Arizona e instalaciones artísticas de varios países. En todas partes es el símbolo del genio, un signo de poder y, a la vez, el heraldo de la destrucción.

Quizá no deberíamos sorprendernos. A diferencia de otras ecuaciones físicas, $E = mc^2$ conecta con la cultura popular por cuatro razones. En primer lugar, la ecuación en sí es compacta, fácil de escribir y dramática en sus consecuencias, tanto en el laboratorio como para el mundo. La ecuación de Einstein para el campo gravitatorio, por el contrario, es casi impronunciable para el hombre de la calle: $R_{ab} - \frac{1}{2}Rg_{ab} = -8\pi GT_{ab}$, *mon amour* no tendría el mismo atractivo comercial y me temo que tampoco resultaría como título de una canción de rock, aunque los físicos opinen que la ecuación que gobierna la relatividad general merece aún más respeto que la equivalencia entre masa y energía.

En segundo lugar, la ecuación $E = mc^2$ plasmaba, al menos en parte, la extraordinaria fascinación que en la cultura popular ejercían las nuevas ideas sobre el espacio y el tiempo debidas a la teoría de la relatividad. Antes, incluso de que ésta fuera formulada, al pintor Claude Monet le habían atraído los conceptos de simultaneidad, velocidad y tiempo y la alteración del espacio. Cuando la física proporcionó un nuevo marco espaciotemporal no euclidiano y la fusión de temporalidad y espacialidad, esos conceptos, o al menos sus equivalentes metafóricos, hallaron un suelo fértil.

En tercer lugar, tras la expedición del astrónomo británico Arthur Eddington para observar el eclipse de 1919, que demostró que la teoría einsteniana predecía correctamente la curvatura de la luz solar, Einstein se convirtió en una figura de culto indiscutible (al menos, para sus admiradores), el prototipo de genio individual. Pacifista en los años previos a la guerra y conciliador después, ejemplo ético, incomprendido, vilipendiado y alabado después sin medida, Einstein se transformó en un símbolo de esperanza para todo aquel que fuera contracorriente. Para sus enemigos, por supuesto, era el antihéroe: cosmopolita, antinacionalista, judío, teórico abstracto, demócrata y ajeno a todo sentimiento de raza o de nación. Incluso con anterioridad a la segunda guerra mundial, Einstein y, con él, sus ecuaciones más famosas, arrastraban una mezcla de filosofía, física y modernidad que unas veces seducía y otras tantas aterraba a quienes le rodeaban.

En el largo periodo de tensión comprendido entre 1939 y 1989, primero de guerra real y luego de guerra fría, la ecuación llegó a implicar algo más —las armas nucleares—, encerrando en sus escasos símbolos tanto el poder como el conocimiento. La «ecuación del sextante» adquiriría así un cuarto significado, pues esas armas parecían conjugar el conocimiento más esotérico con el más terrible poder destructivo. La ecuación se asociaba de pronto a una fuerza casi mística, a la expresión más sintética y tangible del Apocalipsis. Todas esas corrientes culturales, todos esos conceptos parecen hoy orbitar en torno a la ecuación. A la vez fantasía filosófica y genial, física práctica y arma aterradora, $E = mc^2$ se ha convertido en sinónimo del conocimiento técnico por antonomasia. Nuestras ambiciones científicas, nuestras ansias de conocer y nuestras peores pesadillas están encerradas en esos breves trazos de la pluma.

Lecturas recomendadas

La edición más exhaustiva y erudita de la obra de Einstein es, con diferencia, *Einstein's Collected Papers* (Princeton University Press); para el periodo en torno a 1905, véase el volumen 2 y la traducción al inglés de ese volumen realizada por Anna Beck.

La mejor biografía científica de Einstein es *Subtle is the Lord...: Science and the Life of Albert Einstein*, de A. Pais (Oxford University Press, 1982). [Trad, esp.: *El Señor es sutil...*, Barcelona, Ariel, 1984. Traducido por F. Alsina].

Gerald Holton, *Thematic Origins of Scientific Thought*, Harvard University Press, 1973. En este libro hallamos excelentes ensayos sobre la historia cultural y científica de la obra de Einstein.

Una biografía general muy útil: Albrecht Fölsing, *Albert Einstein: A biography*, Nueva York, Viking, 1997.

Una colección de excelentes ensayos desde distintas perspectivas históricas sobre Einstein: Peter Galison, Michael Gordin y David Kaiser, *Einstein's Relativities*, Routledge [en prensa].

A. I. Miller, *Albert Einstein's Special Theory of Relativity: Emergence (1905) and Early interpretation (1905-1911)*, Reading, Addison-Wesley, 1981. Esta obra constituye una historia técnica y muy útil de la teoría especial de la relatividad.

Sobre Wheeler y la bomba H: Peter Galison, *Image and Logic: A Material Culture of Microphysics*, University of Chicago Press, 1997.

Para situar a Einstein respecto a la tradición de la electrodinámica, véase: Oliver Darrigol, *Electrodynamics from Ampère to Einstein*, Clarendon Press, 2000.

Roger Penrose y John Stachel, *Einstein's Miraculous Year: Five Papers that Changed the Face of Physics*, Princeton University Press, 1998. Se trata de un breve pero útil volumen que contiene la traducción al inglés de los cinco artículos de 1905. [Trad, esp.: *Einstein 1905: Un año milagroso*, Barcelona, Crítica, 2001].

No debemos olvidar la exposición que el propio Einstein realizó de su teoría con destino al público en general en *Relativity: The Special and the General Theory*, Nueva York, Crown Publishers, 1961.

Albert Einstein, *Ideas and Opinions*, Nueva York, Bonanza Books, 1954. Contiene ensayos sobre toda clase de temas, desde la política hasta la filosofía de la física. [Trad, esp.: *Mis ideas y opiniones*, Barcelona, Editorial Bon Ton, 2000].

Por último, mi libro de texto elemental favorito sigue siendo, a pesar de los años transcurridos, *Special Relativity*, de A. P. French, perteneciente a la MIT Introductory Physics Series (Nueva York, W. W. Norton, 1968).

Parte 3

El redescubrimiento de la gravedad

La ecuación de Einstein para la relatividad general

Roger Penrose

Introducción

La teoría general de la relatividad revolucionó por completo nuestra forma de ver el mundo físico. Sin embargo, no surgió de las observaciones hechas por experimentadores en laboratorios. Fue simplemente el producto de la perspicacia e imaginación de un teórico. Se trató, por lo tanto, de una revolución en marcado contraste con la imagen convencional del modo en que una revolución científica suele tener lugar. Según esa imagen, sólo abandonamos un punto de vista científico previamente aceptado cuando hay una acumulación suficiente de datos experimentales en contradicción con él. El siglo XX, de hecho, contempló varias revoluciones trascendentales en física fundamental, cada una de las cuales obligó a revisar cuidadosamente principios básicos e hizo añicos puntos de vista previos sobre la naturaleza de la realidad física. En general, todas se ajustaban a la imagen convencional citada. Como veremos, el caso de la relatividad general fue muy diferente.

En términos generales, en la física del siglo XX hubo dos revoluciones fundamentales completamente distintas. La primera fue la relatividad, sobre la naturaleza del espacio y el tiempo, y la

segunda, la teoría cuántica, sobre la naturaleza de la materia. No obstante, podría decirse que la teoría de la relatividad por sí sola produjo *dos* revoluciones, la de la «relatividad especial» y la de la «relatividad general».

La relatividad especial se ocupa de los extraños retoques que hay que hacerle a la física newtoniana cuando los cuerpos viajan a velocidades próximas a la de la luz, circunstancia en la que el espacio y el tiempo se coordinan misteriosamente, transformándose uno en otro y llevando a la noción conjunta de *espacio-tiempo*. La teoría nació, en esencia, de las observaciones en conflicto con la idea de un «éter» omnipresente y equivalente a un estado absoluto de reposo. El conflicto más famoso con dicha idea lo constituía el experimento de Michelson-Morley (1887), que había tratado de medir, sin éxito, la velocidad de la Tierra respecto al éter. Ese y otros experimentos hacían cada vez más difícil sostener el punto de vista newtoniano sobre el espacio y el tiempo. La revolución encarnada por la relatividad especial avanzó inexorablemente gracias al trabajo de varios científicos: George Fitzgerald, Joseph Larmor, Hendrik Lorentz, Henri Poincaré, Albert Einstein y Hermann Minkowski. Según esto, creo que debería ser considerada una revolución de tipo «convencional», en la que hubo experimentos que llevaron a los teóricos a apartarse del planteamiento newtoniano (aunque el camino seguido por Einstein no estuviera precisamente basado en los experimentos).

La teoría cuántica, por su parte, tuvo también una base experimental. De hecho, la tuvo en mucho mayor grado que la relatividad especial. Los físicos se vieron obligados a introducir esa nueva teoría a fin de explicar el comportamiento de la materia a escala muy pequeña ante la avalancha de datos experimentales en abierta contradicción con las ideas de Newton.

Sin embargo, la teoría *general* de la relatividad, con su descripción de la gravedad como efecto de la «curvatura del espacio-tiempo» en vez de la *fuerza* gravitatoria newtoniana, pareció habersele revelado a Einstein sin que existiera necesidad alguna de un nuevo enfoque tan revolucionario. A principios del siglo XX, la elegante teoría newtoniana de la gravitación universal, actuando según una ley cuadrática inversa para las fuerzas entre partículas, concordaba maravillosamente con las observaciones con un margen de error inferior a una parte en diez millones. Existían algunas anomalías menores pero, en última instancia, todas resultaban ser errores de observación o de cálculo o deberse a alguna influencia perturbadora que no se había tenido en cuenta. Para ser rigurosos, no todas, ya que había ciertos minúsculos detalles en relación con la órbita del planeta Mercurio que no acababan de estar justificados. De todos modos, los astrónomos de la época no estaban demasiado preocupados, pues se hallaban convencidos de que un análisis más cuidadoso de la situación resolvería algún día ese problema aparentemente insignificante en el marco de la teoría de Newton.

Desde el punto de vista experimental, por lo tanto, nada hacía prever que la teoría de Newton fuera a ser cuestionada.

Pero Einstein se había visto llevado a un concepto de gravitación muy distinto del newtoniano. No eran los datos de las observaciones lo que inquietaba a Einstein, aunque quizás esta afirmación no sea cierta del todo. En realidad, *había* un dato experimental que le servía de base, pero no provenía del siglo XX ni del XIX; ni siquiera del XVIII o del XVII. Lo que preocupaba a Einstein había quedado bien establecido por parte de Galileo a finales del siglo XVI (y ya había sido observado por otros con anterioridad) y formaba parte de la física gravitatoria universalmente aceptada. Durante más de cuatro siglos, el verdadero significado de la observación de Galileo había permanecido oculto. Pero Einstein supo verla con nuevos ojos y percibir ese significado. Y éste le condujo a la extraordinaria visión de que la gravitación era una característica intrínseca de la *geometría curva del espacio-tiempo* y a la formulación de una ecuación —hoy conocida como ecuación de Einstein— de una elegancia y una simplicidad geométrica sin precedentes. Calcular sus implicaciones presenta, no obstante, una enorme dificultad técnica y los resultados son casi siempre indistinguibles de los de Newton. En los ocasionales casos en que no es así, sin embargo, de la teoría de Einstein se deducen nuevos y notables efectos. En uno de éstos, la precisión de la fórmula einsteniana representa una mejora en un factor de diez millones sobre la teoría de Newton.

¿Cómo es ese paradigma de bella ecuación, la *ecuación de Einstein* que gobierna la relatividad general? Habitualmente se escribe así:

$$R_{ab} - \frac{1}{2} R g_{ab} = -8\pi G T_{ab}$$

Pero ¿qué significa? ¿Por qué nos parece bella esa acumulación de símbolos? Es obvio que, sin el significado que subyace bajo ellos, no existe belleza ni significado físico alguno. Antes de intentar desentrañar dicho significado, detengámonos a hacer una breve interpretación. Las variables del lado izquierdo de la ecuación se refieren a ciertas magnitudes de esa misteriosa curvatura del «espacio-tiempo»; las del derecho, a la densidad de energía de la materia. La ecuación $E = mc^2$ nos dice que la energía es esencialmente equivalente a la masa y, de modo similar, los términos del lado derecho se refieren también a la densidad de *masa*. Recordemos, asimismo, que la masa es la fuente de la gravedad. La ecuación de campo de Einstein³⁹ indica, pues, que la curvatura del espacio-tiempo (lado izquierdo) está directamente relacionada con la distribución de la masa en el universo (lado derecho).

Antes de comenzar, y dado que en los próximos párrafos van a aparecer algunas ecuaciones matemáticas, tal vez sean útiles algunas consideraciones sobre su lectura. Si al lector le intimidan,

³⁹ Actualmente se prefiere hablar de esta ecuación en singular en vez de en plural, como se hacía originalmente, pues es mejor concebirla como una única ecuación aplicable en su totalidad a los tensores involucrados (v. los comentarios sobre la geometría curva del espacio-tiempo de más adelante), en lugar de considerar las componentes de esos tensores.

le recomiendo una práctica que yo mismo sigo cuando me encuentro con una línea así. Consiste, más o menos, en ignorar la línea y saltar a la siguiente que contenga texto. Bueno, quizá se debería echarle al menos un pequeño vistazo a la ecuación antes de continuar. Más tarde, familiarizados algo más con el contexto, podemos regresar a la ecuación descartada y tratar de captar sus características más destacables. El propio texto debería ayudarnos a decidir qué es lo importante y qué cabe ignorar sin más problemas. Si no es así, no hay que tener miedo de dejarla atrás del todo.

El principio de equivalencia

Intentemos comprender lo que Einstein trataba de lograr al proponer su teoría general de la relatividad. ¿Por qué pensaba que era necesario ir más allá de la exitosa teoría de Newton? ¿Por qué introdujo la noción de curvatura espaciotemporal? ¿Qué es realmente, esa curvatura?

El principio fundamental que, en opinión de Einstein, tenía que ser incorporado al núcleo de la teoría gravitatoria es lo que él mismo denominaba *principio de equivalencia*. El ingrediente esencial de este principio ya era, en efecto, conocido por Galileo a finales del siglo XVI (y, antes que él, por Simón Stevin en 1586 y por muchos otros, hasta remontarnos a Ioannes Philiponos en el siglo V o VI). Imaginemos que dejamos caer a la vez un objeto grande y otro pequeño, cada uno de ellos de un material arbitrario, desde —por ejemplo— lo alto de la torre inclinada de Pisa. Si ignoramos los

efectos de la resistencia del aire, los dos objetos caerán con la misma velocidad y llegarán al suelo al mismo tiempo. Imaginemos una cámara de vídeo amarrada al objeto grande y enfocada hacia el pequeño. Como ambos caen exactamente a la vez, la imagen del objeto pequeño captada por la cámara será la de algo que flota inmóvil, aparentemente estacionario y, por lo tanto, ajeno a la acción de la gravedad. Para los dos objetos —hasta el momento en que alcanzan el suelo, por supuesto— la gravedad terrestre parece haberse desvanecido.

Esta observación contiene la esencia del principio de equivalencia. Al caer libremente por acción de la gravedad, se pueden eliminar sus efectos locales, con lo que, aparentemente, la fuerza gravitatoria ha desaparecido. A la inversa, es posible producir efectos indistinguibles de los de la gravedad si se toma como referencia un marco acelerado. Esta gravedad aparente debida a la aceleración es una característica típica de los modernos medios de transporte de alta velocidad. Cuando un coche acelera, sus ocupantes se ven empujados hacia el respaldo de sus asientos como si una nueva fuerza gravitatoria hubiera aparecido de repente y tirara de los pasajeros hacia atrás. De manera similar, si el conductor pisa de pronto el freno, los ocupantes parecen empujados hacia delante, como si una fuerza gravitatoria repentina tirara de ellos en ese sentido. Si el vehículo gira hacia la derecha, aparecería una fuerza gravitatoria que tiraría de los pasajeros hacia la izquierda, y así sucesivamente. Estos efectos se manifiestan de forma particular en

un avión, pues en él es a menudo difícil saber qué dirección corresponde a «abajo» —es decir, hacia el centro de la Tierra—, debido a la confusión entre las sensaciones originadas por la aceleración del avión y las de la gravedad terrestre. El principio de equivalencia dice que esa confusión es una propiedad fundamental de la gravedad. Las leyes físicas que se evidencian cuando tomamos medidas respecto a un marco de referencia acelerado son exactamente las mismas que cuando consideramos un marco *no* acelerado e introducimos el campo gravitatorio adecuado, además de las fuerzas que estuviesen ya presentes.

Hay que subrayar que esta propiedad de «equivalencia» sólo es válida para el campo gravitatorio y para ningún otro tipo de fuerza. Por ejemplo, *no* es aplicable si en lugar de un campo gravitatorio tomamos un campo eléctrico. Consideremos un escenario equivalente al propuesto anteriormente, con los objetos cayendo desde la torre inclinada, pero sometidos a fuerzas eléctricas en vez de gravitatorias. La aceleración con la que cada objeto «cae» en el seno de un campo eléctrico no es independiente en absoluto del material de que está hecho: depende de lo que se denomina relación carga/masa del objeto. Para tomar un caso extremo, imaginemos que los dos cuerpos tienen igual masa, pero sus cargas son opuestas (es decir, una tiene carga positiva y la otra, negativa). ¡Los objetos se acelerarían en el seno del campo eléctrico, pero en sentido contrario uno de otro! Una videocámara situada en uno de

ellos, desde luego que no registraría al otro como si estuviera en reposo.

Lo que diferencia a los cuerpos cargados en el seno de un campo eléctrico respecto a las masas en el seno de un campo gravitatorio es que la fuerza sobre un cuerpo cargado es proporcional a su *carga*, mientras que su resistencia al movimiento —es decir, su *inercia*— es proporcional a su *masa*. Lo que hace especial al caso gravitatorio es que tanto la fuerza sobre el cuerpo como la resistencia de éste a moverse son, *ambas*, proporcionales a la masa. Desde la perspectiva newtoniana, este hecho parece ser enteramente fortuito. La igualdad entre *masa gravitatoria* (que determina la intensidad de la fuerza gravitatoria sobre un cuerpo) y *masa inercial* (que determina la resistencia del cuerpo a cambiar su movimiento) en absoluto es un requisito esencial en una teoría dinámica de corte newtoniano pero, en el caso de la gravedad, hace las cosas algo más simples, ya que no hay dos clases de masa de las que preocuparse.

Aunque estos conceptos eran ya conocidos desde hacía mucho tiempo —básicamente, desde las consideraciones hechas por Galileo y, por supuesto, Newton era consciente de ellos—, fue Einstein el primero en darse cuenta de la profunda trascendencia *física* del principio de equivalencia. ¿En qué consiste esa trascendencia? Recordemos ante todo el desarrollo einsteniano de la relatividad *especial*. En ella, Einstein había considerado como principio fundamental el de «relatividad especial». Según este principio, las

leyes físicas son las mismas para cualquier observador que se mueva uniformemente (sin aceleración alguna). Aunque Larmor, Lorentz y Poincaré habían manejado antes que él las leyes de transformación básicas de la relatividad especial, ninguno de ellos había adoptado el punto de vista de Einstein en el sentido de que ese *principio de relatividad* fuera fundamental y, por consiguiente, respetado por todas las fuerzas de la naturaleza. El «relativismo» estricto de Einstein le llevó a cuestionar la restricción al movimiento *uniforme* del principio de relatividad. ¿Cómo percibe las leyes físicas un observador *acelerado*?

A primera vista, parece que un observador acelerado percibiría leyes *distintas* de las que constata un observador en movimiento uniforme. En lenguaje newtoniano, haría falta introducir «fuerzas ficticias» (es decir, fuerzas imaginarias) para explicar los efectos de la aceleración. Aquí es donde interviene el principio de equivalencia. Según Einstein, esas fuerzas ficticias no son ni más ni menos reales que la fuerza de gravedad que todos sentimos tirando de nosotros hacia el centro de la Tierra. Esa fuerza con la que nuestro planeta nos atrae parece desaparecer si caemos libremente en su seno. Recordemos nuestra imaginaria cámara de vídeo amarrada a uno de los objetos que caían de la torre. En el marco acelerado de la cámara, el campo gravitatorio terrestre parece haber desaparecido. Se ha convertido en «ficticio» por el simple hecho de haber referido las cosas a un marco que está en reposo respecto a la videocámara.

Bajo el enfoque einsteniano, un observador acelerado percibe las mismas leyes que el que se mueve uniformemente si se añade el *campo de fuerzas gravitatorio* adecuado, proveniente de la aceleración, a las demás fuerzas en juego. En el caso de la videocámara que cae, este campo adicional sería un campo gravitatorio dirigido hacia arriba que compensa exactamente el campo gravitatorio terrestre (dirigido hacia abajo). En el marco de referencia de la cámara, por lo tanto, el campo gravitatorio es nulo.

En una conferencia celebrada en Japón en 1922, Einstein recordaba el momento en el que tuvo la idea, a finales de 1907:

«Me hallaba sentado ante mi mesa en la oficina de patentes cuando, de repente, un pensamiento se me vino a la cabeza: "si alguien cae libremente, no siente su propio peso". Me quedé sobrecogido. Esa idea tan simple me dejó una profunda huella y fue la que me impulsó hacia una teoría de la gravitación».

En otra ocasión, Einstein se refería al hecho como *«el pensamiento más afortunado de mi vida»*, pues no en vano contenía la semilla de su maravillosa teoría general de la relatividad.

Al lector tal vez le parezca que, con su teoría, Einstein borró la gravedad de un plumazo. ¡Por supuesto que *existe* un efecto que denominamos gravedad! Los planetas se mueven *realmente* de un modo que la teoría de Newton es capaz de describir con éxito. ¡Y, sin duda, *hay* algo que parece sujetarnos a nuestras sillas! La teoría de

Einstein aparenta decirnos que la gravedad es algo inexistente, dado que siempre podemos eliminar la fuerza gravitatoria eligiendo simplemente un marco de referencia en caída libre. ¿Adónde ha ido a parar la gravedad en el enfoque einsteniano? En realidad, no ha ido a ninguna parte, sino que se halla escondida tras algunas sutilezas que hasta ahora hemos pasado por alto. En la próxima sección veremos dónde se encuentra oculto el campo gravitatorio.

Las fuerzas de marea

Las consideraciones de los párrafos anteriores son esencialmente locales. Hemos ignorado el modo en el que un campo de fuerzas gravitatorio newtoniano podría variar de un punto a otro. La dirección «hacia abajo» no es exactamente la misma aquí, en Oxford, que en Londres, debido a nuestras distintas posiciones sobre el globo terrestre. Si trato de eliminar el campo gravitatorio aquí, donde me encuentro, según mis consideraciones relativas a un marco de referencia rígido que cae libremente hacia el suelo de Oxford, ese mismo marco no será totalmente válido para alguien que se halle en Londres. Vemos, pues, que la «eliminación» del campo gravitatorio mediante la adopción de un marco en caída libre no es tan fácil como parece.

Para hacer la situación un poco más específica, imaginemos a un astronauta llamado Albert —al que, para abreviar, nos referiremos por su inicial, *A*— que cae libremente en las proximidades de la Tierra. Podríamos imaginar que *A* se precipita hacia el suelo sin

más, pero resultaría un tanto inhumano. Nos interesan las aceleraciones y no las velocidades directamente, así que también sería válido suponer que Albert se halla a salvo, en órbita alrededor de la Tierra. Supongamos que A está rodeado de una pequeña esfera de partículas, inicialmente en reposo respecto a él. Cada partícula tendrá una aceleración hacia el centro de la Tierra, C , de acuerdo con la ley cuadrática inversa de Newton. Las dos partículas P_1 y P_2 situadas en línea con CA tendrán aceleraciones en la dirección de C , pero la del punto más bajo, P_1 , será un poco mayor que la de A . Es decir, respecto a Albert, P_1 se acelerará más lentamente hacia el centro de la Tierra, C , mientras que P_2 lo hará más deprisa. A Albert le parecerá que tanto P_1 como P_2 se aceleran, alejándose de A . Por otra parte, cualquier partícula P_3 situada en el círculo horizontal de partículas con centro en A se acelerará ligeramente hacia dentro, al ser atraída por el centro de la Tierra, C , debido a que éste se halla a una distancia finita de A , lo que hace que la dirección «hacia abajo» sea ligeramente distinta. Respecto a A , la aceleración de esa partícula P_3 parecerá acercarla. La esfera de partículas, por lo tanto, experimentará una distorsión, adoptando una forma elipsoidal; en el eje horizontal, las partículas tenderán a acercarse a A , y en el eje vertical paralelo a AC , se moverán hacia fuera (*figura 3.1*).

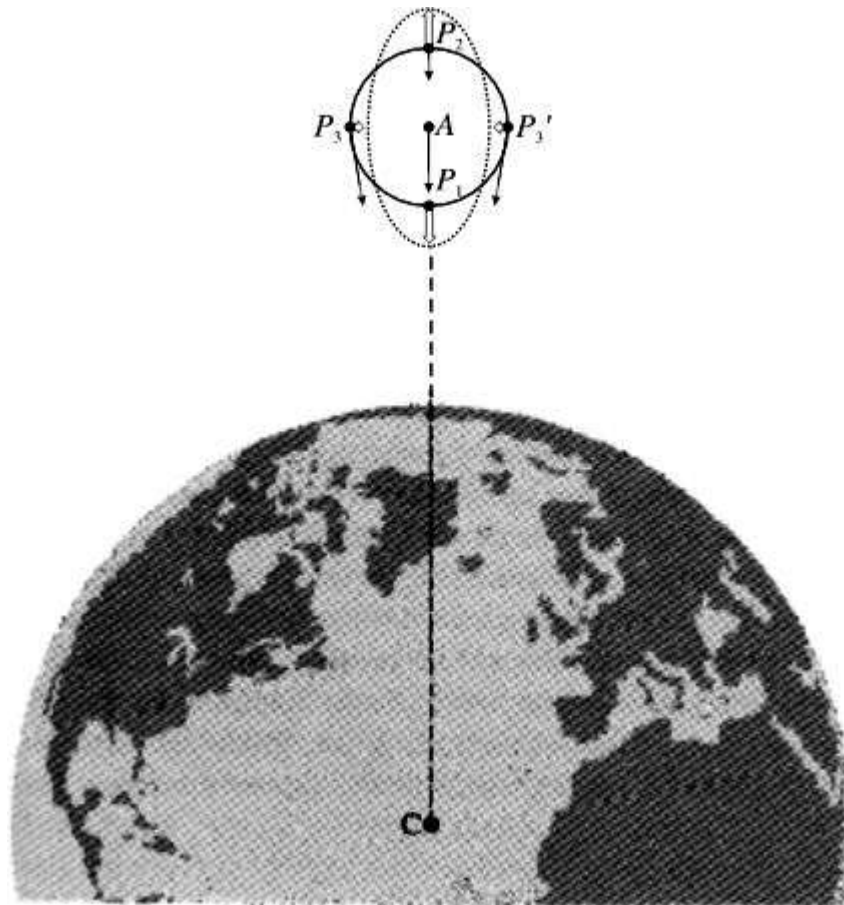


Figura 3.1. El efecto marea. Las flechas muestran la aceleración relativa.

Esta distorsión es conocida como *efecto marea* de la gravedad. La razón de emplear el término «marea» es que precisamente este efecto es el responsable de las mareas en los océanos terrestres, gobernado en este caso por la posición de la Luna. Para visualizarlo, imaginemos que *A* representa el centro de la Tierra y que la esfera de partículas representa la superficie de los océanos. Supongamos que *C* determina la posición de la Luna. De nuevo, las aceleraciones hacia *C* de los distintos puntos de la superficie oceánica varían. El

efecto resultante, respecto al centro de la Tierra, A , será una distorsión elipsoidal de la superficie de los océanos, que será más prominente en dirección a la Luna y, también, en la dirección opuesta. Se trata justamente del principal efecto que da lugar a las mareas. (Existen influencias secundarias, por ejemplo, el —más pequeño— efecto de marea debido al Sol y los efectos inerciales y de fricción debidos al movimiento real del agua en los océanos).

Una característica particular (e implícita) en la *ley cuadrática inversa* de Newton es que el *volumen* de la esfera de partículas permanece inicialmente constante en su momentánea distorsión para convertirse en elipsoide. (Lo cual equivale a decir que la aceleración hacia fuera en P_1 o P_2 es dos veces la aceleración hacia dentro de cualquier punto horizontal como P_3). Este hecho depende de la no existencia de una densidad de masa dentro de la esfera misma. Si hubiera una cantidad significativa de masa *dentro* de la esfera, existiría una aceleración hacia dentro adicional que serviría para *reducir* el volumen de la esfera en su movimiento inicial. La magnitud de esta reducción (inicial) del volumen es, en general, proporcional a la masa total englobada por la esfera. La magnífica teoría de la gravitación de Newton está avalada por hechos como el que acabamos de examinar.

Un caso particular de esta reducción de volumen tendría lugar si considerásemos que la esfera de partículas rodea por completo la Tierra, donde lo que nos preocupa es el propio campo gravitatorio terrestre y no las pequeñas correcciones debidas a la Luna, las

cuales son las (principales) responsables de las mareas. La distorsión de la esfera es, en este caso, una simple reducción de volumen. Se trata de una aceleración «hacia dentro» en toda la superficie terrestre, la cual constituye el familiar campo gravitatorio que nos sujeta a nuestras sillas.

La curvatura espaciotemporal

Aunque el concepto de espacio-tiempo aún no haya sido esbozado en nuestras consideraciones y se examine con más detalle en la sección siguiente, será útil detenerse en él por un momento, ya que el modo de contemplar la gravitación newtoniana en el apartado anterior anticipa cómo la perspectiva de Einstein —de la que el principio de equivalencia es uno de sus pilares— conduce de forma natural a la noción de que la gravitación es una *curvatura del espacio-tiempo*. Imaginemos que la historia del universo se hallara en su totalidad ante nosotros como un continuo *tetradimensional*. Por el momento, no estamos abandonando la física newtoniana; nos limitamos a contemplar el universo de Newton de una manera inusual, como un pedazo de geometría tetradimensional. Además de las tres coordenadas espaciales, x , y , z , introduciremos la coordenada temporal t , a modo de cuarta dimensión. Obviamente, visualizar esas cuatro dimensiones es bastante difícil, pero esta visualización no es realmente necesaria. «Olvidemos» temporalmente la coordenada espacial y , de manera que tengamos un espacio-tiempo tridimensional, representado por x , z y t . La *figura 3.2* nos

da una idea de lo que pretendemos. Una partícula puntual individual está aquí representada por una *curva* en el espacio-tiempo; esta curva que describe la historia de la partícula recibe el nombre de *línea de universo* de la partícula.

Intentaremos describir la historia de la esfera de partículas que rodeaba a Albert en la *figura 3.1* y ver qué relación tiene con la curvatura del espacio-tiempo. A la derecha de la *figura 3.2* he tratado de representar la historia de la evolución de esa esfera, eliminando una de las dimensiones espaciales (concretamente, la dimensión horizontal según el eje y). La esfera (reducida dimensionalmente) aparece como un círculo y, a medida que transcurre el tiempo, se va convirtiendo en una elipse. Obsérvese que las líneas de universo de las partículas desplazadas en sentido vertical, P_1 y P_2 (el eje mayor de la elipse), se curvan hacia fuera, mientras que las de las partículas desplazadas en sentido horizontal, P_3 y P'_3 (el eje menor), lo hacen hacia dentro.

Comparemos este «efecto de combadura» con el comportamiento de las geodésicas en una superficie curva. Una geodésica es una curva de longitud mínima sobre esa superficie. Imaginemos que mantenemos tenso un trozo de cuerda sobre una superficie así: describirá una geodésica. Si la superficie tiene lo que denominamos *curvatura positiva* (la curvatura de una superficie esférica ordinaria), dos geodésicas ligeramente separadas que al comienzo sean paralelas entre ellas tenderán a irse combando y acercando la una a la otra. Si la superficie, en cambio, posee lo que conocemos como

curvatura negativa (la de una silla de montar), las geodésicas inicialmente paralelas tenderán a divergir, apartándose una de otra (*figura 3.3*). Esta manifestación de la curvatura recibe el nombre de *desviación geodésica*.

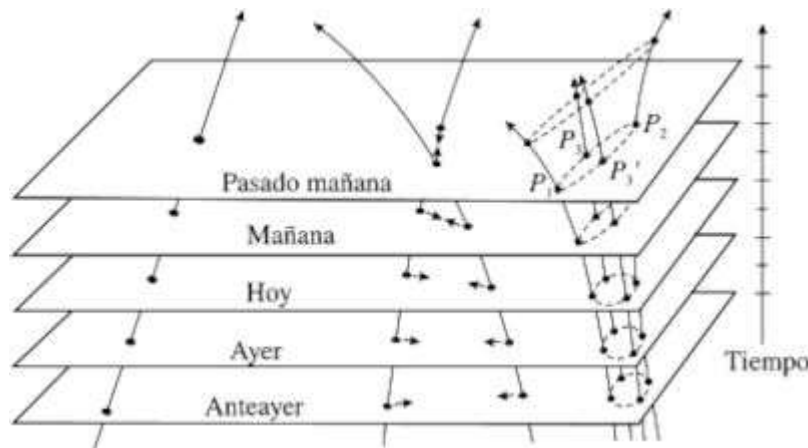


Figura 3.2. Espacio-tiempo (caso newtoniano). A la derecha se ilustra la desviación geodésica (efecto marea).

En nuestro modelo de espacio-tiempo para la distorsión de marea, ilustrado a la derecha de la *figura 3.2*, apreciamos una combinación de esos dos tipos de curvatura. Hay una curvatura positiva (una combadura hacia dentro) para las líneas de universo de las partículas desplazadas horizontalmente, P_3 y P'_3 , mientras que las líneas de universo correspondientes a las partículas desplazadas en sentido vertical, P_1 y P_2 , tienen curvatura negativa (se comban hacia fuera). Interpretar la distorsión de las líneas de universo en el *efecto marea* como cierta forma de desviación geodésica tiene sentido si asumimos que las líneas de universo de unas partículas que se

mueven libremente bajo el efecto de la gravedad son geodésicas espaciotemporales. Para ello es preciso disponer de la adecuada noción de «distancia» en el espacio-tiempo.

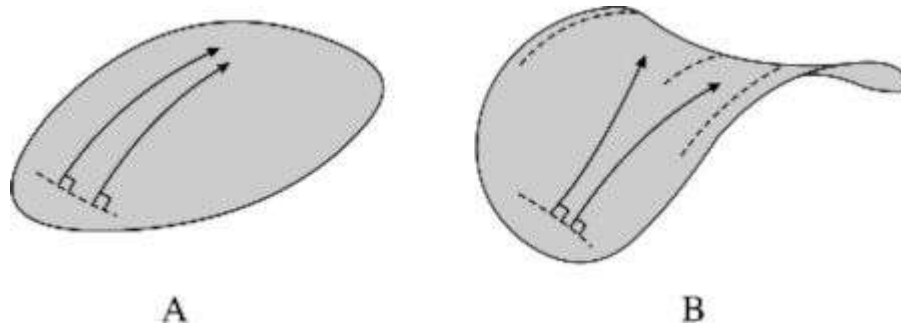


Figura 3.3. A) Una curvatura positiva (como la superficie de una naranja) hace que las geodésicas converjan. B) Una curvatura negativa (como una silla de montar) hace que las geodésicas diverjan.

Examinaremos este punto en las próximas dos secciones y comprobaremos que el efecto de marea es un ejemplo de desviación geodésica y, por lo tanto, una medida directa de la curvatura del espacio-tiempo.

La noción de curvatura en un número mayor de dimensiones es algo más complicada que en el caso bidimensional. En dos dimensiones, la curvatura en un punto dado viene expresada mediante un *único número*,⁴⁰ que será *positivo* para una superficie de tipo esférico y *negativo* para una similar a una silla de montar. Si el número de dimensiones es superior a dos, la curvatura viene expresada por

⁴⁰ Técnicamente, ese número es denominado curvatura *intrínseca* o curvatura *gaussiana* de la superficie. Examinaremos la noción de *intrínseca* con mayor detalle en posteriores secciones.

varios números, denominados *componentes* de curvatura, los cuales miden básicamente el tipo de curvatura bidimensional en distintas direcciones. En el ejemplo que acabamos de considerar hemos visto, de hecho, una componente positiva de curvatura en la dirección horizontal que va de A a P_3 y P'_3 y una componente negativa asociada a la dirección vertical que va de A a P_1 y P_2 . En el espacio-tiempo tetradimensional, la curvatura tiene *veinte* componentes independientes que pueden ser englobadas en un único ente matemático denominado «tensor de curvatura de Riemann». Discutiremos el concepto de tensor en una sección posterior, pero señalaremos de momento que la ecuación de Einstein es una ecuación de tensores en la que los subíndices (como la a y la b en R_{ab}) identifican las componentes en distintas direcciones.

Hasta ahora no hemos hecho uso realmente de la relatividad general: nos hemos limitado a contemplar la teoría de la gravitación de Newton desde la perspectiva einsteniana.⁴¹ Para penetrar en aquella tendremos que saber algo más sobre la relatividad especial: por qué se trata de una teoría sobre un espacio-tiempo tetradimensional, y cuál es la apropiada noción de «distancia» en esta geometría ψ . Abordamos estos temas seguidamente.

La noción de geometría ψ de Minkowski

⁴¹ Una teoría matemática completa para una geometría tetradimensional como la que se deriva de la teoría de Newton fue formulada por primera vez por el gran matemático francés Élie Cartan en 1923-1924.

Einstein basó su teoría especial de la relatividad de 1905 en dos principios fundamentales. El primero ya había sido establecido con anterioridad: todos los observadores en movimiento uniforme perciben las mismas leyes de la naturaleza. El segundo afirmaba que la velocidad de la luz es una constante fundamental y su valor no depende de la velocidad de la fuente. Algunos años antes, el gran matemático francés Henri Poincaré había planteado un esquema similar (y otros, como el físico holandés Hendrik Lorentz, habían avanzado en él). Pero Einstein fue el primero en ver que los principios básicos de relatividad tenían que ser aplicables a *todas* las fuerzas de la naturaleza.

Los historiadores aún discuten acerca de si Poincaré había comprendido la relatividad especial antes de que Einstein entrara en escena. En mi opinión, aunque esto fuera así, la relatividad especial no fue entendida en su *totalidad* (*ni* por Poincaré *ni* por Einstein) hasta que Hermann Minkowski presentó, en 1908, su modelo de espacio-tiempo tetradimensional. En una famosa conferencia dada en la Universidad de Gotinga, el matemático lituano proclamó: «En lo sucesivo, el espacio en sí y el tiempo en sí están condenados a desvanecerse en las sombras y sólo una especie de unión entre ambos conservará una realidad independiente».

Einstein no pareció apreciar al principio la trascendencia de la contribución de Minkowski, y durante un par de años no la consideró seriamente. Cuando después comprendió su potencia, la convirtió en plataforma sobre la que desarrollaría su relatividad

general, en la que la geometría del espacio-tiempo tetradimensional de Minkowski se hace *curva*.

La interpretación física de esta curvatura es, en esencia, la que ya le hemos dado, pero aún falta un ingrediente: la interpretación de las líneas de universo de las partículas que se mueven libremente bajo la acción de la gravedad como *geodésicas* en la geometría del espacio-tiempo. Ejemplos de las citadas geodésicas serían las líneas de universo de nuestro astronauta A y de la esfera de partículas que le rodea. Para entender esta interpretación debemos examinar primero la estructura matemática tetradimensional *plana* que Minkowski introdujo en realidad con el fin de describir la relatividad especial.

A tal efecto, será de ayuda empezar considerando la geometría euclídea tridimensional ordinaria. Utilizando coordenadas cartesianas, x, y, z , para identificar cada punto del espacio euclídeo tridimensional, la *distancia* l desde el origen (con coordenadas $x = y = z = 0$) al punto (X, Y, Z) (con coordenadas $x = X, y = Y, z = Z$) viene dada por la fórmula de Pitágoras:

$$l^2 = X^2 + Y^2 + Z^2$$

El lector recordará el teorema de Pitágoras, el cual establece que el cuadrado de la hipotenusa de un triángulo rectángulo es igual a la suma de los cuadrados de los catetos. Esto da lugar a la fórmula bidimensional $l^2 = X^2 + Y^2$, ya que la distancia entre dos puntos en

el plano es la hipotenusa l de un triángulo rectángulo cuyos catetos son X e Y . La extensión a tres dimensiones es una consecuencia inmediata.

Podemos usar la fórmula de arriba para calcular la distancia al cuadrado entre dos puntos cualesquiera sólo con sustituir X por la diferencia entre las abscisas de ambos puntos y hacer lo mismo con Y y Z .

Es fácil generalizar la fórmula a cuatro dimensiones y obtener el cuadrado de la distancia desde el origen al punto $w = W, x = X, y = Y, z = Z$ en el espacio euclídeo tetradimensional como:

$$l^2 = W^2 + X^2 + Y^2 + Z^2$$

Sin embargo, la geometría del espacio-tiempo de Minkowski difiere sutil pero trascendentalmente. Aunque las coordenadas ψ resultan ligadas unas a otras en la teoría de la relatividad mediante una especie de rotación (la «transformación de Lorentz»), el modo en el que una rotación euclídea ordinaria mezcla las coordenadas w, x, y, z no nos da la receta correcta. Hay una distinción cualitativa importante en las coordenadas espaciotemporales de la descripción de Minkowski, que se manifiesta como una diferencia de *signos* en la fórmula de la distancia.

En lugar de la cuarta coordenada espacial w , introducimos una coordenada temporal, t . ¿En qué sentido hay que modificar la fórmula anterior para obtener la medida minkowskiana correcta de

la «distancia» τ ? Para llegar a obtener dicha medida debemos invertir los signos de *todas* las contribuciones espaciales, dejando que sólo la coordenada temporal $t = T$ contribuya con signo positivo:

$$\tau^2 = T^2 - X^2 - Y^2 - Z^2$$

Como estamos utilizando unidades tanto de distancia como de tiempo, la velocidad de la luz se convierte en el *nexo* de unión. Es decir, si estuviéramos usando el año como unidad de tiempo, deberíamos emplear como unidad de distancia espacial el año-luz. Si la unidad temporal es el segundo, debemos usar el segundo-luz (unos 300.000 kilómetros) como unidad de medida espacial.

¿Qué clase de «distancia» es entonces la definida por τ ? En realidad, τ es una medida de *tiempo*, lo que se conoce como *tiempo propio*. Si el punto del espacio-tiempo P , de coordenadas $t = T$, $x = X$, $y = Y$, $z = Z$, es tal que el resultado de la operación de la derecha es *positivo*, se dice que P presenta una *separación tipo tiempo* respecto al origen O , lo que, desde el punto de vista físico, significa que para la línea de universo de una partícula es teóricamente posible pasar de O a P (si T es positivo) o de P a O (si T es negativo). Si esa partícula se mueve uniformemente y en línea recta de O a P , la variable τ (en valor absoluto) es el *tiempo* (el tiempo propio) realmente experimentado por la partícula entre O y P , medido por un reloj ideal situado en la propia partícula. (El hecho de que este tiempo no sea simplemente el t newtoniano, sino que dependa también de las

coordenadas espaciales, es una expresión de la «relatividad del tiempo» implícita en la relatividad especial). Como en la geometría euclídea, el razonamiento es válido también si sustituimos el origen O por cualquier otro punto arbitrario y T, X, Y, Z , por las *diferencias* entre las respectivas coordenadas t, x, y, z de los dos puntos P y P' implicados, siendo t el tiempo experimentado por la partícula al moverse por inercia de P a P' .

La geometría minkowskiana tiene la curiosa propiedad de que la «distancia» entre dos puntos P y P' puede ser nula aunque tales puntos no coincidan. Esto sucede cuando un rayo de luz puede contener tanto a P como a P' (pensemos que una «partícula de luz» o fotón viaja a la velocidad de la luz). Así pues, si llamamos *tiempo propio* a la interpretación de la «distancia de Minkowski» expuesta anteriormente, observamos que un fotón no experimentaría en absoluto el paso del tiempo (si los fotones fuesen capaces de experimentar algo). Para un punto P dado, el lugar geométrico de esos puntos P' constituye el *cono de luz* (futuro) de P . Los conos de luz son importantes porque determinan las *propiedades de causalidad* del espacio de Minkowski, pero no nos ocuparemos mucho de ellos aquí. El único aspecto esencial que se debe tener en cuenta es que la línea de universo de una partícula con masa tiene que hallarse *dentro* del cono de luz en todo momento. Esta circunstancia expresa simplemente el hecho de que la partícula nunca supera la velocidad de la luz. Una línea de universo de esta

clase recibe el nombre de curva *tipo tiempo*. Toda línea de universo de una partícula masiva es necesariamente una curva tipo tiempo. Cualquier curva tipo tiempo (es decir, cualquier línea de universo válida para una partícula), sea recta o no, tiene una «longitud» de Minkowski. Una línea de universo curvada describe a una partícula *acelerada*. Su «longitud» es, simplemente, el tiempo (propio) experimentado por la partícula. Para obtener matemáticamente esa longitud, basta con hacer lo mismo que en una geometría euclídea ordinaria, pero teniendo en cuenta las diferencias de signo, indicadas con anterioridad, que aparecen al pasar a la geometría minkowskiana. En la práctica, hay que usar la expresión *infinitesimal* que mide la «distancia» entre dos puntos infinitamente próximos y después «sumar» (en términos matemáticos, *integrar*) todas esas distancias infinitesimales a lo largo de la curva para obtener su longitud total. En la geometría euclídea tridimensional, esa distancia infinitesimal « dl » se obtiene a partir de las coordenadas cartesianas x, y, z mediante la fórmula:

$$dl^2 = dx^2 + dy^2 + dz^2$$

En el caso minkowskiano, la expresión es la siguiente:

$$dt^2 = dt^2 - dx^2 - dy^2 - dz^2$$

pero la interpretación es del todo análoga. (Aquéllos menos acostumbrados a la notación pueden imaginar que dt significa $t' - t$, dx sería $x' - x$, etcétera, con P' infinitamente próximo a P dentro del cono de luz de P). El intervalo total de tiempo (propio) entre dos puntos pertenecientes a una línea de universo, medido por un reloj ideal, es la «longitud» total de la fracción de dicha línea comprendida entre ambos puntos.

Una característica importante de la longitud en la geometría euclídea es que, de todas las curvas que pasan por los dos puntos, la longitud mínima corresponde a una línea recta («la distancia más corta entre dos puntos»). Existe una propiedad análoga en la geometría de Minkowski, aunque las cosas suceden de otra manera. Si elegimos un par de puntos con una separación tipo tiempo entre ellos, de todas las curvas tipo tiempo que los unen, el tiempo propio máximo corresponde a la línea recta. Desde el punto de vista físico, este hecho da lugar a lo que a veces se denomina «paradoja del reloj» (o «paradoja de los gemelos»), según la cual alguien que viaja hasta una estrella lejana y regresa envejece menos (debido a la menor «distancia de Minkowski» recorrida) que su hermano gemelo que ha permanecido todo el tiempo en la Tierra. El gemelo «terricola» tiene una línea de universo recta y, por consiguiente, experimenta un lapso de tiempo mayor que su hermano «viajero», cuya línea de universo se curva a causa de la aceleración. Es erróneo, sin embargo, considerar el hecho una paradoja. Aunque pueda resultar chocante, no se trata de ninguna paradoja y muchos experimentos

han confirmado el fenómeno con gran precisión. La geometría minkowskiana hace que la diferencia temporal entre los dos gemelos parezca casi «ordinaria».

¿Qué es lo que condujo a Einstein a modificar la bella geometría espaciotemporal de Minkowski y a introducir un espacio-tiempo curvado? Hemos visto que, en la relatividad especial, las partículas que se mueven libremente en ausencia de fuerzas —es decir, las partículas que se desplazan *inercialmente*— poseen líneas de universo rectas en el espacio de Minkowski. El deseo de Einstein de incorporar el principio de equivalencia a la teoría física le hizo ver que era preciso un *nuevo* concepto de «movimiento inercial». Como la fuerza gravitatoria puede ser localmente eliminada mediante la adopción de un marco de referencia en caída libre, según el punto de vista de Einstein, no consideraremos que dicha fuerza es «real». De este modo, Einstein vio que necesitaba introducir una noción diferente de movimiento inercial, la denominada *caída libre bajo la acción de la gravedad*, en la que no hubiera otras fuerzas en acción. Debido al efecto marea que hemos descrito anteriormente, no cabe pensar que las partículas «inerciales» (en el sentido einsteniano) tengan líneas de universo rectas (es decir, geodésicas) en la geometría de Minkowski. Así pues, necesitamos generalizar esta geometría, lo que la convierte en *curvada*. Einstein halló que, en efecto, las líneas de universo de sus partículas inerciales podían ser *geodésicas* en esta nueva geometría —maximizando localmente la «longitud», en vez de hacerla mínima, según lo que habíamos

dicho— y la distorsión de marea, un ejemplo de desviación geodésica, la cual proporciona una medida directa de la curvatura espaciotemporal. Examinemos algo más en detalle esta curvatura.

Geometría espaciotemporal curva

Dos grandes matemáticos alemanes del siglo XIX, Carl Friedrich Gauss y Bernhard Riemann, introdujeron la noción general de «geometría curva». Para hacernos una idea de este tipo de geometría, pensemos en la superficie de una pelota de tenis partida por la mitad. La semiesfera puede ser deformada de muchas maneras, pero lo que denominamos su geometría *intrínseca* permanece inalterada bajo tales deformaciones. La geometría intrínseca está relacionada con las distancias medidas *a lo largo* de la superficie y no tiene que ver con el espacio (en nuestro ejemplo, el espacio euclídeo tridimensional ordinario) en el que cabe pensar que dicha superficie está inserta. Las distancias medidas directamente desde un punto a otro, atravesando el espacio *fuera* de la superficie, no forman parte de la geometría intrínseca. La longitud de una línea dibujada *sobre* media pelota de tenis no varía por más que la doblemos, aplastemos o flexionemos.

Gauss presentó el concepto de geometría intrínseca para el caso bidimensional —el de la pelota de tenis— en 1827. Demostró que en esta geometría existe una noción de curvatura que es completamente intrínseca y a la que no le afecta cambio alguno en la forma que la superficie pueda adoptar. Esa curvatura puede ser

calculada a partir de medidas de longitud tomadas a lo largo de la superficie, entendiendo dichas medidas de longitud de una curva como el resultado de integrar distancias infinitesimales dl a lo largo de ella, como se decía con anterioridad. En la práctica, y tras aplicar a la superficie un sistema de coordenadas idóneo, por ejemplo, u, v , obtenemos la siguiente expresión:

$$dl^2 = A du^2 + 2B du dv + C dv^2$$

Donde A, B y C son funciones de u y v (la expresión es casi igual a la fórmula de Pitágoras para la distancia,

$$dl^2 = dx^2 + dy^2$$

pero usando coordenadas generales u, v).

En 1854, Riemann mostró el modo de generalizar la geometría intrínseca de Gauss a superficies de más dimensiones. El lector tal vez se pregunte qué interés podrían tener los matemáticos en geometrías intrínsecas de un número de dimensiones mayor. El espacio ordinario tiene sólo tres dimensiones y no parece que tenga mucho sentido «flexionar» una «superficie» tridimensional en su seno, y no hablemos de superficies de más dimensiones. Ante todo debemos puntualizar que el escenario anteriormente descrito es útil sólo para empezar a entender el concepto de «geometría intrínseca». En realidad, habría que pensar en la geometría intrínseca de

nuestra superficie como en algo que tiene sentido por sí mismo, sin necesidad de que haya un espacio de por medio. De hecho, una de las motivaciones originales de Riemann fue pensar que el espacio físico tridimensional que habitamos podría tener una geometría intrínseca curva, sin que deba «residir» en el seno de un espacio de más dimensiones.

No obstante, Riemann consideró también geometrías intrínsecas n -dimensionales y cabría preguntarse por qué. En este punto hay que hacer dos consideraciones importantes. En primer lugar, resulta que el formalismo matemático desarrollado para manejar espacios curvos tridimensionales es básicamente el mismo que para manejar espacios n -dimensionales curvos en general, con lo que no se gana nada restringiendo el caso a $n = 3$. En segundo, la geometría (intrínseca) n -dimensional curva es importante en muchos contextos en los que n no se refiere al número de dimensiones de un espacio ordinario, sino al de grados de libertad de un sistema. Existen espacios matemáticos abstractos denominados «espacios de configuraciones», en los que cada punto representa una disposición concreta de todas las partes de cierta estructura física. En estos espacios, de hecho, la magnitud n puede ser muy grande —en el caso de que el sistema conste de muchas partes— y la geometría de Riemann resulta de gran ayuda.

Las nociones de «métrica» y «curvatura» en el caso n -dimensional son generalizaciones naturales de las introducidas por Gauss para las superficies bidimensionales ordinarias pero, debido al gran

número de componentes involucradas, es necesaria una notación capaz de manejarlas adecuadamente en su conjunto. En lugar de las tres «componentes métricas» A , B y C que aparecen en la expresión anterior para $d\ell^2$ en el caso bidimensional, para tres dimensiones necesitamos seis de esas componentes. Son las componentes de un *tensor métrico*, representado generalmente por g_{ab} . Esta variable sirve para definir la adecuada noción de «distancia» entre puntos vecinos, denotada habitualmente por ds .⁴²

En la geometría de Riemann, obtenemos la *longitud de una curva* en el espacio *integrando* ds a lo largo de la curva, igual que en caso del espacio plano ya analizado. Una geodésica en una variedad de Riemann es una curva que minimiza (localmente) la longitud, es decir, que describe «la distancia más corta entre dos puntos», en el sentido adecuado. La *curvatura* del espacio de Riemann es la magnitud que describe la cantidad de desviación geodésica en todas las direcciones posibles del espacio (según se indicaba con anterioridad). Como era de esperar, la curvatura tiene muchas componentes, cada una de las cuales representa la curvatura asociada a una de las muchas direcciones posibles en las que cabe

⁴² La anterior expresión euclídea para $d\ell^2$ se convierte ahora en la conocida fórmula $ds^2 = g_{ab} dx^a dx^b$ (en casi toda la bibliografía, « $d\ell$ » suele ser reemplazado por « ds »). Para un espacio n -dimensional se necesitan n coordenadas independientes, denotadas aquí por x^1, x^2, \dots, x^n . Esto puede resultar un poco confuso, ya que la notación « x^2 » no significa aquí « x al cuadrado», ni « x^3 » quiere decir « x al cubo», etc. La notación « x^a » (o « x^b », etc.) es un símbolo genérico para una de esas coordenadas. Del mismo modo, « g_{ab} » es un símbolo genérico para una de las magnitudes $g_{11}, g_{12}, \dots, g_{nm}$, cada una de las cuales es una de las $n(n+1)/2$ funciones independientes, ya que $g_{ab} = g_{ba}$. Se adopta aquí la *convención de suma de Einstein*, por la cual los índices repetidos se suman. Así pues, la expresión « $g_{ab} dx^a dx^b$ » equivale a « $g_{11} dx^1 dx^1 + g_{12} dx^1 dx^2 + \dots + g_{nm} dx^n dx^m$ ». En el caso bidimensional, $g_{11} = A$, $g_{12} = g_{21} = B$ y $g_{22} = C$ son funciones de las dos coordenadas u y v , donde $x^1 = u$ y $x^2 = v$.

medir la desviación geodésica. En la práctica, toda esa información queda recogida en una variable denominada *tensor de Riemann*. El tensor de Riemann (o su colección de componentes) se representa por R_{abcd} , donde los subíndices se refieren a todas las maneras distintas en las que se puede medir la desviación geodésica.⁴³

La relatividad general de Einstein está formulada en términos de un concepto de espacio-tiempo curvo tetradimensional que guarda la misma relación con el espacio-tiempo plano de Minkowski que el concepto de geometría curva de Riemann respecto a la geometría euclídea plana. La métrica g_{ab} puede ser utilizada para definir longitudes de curvas pero, al igual que en la geometría espaciotemporal plana de Minkowski, es mejor pensar que esa «longitud» corresponde a un *tiempo* medido por una partícula a lo largo de su línea de universo. Aquellas líneas de universo que maximizan localmente esa medida de tiempo son las geodésicas espaciotemporales y corresponden a partículas que se mueven inercialmente (en el sentido einsteniano de que «se mueven libremente bajo la acción de la gravedad», como ya hemos indicado). Recordemos ahora que, en la teoría de Newton, la desviación geodésica en el espacio-tiempo causada por la gravedad tiene la propiedad de que en el vacío no hay inicialmente cambio de volumen alguno, mientras que cuando existe materia en las proximidades de las geodésicas desviadas, la *reducción de volumen* es proporcional a la masa total rodeada por las geodésicas. Esta

⁴³ Hay expresiones explícitas, aunque complicadas, que nos dicen cómo se calcula R_{abcd} a partir de g_{ab} y sus derivadas parciales primera y segunda con respecto a las coordenadas x^a .

reducción de volumen es un *promedio* de la desviación geodésica en todas las direcciones alrededor de la geodésica central —la línea de universo del astronauta A , en nuestro caso—. Así pues, necesitamos una entidad que mida el citado promedio. Esta entidad es conocida como *tensor de Ricci* y se obtiene a partir de R_{abcd} . Su colección de componentes se representa habitualmente como R_{ab} . Existe también un promedio global único, R , denominado *curvatura escalar*.⁴⁴ Recordemos que R_{ab} y R , junto con g_{ab} , son precisamente los términos que aparecen en el lado izquierdo de la ecuación de Einstein.

Las variables g_{ab} , R_{abcd} y R_{ab} constituyen (conjuntos de componentes de) un tipo de ente matemático denominado tensor. Los tensores son fundamentales en el estudio de la geometría de Riemann. La razón tiene que ver con el hecho de que, en este contexto, es indiferente el conjunto de coordenadas elegido para describir la variedad (lo cual es una consecuencia de la aplicación estricta del principio de equivalencia). Es posible usar cualquier juego de coordenadas, según convenga a cada uno. El *cálculo tensorial* constituyó un logro técnico extraordinario y fue desarrollado a finales del siglo XIX por varios matemáticos como un medio de extraer *información invariante* sobre la variedad, su métrica y su curvatura («invariante» quiere decir, esencialmente, «independiente de cualquier elección concreta de coordenadas»).

⁴⁴ Mediante la convención de suma de Einstein podemos definir R_{ab} y R mediante las relaciones $R_{ab} = R_{abcd} g^{cd}$ y $R = R_{ab} g^{ab}$, donde g^{ab} es la *inversa* de g_{ab} , en el sentido empleado en álgebra matricial.

Al tratar de incorporar plenamente el principio de equivalencia a una teoría física de la gravitación, Einstein se dio cuenta de que necesitaba una formulación que fuese «invariante» en el sentido antes indicado. Llamó a este requisito *principio de covarianza general*. Las coordenadas espaciotemporales empleadas para describir dos marcos de referencia acelerados diferentes pueden estar relacionadas entre ellas de alguna manera (muchas veces complicada), pero ninguno de ambos conjuntos es «preferible» al otro. Einstein tuvo que pedirle a su colega Marcel Grossmann que le enseñara los rudimentos del «cálculo de Ricci» (que era el nombre que tenía entonces el cálculo tensorial). La única diferencia importante entre la geometría espaciotemporal curva que él precisaba y la geometría de Riemann para la que había sido diseñado el cálculo de Ricci (en el caso tetradimensional) era el cambio de «signatura» necesario para pasar de la estructura localmente euclídea de los espacios de Riemann a la estructura localmente minkowskiana requerida por un espacio-tiempo relativista.

La relatividad general

Volvamos a Albert, nuestro astronauta *A* rodeado de una esfera de partículas. Tanto *A* como las partículas se mueven inercialmente en el sentido einsteniano (es decir, libremente bajo la acción de la gravedad) y, según lo postulado, sus líneas de universo deberían ser

geodésicas en el espacio-tiempo.⁴⁵ Recordemos que, en la teoría de Newton, la reducción inicial del volumen de esa esfera es proporcional a la masa que engloba y que el tensor de Ricci es la variable que mide ese cambio de volumen. Según esto, cabría esperar que la generalización relativista adecuada de la teoría de Newton fuera aquella en la que hubiera una ecuación que relacionara el tensor espaciotemporal de Ricci con otra variable tensorial que midiera la densidad de masa de la materia. Esta última variable recibe el nombre de *tensor de energía-momento* y su conjunto de componentes se representa por T_{ab} . Una de esas componentes mide la densidad de masa-energía; las otras miden densidades de momento, esfuerzos y presiones en el material. En la teoría de Newton hay un factor de proporcionalidad entre la aceleración interna y la densidad de masa: la constante de la gravitación universal, G . Esto condujo a Einstein a proponer una ecuación como la siguiente:

$$R_{ab} = -4\pi G T_{ab}$$

El término 4π proviene del hecho de que estamos tratando con densidades y no con partículas individuales y el signo menos, de que la aceleración es interna y la convención que he adoptado para el signo del tensor de Ricci es tal que las aceleraciones externas

⁴⁵ Einstein demostraría más tarde que este postulado puede ser deducido de su ecuación de campo, al igual que otras asunciones razonables.

aparecen con signo positivo, aunque existen innumerables opciones para ésta y otras convenciones aplicables al caso.

La ecuación es, de hecho, la primera que Einstein presentó, hasta que poco después se dio cuenta de que no era consistente con cierta ecuación,⁴⁶ que necesariamente debe satisfacer T_{ab} y que expresa una ley fundamental de *conservación de la energía* para las fuentes de materia. Esto le obligó, tras varios años de vacilaciones, a sustituir la magnitud R_{ab} de la izquierda por otra ligeramente distinta, $R_{ab} - 1/2R g_{ab}$, la cual, por razones puramente matemáticas, satisface *también*, milagrosamente, la misma ecuación que T_{ab} . Mediante este cambio, Einstein obtuvo la consistencia necesaria en la ecuación resultante, la muy notable y justamente famosa *ecuación de Einstein*:⁴⁷

$$R_{ab} - 1/2R g_{ab} = -8\pi G T_{ab}$$

Esta «reducción de volumen» en la desviación geodésica a la que la ecuación da lugar es sólo ligeramente distinta de la que cabría esperar en la teoría de Newton, debido al término adicional « $-1/2Rg_{ab}$ » del lado izquierdo de la ecuación. La «fuente de la gravitación» (es decir, la fuente de la reducción del volumen), en

⁴⁶ La desaparición de la «divergencia covariante» de T_{ab} .

⁴⁷ En el otoño de 1915, el matemático David Hilbert llegó a esta misma ecuación casi al mismo tiempo que Einstein, pero por diferente camino. El hecho daría lugar a una desagradable disputa sobre prioridades. Pero la contribución de Hilbert, aunque técnicamente importante, no menoscaba realmente la prioridad fundamental de Einstein en esta materia. Véase, en particular, *New Light on the Einstein-Hilbert Priority Question*, de J. Stachel, aparecido en *Journal of Astrophysics and Astronomy*, 1999, 20 (3-4), págs. 91-101.

lugar de ser simplemente $4\pi G$ multiplicado por la densidad de *masa* (en el sentido de masa-energía expresado por T_{ab}), resulta ser ahora $4\pi G$ multiplicado por la densidad de masa más la *suma de las presiones* en el material, en tres direcciones mutuamente ortogonales (provenientes de las otras componentes de T_{ab}). En los materiales ordinarios, tales como los que forman las estrellas comunes y los planetas, las presiones son muy pequeñas comparadas con las densidades de masa (debido a que las partículas que constituyen estos cuerpos se mueven lentamente en comparación con la velocidad de la luz) y la teoría de Newton predice su comportamiento de forma muy precisa. Hay, sin embargo, algunas circunstancias (p. ej., la fase de inestabilidad de una estrella supermasiva, a medida que colapsa para convertirse en un agujero negro) en las que esa sutil diferencia se hace muy relevante.

Comprobaciones clásicas de la relatividad general

De la discusión anterior podría parecer que la relatividad general de Einstein es un mero ajuste técnico de la teoría de Newton para hacerla concordar con los principios de relatividad y equivalencia. En efecto, se podría decir así, aunque la forma en que he presentado la comparación entre ambas iba orientada a facilitar su comprensión y no corresponde al modo en que se desarrollaron las cosas. Concentrándonos en la fuerza de marea de la gravitación de Newton como en algo que no puede ser eliminado en una caída libre, hemos podido ver con más claridad su relación con la

curvatura espaciotemporal y, por lo tanto, con el marco einsteniano de la relatividad general.

De hecho, resulta notablemente difícil encontrar diferencias claramente observables entre las dos teorías. Originalmente, existían las llamadas «tres pruebas» de la relatividad general. La más impactante de las tres fue la explicación del avance del perihelio del planeta Mercurio en su órbita alrededor del Sol. Se conocía, desde el siglo XIX, una curiosa discrepancia con la teoría de Newton en el movimiento de Mercurio. Descontando los efectos perturbadores de todos los demás planetas, quedaba aún una pequeña componente adicional en el movimiento de Mercurio que suponía una oscilación en el eje de su elipse orbital de 43 segundos de arco por siglo. Esta cantidad es tan diminuta que la citada elipse tardaría tres millones de años en completar el giro debido en exclusiva a este efecto. Los astrónomos habían conjeturado diversas explicaciones, incluida la existencia de otro planeta en la órbita de Mercurio, al que incluso se le llegó a dar nombre (Vulcano). Ninguna de esas ideas funcionó, hasta que la teoría de Einstein explicó con precisión la discrepancia en lo que constituyó una comprobación espectacular de su validez.⁴⁸ Las otras dos pruebas se refieren al enlentecimiento de los relojes ideales en un campo gravitatorio y a la desviación de la luz por parte del campo gravitatorio solar. El

⁴⁸ En 1966, Robert Dicke «sobresaltó» a la comunidad científica al anunciar que cuidadosas observaciones del achatamiento solar realizadas por Goldenberg y por él mismo demostraban que el Sol poseía un momento cuadrupolar de tal magnitud que el avance del perihelio de Mercurio predicho por la relatividad general se iba por completo al traste. Afortunadamente, posteriores observaciones y consideraciones teóricas demostraron que la conclusión de Dicke era errónea.

efecto de enlentecimiento de los relojes fue confirmado de manera convincente mediante un experimento realizado en 1960 por Pound y Rebka, aun admitiendo que era una comprobación un tanto débil de la relatividad general, pues se trataba de una consecuencia directa del principio de conservación de la energía y de la ecuación $E = hf$ para la energía de un fotón.

El efecto de la desviación de la luz tiene una historia más interesante. Antes de llegar a la relatividad general, Einstein había hecho uso del principio de equivalencia para predecir, en 1911, que el Sol desviaría la luz procedente de una estrella una cantidad que es sólo la mitad de lo que la teoría completa predice. El efecto sería observable durante un eclipse total de Sol, así que en 1914 se decidió organizar una expedición a Crimea para comprobar la predicción. Desde el punto de vista de Einstein, fue una suerte que la primera guerra mundial impidiera llevarla a cabo. Cuando, finalmente, Arthur Eddington viajó a la Isla del Príncipe para observar la desviación de la luz durante el eclipse de 1919, Einstein había encontrado ya la teoría correcta —en 1915— y los resultados sirvieron para confirmarla de manera triunfal. Desde la perspectiva actual, esas observaciones pueden parecer menos convincentes de lo que lo fueron en su tiempo, cuando supusieron el éxito de la teoría einsteniana. En cualquier caso, las observaciones modernas del efecto y de un retardo temporal constatado por Shapiro añaden crédito adicional a la predicción de Einstein.

La desviación einsteniana de la luz actualmente está tan bien establecida que se emplea de forma rutinaria como herramienta en las observaciones astronómicas y cosmológicas. Las galaxias lejanas se convierten en complejas lentes para fuentes de luz aún más distantes, proporcionando información, que no es posible obtener fiablemente por otros medios, sobre la distribución de la masa en el universo. La predicción de Einstein ha resultado ser una extraordinaria sonda para detectar materia en el último rincón del cosmos.

Ondas gravitatorias

Una de las predicciones más notables de la teoría de Einstein es la existencia de *ondas gravitatorias*.⁴⁹ La teoría del electromagnetismo de Maxwell condujo a la predicción de que las ondas producidas por campos eléctricos y magnéticos oscilantes debían propagarse a través del espacio a la velocidad de la luz y Maxwell había postulado en 1865 que la propia luz era un efecto de esta naturaleza. En la actualidad, la predicción de Maxwell ha sido confirmada exhaustivamente en muchas situaciones experimentales. La teoría de la gravitación de Einstein guarda muchas semejanzas con la teoría de Maxwell, y una de ellas es la existencia de las ondas gravitatorias, distorsiones espaciotemporales que se propagan a la velocidad de la luz. Esas ondas serían emitidas por cuerpos

⁴⁹ Curiosamente y sin contar con la plataforma de la relatividad general, Poincaré había predicho ya en 1905 la existencia de ondas gravitatorias, basándose en analogías con la teoría del electromagnetismo de Maxwell.

gravitatorios en órbita unos alrededor de otros, pero el efecto es generalmente muy pequeño. En nuestro sistema solar, la mayor emisión de energía en forma de ondas gravitatorias proviene del movimiento de Júpiter alrededor del Sol. Esta pérdida de energía equivale solamente a la luz de una bombilla de 40 watios.

De hecho, e influido tal vez por su colega el físico polaco Leopold Infeld, Einstein llegó a cuestionar su creencia de que un sistema gravitatorio libre pudiera perder realmente energía en forma de ondas gravitatorias. A principios de la década de 1960, cuando comenzaba a interesarme por la teoría de Einstein, existía un áspero debate en relación con el tema. A la vez, tenían lugar los primeros avances importantes sobre la relatividad general. Durante muchos años, desde la época en la que fue concebida, la teoría había atraído la atención de muy pocos físicos relevantes y solía ser encuadrada en el ámbito de la matemática pura. Pero a comienzos de la citada década brotó un repentino interés por ella. En particular, los trabajos de varios teóricos proporcionaron lo que en mi opinión era una demostración convincente de que las ondas gravitatorias eran un fenómeno físico real y de que la pérdida de energía debida a esas ondas estaba de acuerdo con la fórmula que Einstein propusiera muchos años antes, en 1918.

Con posterioridad, la teoría de Einstein ha adquirido un extraordinario auge gracias a las observaciones (y al análisis teórico) de Joseph Taylor y Russell Hulse. Ellos fueron los primeros en observar, en 1974, las señales pulsantes procedentes del sistema

PSR 1913+16, formado por dos estrellas de neutrones. Las variaciones de esas señales proporcionan información detallada sobre las masas de las estrellas y sobre sus órbitas y es posible cruzar dicha información con lo que predice la relatividad general. La concordancia entre ambas es extraordinariamente precisa. En los veinticinco años en que el citado sistema ha estado siendo observado, la variación en el periodo de las señales ha sido de una parte en 10^{14} , es decir, de una parte en cien billones. En una primera aproximación, el hecho verifica las órbitas newtonianas de las estrellas. En segundo lugar, proporciona una confirmación detallada de las correcciones previstas por la relatividad general para esas órbitas (del tipo de las que se aplican al avance del perihelio de Mercurio). Finalmente, la pérdida de energía por parte del sistema en forma de ondas gravitatorias concuerda con precisión con la teoría einsteniana. En 1993, Hulse y Taylor fueron galardonados con el Premio Nobel de Física por el descubrimiento y análisis de ese notable sistema estelar. Desde sus inciertos comienzos, cuando parecía una teoría extravagante y de débil fundamento, la relatividad general ha llegado a convertirse hoy en la teoría física que se ajusta con mayor grado de precisión al comportamiento de la naturaleza.

La existencia de ondas gravitatorias parece estar fuera de toda duda en el caso del sistema PSR 1913+16. Pero estas ondas todavía no han sido observadas de manera convincente y *directa* aquí, en la Tierra. Hay varios detectores en diferente estado de construcción

que deberían poder registrar estas ondas en el futuro. Más aún, el conjunto de esos detectores, ubicados en distintos puntos del globo, se podría convertir dentro de unos años en un telescopio de ondas gravitatorias a escala planetaria, capaz de obtener información sobre grandes cataclismos (como colisiones de agujeros negros) que tengan lugar en galaxias muy lejanas. Esto abriría un nuevo tipo de ventana hacia el universo, en el que las familiares ondas electromagnéticas serían sustituidas por ondas gravitatorias. Como en el caso de la desviación de la luz, la predicción de Einstein sobre la existencia de ondas gravitatorias puede llegar a convertirse en una nueva y poderosa herramienta que nos permita conocer lo que hay en los confines del universo.

Algunos problemas de la relatividad general

Hemos visto hasta ahora algunos de los extraordinarios éxitos de la relatividad general. ¿Qué hay acerca de sus limitaciones? Un lugar común en torno a la teoría es que sus ecuaciones son notoriamente difíciles de resolver. En efecto, a pesar de su apariencia relativamente simple, la ecuación de Einstein encierra una enorme complejidad, la cual se pone de manifiesto cuando escribimos la expresión $R_{ab} - \frac{1}{2}Rg_{ab}$ de forma explícita en términos de las componentes de g_{ab} y sus derivadas parciales primera y segunda respecto a las coordenadas. Durante muchos años sólo se conocieron explícitamente unas pocas soluciones de la ecuación, pero recientemente se han encontrado muchas otras mediante

nuevos procedimientos de cálculo. La mayoría de ellas tienen un interés puramente matemático y no son aplicables a ningún escenario físico relevante. Sin embargo, de la naturaleza de las soluciones exactas se ha obtenido valiosa información relativa, en particular, a los cuerpos rotatorios, a los agujeros negros, a las ondas gravitatorias y a la cosmología.

En cualquier caso, sigue siendo difícil encontrar soluciones exactas que describan situaciones que nos puedan interesar. Entre las más notables está el denominado «problema de los dos cuerpos»: hallar una solución exacta de la ecuación de Einstein que describa, por ejemplo, dos estrellas orbitando una alrededor de otra. La dificultad estriba aquí en que, debido a la emisión de ondas gravitatorias, ambas estrellas trazarían una espiral hacia la otra, por lo que la situación no posee simetría. (La simetría suele ser de gran ayuda a la hora de resolver ecuaciones). No obstante, en la actualidad, la dificultad para encontrar soluciones exactas de sus ecuaciones no se considera una limitación de una teoría física. Con el advenimiento de los modernos ordenadores, a menudo los físicos obtienen información mucho más útil sobre una ecuación mediante una simulación numérica que a partir de una solución exacta explícita. Se han hecho esfuerzos considerables para desarrollar técnicas de computación aplicables a la relatividad general y ha habido ya grandes progresos en este área.

Algunos de los principales problemas asociados a la resolución de las ecuaciones de Einstein son de una naturaleza ajena a la mera

complejidad y provienen de un ingrediente específico de la relatividad general: el principio de covarianza general. Cuando se encuentra una solución, ya sea por métodos analíticos o computacionales, puede que no esté del todo claro lo que *significa*. Muchas características de la solución podrían reflejar simplemente ciertos aspectos de las coordenadas adoptadas, en vez de expresar algo de interés relativo a la física del problema. Se han desarrollado algunas técnicas para responder a estas cuestiones, pero aún queda mucho por hacer en este sentido.

Existe, por último, el profundo asunto de las *singularidades* en las soluciones de la ecuación de Einstein. Son puntos en los que la solución «diverge», dando por resultado un infinito en lugar de algo razonable desde el punto de vista físico. Durante muchos años existió una gran confusión en torno al tema, ya que tales singularidades pueden resultar «ficticias», es decir, ser simplemente el resultado de una inadecuada elección de coordenadas y no una característica verdaderamente singular del espacio-tiempo. El ejemplo más famoso de esta confusión tuvo lugar en relación con la célebre *solución de Schwarzschild* —la más importante de todas las soluciones de la ecuación de Einstein—. La solución describe el campo gravitatorio estático que rodea a una estrella con simetría esférica y fue hallada por Karl Schwarzschild en 1916, a punto de morir como consecuencia de una extraña enfermedad contraída en el frente oriental de la primera guerra mundial y en el mismo año en que Einstein publicó su primera formulación completa de la

relatividad general. Para un cierto radio, hoy conocido como *radio de Schwarzschild*, aparecía una singularidad en las componentes métricas y esta región del espacio-tiempo era denominada habitualmente «singularidad de Schwarzschild». A los científicos no les preocupaba demasiado esta singularidad, no obstante, ya que la región se hallaba en las profundidades de la estrella, lugar donde, debido a la presencia de una densidad de materia (el término T_{ab} de la ecuación de Einstein), la solución de Schwarzschild dejaba de ser válida. Pero en la década de 1960, el descubrimiento de los cuásares condujo a los astrónomos a preguntarse si podrían existir objetos estelares tan densos que su superficie exterior quedara dentro de su radio de Schwarzschild.

En realidad, ya en 1933 el sacerdote belga Georges Lemaître había demostrado que, mediante el cambio de coordenadas adecuado, la singularidad de Schwarzschild podía considerarse ficticia. Como consecuencia de ello, esta región ya *no* es considerada hoy como una singularidad, pero se denomina *horizonte de Schwarzschild* —el horizonte de sucesos de un agujero negro: todo cuerpo comprimido más allá de su radio de Schwarzschild colapsa inevitablemente hacia su centro, dando lugar a un objeto de estas características—. Ningún tipo de información puede escapar del interior del radio de Schwarzschild, motivo por el cual la región se denomina «horizonte».

Singularidades espaciotemporales

Llegados a este punto, me parece oportuno relatar cómo llegué a involucrarme profesionalmente en la relatividad general. A finales de

los años cincuenta, yo era un joven investigador del St. John's College de Cambridge. Mi área «oficial» de interés eran las matemáticas puras, pero un colega amigo mío, Dennis Sciama, se había propuesto mantenerme al día de las muchas cosas interesantes que estaban teniendo lugar en física y en astronomía. En su momento, había tenido cierto interés, como aficionado, por la relatividad general, una teoría cuya elegancia podía ser apreciada por alguien como yo, un simple amante de la geometría al que le atraían las ideas físicas novedosas. Aunque Dennis logró despertar en mí el interés por la física, no pensaba en la relatividad general como en algo a lo que pudiera dedicarme en serio, sobre todo porque la consideraba un tema secundario frente a los grandes asuntos de la física cuántica fundamental y el universo a escala microscópica.

Sin embargo, en algún momento de 1958, Dennis me convenció para que le acompañara a un seminario que David Finkelstein daba en Londres. Trataba sobre la extensión de la solución de Schwarzschild más allá del radio de Schwarzschild. Recuerdo que la conferencia me impresionó bastante, pero lo que más me intrigó fue el hecho de que, aunque la «singularidad» en el radio de Schwarzschild podía ser eliminada mediante un cambio de coordenadas, la singularidad en el *centro* (radio cero) seguía existiendo y no podía ser cancelada de esa manera. ¿Y si hubiera un principio subyacente que impidiese la eliminación total de singularidades para toda una clase de soluciones de la ecuación de

Einstein, a la cual pertenecería la de Schwarzschild?, pensé para mí.

Al volver a Cambridge, traté de reflexionar sobre el problema, pero mi formación era completamente inadecuada para abordarlo. En aquella época me hallaba trabajando en un formalismo conocido como *cálculo 2-espínor*, el cual es aplicable al estudio de las partículas cuánticas dotadas de espín. Mi trabajo puramente matemático me había llevado a estudiar álgebra de tensores de forma muy general y me habían fascinado los 2-espinores porque parecían ser, en cierto sentido, la raíz cuadrada de los vectores y los tensores. En cierta manera, los 2-espinores constituyen un sistema que es incluso más primitivo y universal en su descripción de las estructuras del espacio-tiempo que el proporcionado por los tensores. Así pues, traté de ver si el empleo de los espinores podía arrojar nueva luz sobre la relatividad general y si eran útiles para resolver el problema de la singularidad.

Aunque los espinores no me decían gran cosa acerca de las singularidades, encontré que encajaban extraordinariamente bien en la ecuación de Einstein, poniendo en evidencia algunos aspectos a los que no era fácil llegar por otros caminos. La elegancia de las expresiones resultantes era sorprendente, así que quedé enganchado. Durante los siguientes cuarenta y dos años, la relatividad general ha sido una de mis mayores pasiones, en particular en lo que se refiere a su afinidad con ciertas técnicas matemáticas poco frecuentes.

En 1964 volví a interesarme por las singularidades, debido en gran parte a que John A. Wheeler señaló que las observaciones recientes de lo que hoy conocemos como cuásares indicaban que había objetos astrofísicos reales cuyo tamaño se acercaba al radio de Schwarzschild. ¿Podía ser evitada la singularidad que aparece cuando un cuerpo colapsa *más allá* de ese radio —la singularidad central que me había preocupado a raíz del discurso de Finkelstein?—. La solución exacta correspondiente a tal colapso (denominada actualmente «agujero negro»), encontrada por Oppenheimer y Snyder en 1939, presenta, en efecto, una singularidad genuina en su centro. Pero una premisa fundamental en su modelo era la existencia de simetría esférica exacta. Se podría pensar que, de haber irregularidades, la materia no caería hacia un punto de densidad infinita ubicado en el centro, sino que pasaría a través de una configuración central más o menos complicada y sería despedida de nuevo hacia el exterior, con lo que no se produciría una singularidad real.

Mis antiguos temores de que dichas singularidades quizá fuesen inevitables me habían llevado a dudar de esa posibilidad, así que empecé a preguntarme si algunas ideas sobre las que había estado trabajando últimamente —en esencia, consideraciones cualitativas de tipo topológico, en lugar de buscar directamente una solución exacta para la ecuación de Einstein— podrían resolver el problema. Más adelante, esta heterodoxa línea de razonamiento me condujo hasta el primer «teorema de la singularidad», el cual mostraba que,

partiendo de ciertos supuestos generales muy razonables y sin que haya que adoptar premisa alguna sobre simetrías, *todo* colapso gravitatorio en el interior de una región cualitativamente similar al radio de Schwarzschild da lugar a una singularidad espaciotemporal verdadera.

Este resultado se generalizó en posteriores trabajos realizados por Stephen Hawking y otros que realizamos conjuntamente, en los cuales se demostraba que, además de en el escenario del agujero negro, estas singularidades también son inevitables en el *Big Bang* que dio origen al universo, independientemente de cualquier consideración respecto a las simetrías. Los modelos cosmológicos estándares derivan de las soluciones originales de la ecuación de Einstein que el matemático ruso Alexander Alexandrovich Friedmann halló en 1922. En ellas se supone homogeneidad espacial e isotropía exactas y la solución se expande a partir de la singularidad inicial del *Big Bang*. Lo que los teoremas de la singularidad demuestran es que no es posible eliminar esa singularidad inicial prescindiendo simplemente de las premisas de homogeneidad e isotropía.

Todo ello presupone la validez de la ecuación de Einstein (y de algunos supuestos razonables relativos a T_{ab}). Hay quien opina que esos teoremas ponen de evidencia un fallo de base en la relatividad general. Mi punto de vista personal es otro. Sabemos que, en cualquier caso, la teoría de Einstein no puede ser la última palabra acerca de la naturaleza de la gravedad y el espacio-tiempo. En algún

momento tendrá que llegar la necesaria conciliación de la relatividad general con la mecánica cuántica. Lo que los teoremas de la singularidad revelan es, en realidad, una virtud de la teoría de Einstein: la de señalar claramente cuáles son sus limitaciones, la de decirnos que debemos buscar cómo extenderla al mundo cuántico y la de anticiparnos algo de lo que cabe esperar de esa unión final entre lo cuántico y lo gravitatorio. Trataremos de vislumbrarlo en la próxima sección.

El principio y el fin del tiempo

En la discusión anterior hemos citado dos escenarios en los que surgen singularidades espaciotemporales en la teoría de Einstein: el colapso gravitatorio que conduce a un agujero negro y la explosión que supuestamente dio origen al universo. Al parecer, a Einstein le disgustaba profundamente la existencia de esas dos manchas en su teoría. Estaba convencido de que una modificación realista de las condiciones de alta simetría asumidas en las soluciones exactas estándares debería llevar a soluciones *no* singulares. Desgraciadamente, nunca sabremos su opinión sobre los teoremas de la singularidad a los que nos hemos referido, pero parece ser que una de las razones por las que Einstein luchó denodadamente en sus últimos años para extender su relatividad general hacia alguna clase de «teoría del campo unificado» fue precisamente tratar de llegar a una teoría libre de singularidades.

Al principio se inclinó por un universo espacialmente cerrado y *estático* —sin cambios a lo largo del tiempo—. Llegó entonces a la conclusión (en 1917) de que sólo podía obtenerlo si introducía una *constante cosmológica*, Λ , en su ecuación, la cual se transformaba en:

$$R_{ab} - \frac{1}{2}R g_{ab} + \Lambda g_{ab} = -8\pi G T_{ab}$$

Más tarde consideraría esta modificación «la mayor pifia de su vida». Si no se hubiese aferrado al modelo estático y hubiera dejado la ecuación tal cual era en su origen —permitiendo obtener el modelo de Friedmann de un universo que se expandía desde un *Big Bang* inicial—, Einstein habría podido predecir la expansión del universo, constatada experimentalmente en 1929 por Edwin Hubble.

Existe hoy en día una gran controversia acerca de si las observaciones acumuladas sugieren la existencia de una (diminuta) constante cosmológica. Algunos cosmólogos (en especial, los partidarios de lo que se denomina el «universo inflacionario») afirman que esta constante es necesaria para explicar las observaciones recientes. En cualquier caso, las contradicciones en los datos aconsejan esperar algún tiempo antes de adoptar una postura definitiva al respecto.

En mi opinión, aunque hay que ser prudentes a la hora de hacer afirmaciones basadas en las observaciones del universo a gran escala, debemos aceptar, en cambio, que el *Big Bang* y las

singularidades de los agujeros negros forman parte de la naturaleza. En vez de sobrecogernos, tenemos que tratar de aprender de ellas algo sobre la «geometría cuántica» que debería finalmente reemplazarlas. ¿Qué es posible aprender? Aunque, en detalle, se sepa poco sobre las singularidades, podemos hacer algunas consideraciones generales.

La primera es que, aunque en ciertos casos (tales como el de una estrella supermasiva o un conjunto de estrellas en el centro de una galaxia) el colapso gravitatorio sea inevitable, no es seguro que el resultado final sea un agujero negro, a pesar de que los teoremas nos digan que debemos esperar tales singularidades espaciotemporales. Existe una hipótesis aún no demostrada, conocida como de la «censura cósmica» y que formulé en 1969, según la cual la singularidad resultante no puede estar «desnuda» (es decir, ser «visible desde el exterior»). Si no pueden darse las singularidades desnudas, el resultado será, en efecto, un agujero negro. (En cualquier caso, las singularidades desnudas serían «aún peores» que los agujeros negros). Un agujero negro engulle materia situada en su proximidad inmediata y (suponiendo que existe la censura cósmica) la destruye en la singularidad de su centro. Para la materia que cae a su interior, esa singularidad representa el «fin del universo», y equivale a un *Big Bang* invertido en el tiempo.

A pesar de ese panorama tan desagradable, el espacio-tiempo *exterior* a un agujero negro posee gran cantidad de propiedades interesantes. Por otra parte, en el centro de casi todas las galaxias

parece haber agujeros negros de gran tamaño y la extraordinaria física a la que dan lugar en sus inmediaciones sería la responsable de la gigantesca emisión de energía de los cuásares, cuya magnitud puede superar a la de galaxias enteras. Los agujeros negros son también las regiones de más alta entropía conocida en el universo y una famosa fórmula debida a Bekenstein y Hawking nos dice exactamente cuál es esa entropía a partir del área de la superficie del horizonte del agujero.

En la práctica, la censura cósmica viene a afirmar que en el universo sólo hay dos clases de singularidades espaciotemporales: la de tipo *pasado* (en el *Big Bang*) y la de tipo *futuro* (en los agujeros negros). La materia se crea en la singularidad de tipo pasado y se destruye en las de tipo futuro. A primera vista, los dos tipos de singularidad parecen ser simplemente el inverso temporal uno de otro. Sin embargo, si se analizan en detalle, existen grandes diferencias entre ambos, debidas fundamentalmente a la gigantesca entropía de los agujeros negros. En términos cotidianos, «entropía» es sinónimo de «desorden» y la famosa segunda ley de la termodinámica nos dice que la entropía del universo se incrementa con el tiempo. Según esto, el origen físico de esa segunda ley puede ser atribuido a la gran asimetría que hay entre las estructuras de las singularidades de tipo pasado y tipo futuro: las primeras son particularmente especiales y simples, mientras que las segundas son generales y extraordinariamente complejas. Mediante la fórmula de Bekenstein-Hawking para la entropía de un agujero negro, se

puede concluir que el *Big Bang* fue tan increíblemente especial como una parte en al menos 10 elevado a 10^{123} .

¿Gravitación cuántica?

¿De dónde proviene esta inmensa asimetría temporal en la estructura de las singularidades espaciotemporales? El tema sigue siendo fuente de controversias pero, en mi opinión, tiene que ver con el hecho de que la «gravitación cuántica», que en su momento deberá explicar la naturaleza detallada de las singularidades espaciotemporales, ha de ser asimétrica en el tiempo. Me sorprende el que pocos de los investigadores en gravitación cuántica parezcan haber llegado a la conclusión de que, cualquiera que sea la naturaleza de esta teoría aún inexistente, deberá obedecer a un esquema temporal fundamentalmente asimétrico. Es cierto que la ecuación de Einstein es simétrica respecto al tiempo y también lo es la ecuación de Schrödinger, que gobierna la evolución de un estado cuántico. Según esto, cualquier aplicación «convencional» de las leyes de la mecánica cuántica a la teoría de Einstein debería llevar a conclusiones *simétricas* en el tiempo. En mi opinión, de esto se deduce claramente que la ansiada «gravitación cuántica» tendrá que ser una teoría cuántica *no convencional*, según la cual es probable que las propias leyes de la mecánica cuántica deban *cambiar*, sin contar los cambios que también se produzcan en las leyes clásicas de la relatividad general einsteniana. En este sentido, estoy de acuerdo con Einstein en que la mecánica cuántica es incompleta.

Sin embargo, no es ésta la posición de la gran mayoría de quienes tratan de combinar la teoría cuántica con la relatividad general. A pesar de que hay muchas ideas fascinantes e inéditas propuestas como candidatas a teoría de la «gravitación cuántica» —tales como los «espacio-tiempos» de diez, once o veintiséis dimensiones o los conceptos de supersimetría, cuerdas, etc.—, ninguna de ellas contempla la posibilidad de que las propias leyes de la mecánica cuántica quizá tengan que cambiar. Mi punto de vista (y el de una importante minoría de investigadores de los fundamentos de la mecánica cuántica) es que los cambios en la teoría cuántica se van a producir en cualquier caso, debido al denominado «problema de la medida».

¿En qué consiste ese problema? Para responder a la pregunta necesitamos hablar un poco sobre las leyes de la teoría cuántica. Existe una variable matemática denominada *estado cuántico* (o función de onda), que habitualmente se representa como Ψ , y se supone que contiene toda la información necesaria para definir un sistema cuántico dado. La evolución en el tiempo del estado viene determinada por la ecuación de Schrödinger hasta que efectuamos una medida sobre el sistema, momento en el que el estado *salta* (aleatoriamente) a una de las posibles opciones permitidas por la medida específica que estamos realizando. Ese «salto» no obedece a la ecuación de Schrödinger y el problema de la medida consiste en comprender cómo se produce, ya que se supone que el estado evoluciona realmente según dicha ecuación determinista.

Creo razonable la hipótesis de que la ecuación de Schrödinger tal cual está formulada *no* es rigurosamente aplicable a todas las escalas y requiere una modificación cuando los efectos gravitatorios son significativos. Según esto, dicha modificación formaría parte necesariamente de la «teoría de la gravitación cuántica correcta». Una de las principales razones para pensar así proviene de la probable existencia de un conflicto fundamental entre el principio de covarianza general y los principios básicos de la evolución de la función de onda de Schrödinger. Según este razonamiento, el salto cuántico (que para mí es un fenómeno físico *real* y no una «ilusión», como se supone a menudo) entra en juego como un aspecto de la resolución de dicho conflicto.⁵⁰ En definitiva, cualquiera que sea la modificación a introducir en la ecuación de Schrödinger, tendría que ser asimétrica en el tiempo, dando lugar a esa enorme asimetría entre las singularidades pasadas y futuras a que me he referido.

Hasta la fecha no se ha encontrado modificación alguna plausible de la ecuación de Schrödinger, con lo que la unificación de la teoría cuántica con la relatividad general por esa vía continúa siendo tan esquiva como lo es a través de otras, más convencionales, que los físicos han estado ensayando. Lograr esa unificación se plantea como uno de los más grandes retos científicos para el siglo XXI. El hallazgo tendrá consecuencias profundas, mucho más allá de las que somos capaces de entrever en este momento. Para alcanzarlo,

⁵⁰ He propuesto un experimento, técnicamente difícil pero en apariencia factible, para comprobar si mi propuesta es correcta. Una de sus posibles implementaciones debería desarrollarse en el espacio exterior.

en cualquier caso, tendremos que desentrañar los raros y maravillosos principios que subyacen bajo la bella ecuación de Einstein.

Lecturas recomendadas

W. Rindler, *Relativity: Special, General and Cosmological*, Oxford University Press, 2001.

W. Rindler, *Essential Relativity*, Nueva York, Springer-Verlag, 1997.

L. A. Steen, ed., *The Geometry of the Universe*, Mathematics Today, Twelve Informal Essays, Nueva York, Springer-Verlag, 1978.

K. Thorne y C. W. W. Norton, *Black Holes and Time Warps: Einstein's Outrageous Legacy*, Nueva York, Springer-Verlag, 1994.

A. Einstein, *Relativity: The Special and the General Theory* (Reimpresión de Three Rivers Press, California, 1995).

Parte 4

Erótica, estética y la ecuación de onda de Schrödinger

Arthur I. Miller

Un buen amigo de Erwin Schrödinger relataba que éste «*había realizado su gran obra durante un tardío arrebatado amoroso*». El acontecimiento tenía lugar en las Navidades de 1925, cuando el físico vienés de treinta y ocho años pasaba sus vacaciones con una antigua novia en la estación de esquí suiza de Arosa, cerca de Davos. La pasión fue el catalizador de una explosión de actividad creadora que duraría todo un año. Como en el caso de la dama desconocida que inspiró los sonetos a Shakespeare, la identidad de esa mujer hoy día es un misterio —aunque probablemente no lo fuera para la esposa del físico, que solía estar al tanto de todas sus infidelidades conyugales—. Tal vez debamos a esa dama misteriosa el extraordinario hecho de que varias líneas de investigación, aparentemente inconexas, se fundieran y Schrödinger descubriera la ecuación que lleva su nombre.

En su forma, al menos, la ecuación de Schrödinger era ya conocida por muchos científicos y su presencia resultaba casi reconfortante ante el continuo ataque a los conceptos tradicionales llevado a cabo por los nuevos físicos cuánticos. Esta largamente esperada expresión de la física cuántica fue formulada por primera vez por su contrariado descubridor, Max Planck, en 1900 y refinada después por Albert Einstein y Niels Bohr, entre otros. En esencia, la

ecuación venía a ser para el mundo subatómico lo que las leyes de Newton eran para el universo a gran escala desde hacía siglos: permitía a los científicos hacer predicciones detalladas sobre el comportamiento de la materia, a la vez que se visualizaban los sistemas atómicos objeto de estudio. Mediante la ecuación de Schrödinger se podía entender por primera vez la estructura atómica en detalle, allí donde las ecuaciones de Newton perdían todo sentido.

Seis meses antes de la explosión creativa de Schrödinger se había encontrado una nueva física cuántica para el átomo. El autor del descubrimiento era Werner Heisenberg, un joven y brillante teórico alemán perteneciente a la Universidad de Gotinga. A sus veinticuatro años, Heisenberg había propuesto un enfoque completamente distinto de la física atómica, soportado por una matemática inusual y difícil que no permitía visualizar los procesos atómicos ni ofrecía ecuaciones análogas a las de Newton para los sistemas clásicos. De hecho, uno de los motivos que llevaron a Schrödinger a formular su versión de la física atómica fue el rechazo a la propuesta de Heisenberg. Schrödinger llegó incluso a demostrar que ambas versiones eran equivalentes —¿cuál era, pues, la mejor?—. Ni qué decir tiene que tanto Schrödinger como Heisenberg fueron defensores acérrimos de sus respectivas propuestas y sumamente críticos con la del rival.

Pero existe una paradoja. Aunque, a primera vista, la ecuación de Schrödinger era fácil de utilizar, involucraba un concepto,

denominado función de onda, que era extremadamente difícil de interpretar e imposible de observar directamente. Heisenberg discrepaba abiertamente de la interpretación que Schrödinger daba a la función de onda como representación de la «nube» de carga asociada a un electrón que gira alrededor del núcleo atómico. La áspera controversia surgida aún no está resuelta en la actualidad. El propio Schrödinger nunca estuvo satisfecho con la interpretación más extendida del significado de la función de onda.

En el presente ensayo quisiera explorar el modo en el que la interpretación de Heisenberg llegó a prevalecer sobre la de Schrödinger, a pesar de que el *método* de este último, adecuadamente reinterpretado, sustituyó al del primero en casi todas las áreas de la teoría física. Los temas sometidos a intenso debate en aquella época —cómo visualizar el comportamiento del átomo y si éste se puede explicar en términos de probabilidad exclusivamente— todavía colean hoy. Desde un punto de vista práctico, sin embargo, la teoría cuántica ha demostrado ser extraordinariamente fructífera. Ha puesto las bases de nuestro conocimiento del mundo microscópico, permitiendo a los técnicos desarrollar transistores, microprocesadores, láseres y cables de fibra óptica cada vez más eficientes. La piedra angular de esta teoría es la ecuación de Schrödinger, la cual se ha convertido en una herramienta de investigación habitual para los científicos de todo el mundo.

Nacido en 1887 en Viena, la capital cultural y política del imperio austrohúngaro, Schrödinger asistía a un *gymnasium* (instituto) que hacía un especial énfasis en el estudio de los clásicos griegos y latinos. Schrödinger aprendió también por su cuenta inglés y francés.⁵¹ Tuvo un excelente expediente escolar y se le consideró un alumno superdotado. Esta amplia y profunda educación sirvió para inculcar en él un enorme respeto por la tradición clásica. Su libro *La Naturaleza y los griegos*, publicado en 1948, es una elegante exposición de las antiguas teorías físicas y de su importancia. Schrödinger tuvo también gran interés por la filosofía, lo que le llevó a una lectura más que anecdótica de textos orientales, tales como el Vedanta, sobre el que escribió en 1925 en *Buscando el camino*, una intensa e íntima confesión de sus creencias. Influido por el hinduismo, el ensayo aboga por el carácter unitario de la consciencia humana y por la unión entre humanidad y naturaleza. No sería publicado hasta 1961, un año antes de su muerte, formando parte del volumen titulado *Mi concepción del mundo*.

Aunque Heisenberg asistió también a un *gymnasium* y tenía talento para la música y la filosofía, su personalidad difería radicalmente de la de Schrödinger, que era catorce años mayor que él y tenía un carácter más conservador.⁵² Heisenberg se sentía como pez en el agua en las situaciones de cambio. No en vano alcanzó la madurez en uno de los periodos más turbulentos de la historia alemana,

⁵¹ Para más detalles biográficos, véase: W. Moore, *Schrödinger: Life and Thought*, Cambridge University Press, 1989.

⁵² Véase: D. Cassidy, *Uncertainty: The Life and Science of Werner Heisenberg*, Nueva York, Freeman, 1992.

entre la derrota en la primera guerra mundial, el colapso de la monarquía y la revolución que se extendía por todo el *Reich*. Al igual que Schrödinger, Heisenberg provenía de una familia culta; tocaba el piano casi al nivel de un concertista. La música desempeñó un papel importante en su vida, a diferencia de Schrödinger, que no tuvo interés alguno por ella. Compartían, en cambio, la energía y el espíritu joven, cualidades que mantuvieron hasta avanzada edad.

En 1906 Schrödinger ingresó en la Universidad de Viena, donde tuvo excelentes profesores. Creció intelectualmente en esa atmósfera, ampliando sus conocimientos de física y añadiendo el interés por la biología, a la que aportaría cuarenta años después algunas ideas radicales en su breve pero trascendental libro *¿Qué es la vida?* (James Watson, descubridor de la estructura del ADN junto a Francis Crick, hablaba de él como fuente de inspiración).

Por aquella época, el altamente desarrollado instinto erótico de Schrödinger comenzaba a aflorar. Sus planteamientos al respecto diferían del objetivo machista tradicional: en lugar de tratar de dominar a la mujer, su filosofía se orientaba a explorar la esencia de la sensualidad femenina. Mantenía un cuaderno de bitácora en el que anotaba comentarios, nombres y fechas de los encuentros, su *Ephemeridae*. Como el artista de vanguardia Gustav Klimt, Schrödinger siempre intentaba «atrapar la sensación de feminidad». Cabe imaginar que su calculada naturalidad en la vestimenta y en la apariencia, su frente despejada, su pelo cuidadosamente peinado y su intensa mirada, unidos a una cultura aparentemente

inagotable, debían resultar muy atractivos para las mujeres. A pesar de su porte burgués y correcto, había siempre algo byroniano en él. Al igual que otros compatriotas vieneses, tales como Ludwig Wittgenstein, Schrödinger tuvo un papel activo en la primera guerra mundial, sirviendo con distinción en el frente italiano en una unidad de artillería del ejército austrohúngaro. Schrödinger fue citado por su valentía ante el fuego enemigo en octubre de 1915 en el curso de una de las sangrientas batallas que Ernest Hemingway hizo famosas en *Adiós a las armas*. Poco después fue ascendido a teniente y terminaría la guerra en Viena en el tranquilo puesto de profesor de meteorología elemental para oficiales del ejército, desde el que publicaría artículos sobre la teoría de gases y la relatividad general.

En 1925, Schrödinger era la antítesis de los jóvenes e insolentes científicos recién llegados a la mecánica cuántica. Hasta su elegante forma de vestir contrastaba con la de Heisenberg, de quien se dice que parecía «un muchacho de pueblo, con el pelo corto, los ojos claros y brillantes y una expresión encantadora».⁵³ Comparado con Heisenberg y con su colega y confidente, el hipercrítico y áspero Wolfgang Pauli, Schrödinger era ya todo un personaje al frente de una cátedra en la Universidad de Zurich.

La formación universitaria de Heisenberg fue excepcional. Casualmente, al ingresar en la Universidad de Munich, durante el semestre invernal de 1920-1921, le correspondió como profesor de

⁵³ *Ibid.*, pág. 137, tal como lo describe Max Born.

física atómica dentro del ciclo de física teórica el famoso físico Arnold Sommerfeld. De este modo, Heisenberg se enfrentó directamente a la investigación más avanzada en ese campo. A menudo decía que había aprendido física al revés, empezando por la física atómica antes de estudiar la de Newton, la cual se considera imprescindible para abordar temas más avanzados.

La teoría atómica en vigor en la época había sido formulada en 1913 por el físico danés de veintisiete años Niels Bohr. Bohr buscó obsesivamente la claridad durante toda su vida, sometiendo sus ideas a discusiones profundas y críticas con sus colegas y alumnos. Por esta razón, muchos de sus artículos científicos son crípticos; reelaborados una y mil veces, la omisión de una sola palabra puede alterar por completo su significado. Cuando mejor funcionaba su mente era conversando, cuando tenía alguien en quien «rebotar» sus ideas. En 1913 era un joven apresurado, en buena forma gracias al fútbol que practicaba. Diez años después comenzaría a acusar en su aspecto el peso de los problemas que había cargado sobre sus hombros, la búsqueda del significado de una nueva física, una física que desafiaba todas las reglas sobre lo que se supone debía ser una teoría.⁵⁴

La teoría atómica de Bohr de 1913 es recordada hoy por su metáfora del átomo como un minúsculo sistema solar. Era un magnífico pastiche de la mecánica celeste newtoniana con algunas dosis de la teoría de la radiación de Planck. La aplicación de la

⁵⁴ Para detalles biográficos, véase: A. Pais, *Niels Bohr's Times: In Physics, Philosophy and Polity*, Oxford University Press, 1991.

teoría de Newton permitió a Bohr trasladar el modelo del físico inglés al dominio atómico. Aquí, los electrones se ven constreñidos a seguir determinadas órbitas alrededor de su Sol central, el núcleo. Esas órbitas permitidas se denominan estados estacionarios o niveles de energía. Consideremos el átomo de hidrógeno, el más simple de todos, ya que consta de un único electrón ligado a un núcleo cargado positivamente. Según la teoría de Bohr, ese electrón sólo puede existir en determinadas órbitas. La órbita permitida más baja —la más cercana al núcleo— corresponde al estado fundamental del átomo. Una consecuencia chocante de la teoría de Bohr es que, cuando se halla en un estado permitido, el electrón se limita a estar «posado» como un pájaro en un árbol, no haciendo otra cosa que «esperar». Según la teoría electromagnética aceptada entonces —combinada con la mecánica de Newton—, el electrón debía orbitar en torno al núcleo como un planeta alrededor del Sol. De acuerdo con las leyes físicas tradicionales, el electrón en órbita estaría emitiendo energía de radiación continuamente. Debido a ello, perdería energía para acabar cayendo en espiral hacia el núcleo. El resultado es que la materia sería totalmente inestable. El lector sabe de sobra que no es así, ya que, por ejemplo, está leyendo el presente ensayo en lugar de hacer explosión. Explicar la estabilidad del átomo era considerado un asunto clave en aquel momento. Bohr fue lo suficientemente perspicaz como para darse cuenta de que el problema era, por el momento, insoluble y de que el hecho debía ser aceptado sin más. Por ello, postuló la existencia

de una órbita o estado estacionario fundamental, en el cual el electrón no cae hacia el núcleo ni emite energía alguna —y no se admiten preguntas.

Iluminando el átomo con luz, por ejemplo, es posible excitar el electrón hasta una órbita permitida más alta. Una vez allí, el electrón vuelve a ser como el pájaro en el árbol, a la espera solamente de descender al estado fundamental, estado que finalmente alcanzará, ya sea directamente o mediante transiciones entre estados. Esas transiciones no son suaves, sino discontinuas y se denominan saltos cuánticos. Al realizarlos, el electrón emite radiación en ráfagas —es decir, de forma discontinua—. Uno de los mayores éxitos de la teoría de Bohr fue su capacidad para predecir la longitud de onda de la radiación emitida por el átomo de hidrógeno con un margen de error de un uno por ciento respecto a los valores observados por los experimentadores. La teoría predijo también y con similar precisión longitudes de onda correspondientes a transiciones aún no observadas en aquella época.

La teoría de Bohr causó un gran revuelo en la comunidad científica. Un eminente físico como Max Born —que, más tarde, se mostraría menos drástico— llegó a afirmar que la teoría no era sino «un truco de magia para la mente; en realidad, se basa en la superstición (tan antigua como la historia del pensamiento) de que el destino del hombre está escrito en las estrellas». Einstein, en cambio, saludó de inmediato la teoría como «un gigantesco logro».

En 1925, no obstante, la situación en la física atómica se había vuelto tremendamente confusa. La impresión general entre los físicos era que la teoría de Bohr constituía un callejón sin salida. No podía explicar con precisión más allá de escenarios simples basados en el átomo de hidrógeno. Hacia 1923, los datos que comenzaban a acumularse sobre la interacción de los átomos con la luz parecían sugerir que aquéllos no se comportaban en la práctica como diminutos sistemas solares.

Los físicos improvisaron rápidamente una versión híbrida de la teoría de Bohr que pudiera resolver provisionalmente la situación. En ella no se trataba de visualizar lo que sucedía a nivel subatómico, sino que se asumía que los átomos podían de alguna manera perder energía mediante una transición entre un nivel de energía dado y otro comparativamente más bajo —por medio de un «salto cuántico»—. Del mismo modo, el átomo podía ganar energía saltando desde un nivel a otro más alto. En estos procesos, la energía cedida o adquirida era transportada por una ráfaga de luz, correspondiente a una radiación de una longitud de onda concreta. Esto explicaba por qué los átomos emiten y absorben radiación en longitudes de onda específicas, conocidas como líneas espectrales. Otra característica importante de esta teoría improvisada era la novedosa idea de que no es posible predecir exactamente cuándo se producen los saltos cuánticos —sólo cabía establecer la probabilidad de que tuvieran lugar en un instante determinado—. Bohr adoptó este «probabilismo», que llegaría a ser la piedra angular

del pensamiento cuántico, partiendo de una idea exitosa que Einstein había propuesto en 1916, en el marco de su teoría sobre la interacción entre la radiación y el átomo. Esos tres aspectos de la teoría mejorada —probabilismo, saltos cuánticos y no visualizabilidad— permitieron que la teoría se mantuviera en pie hasta comienzos de 1925, momento en que terminó por venirse abajo.

Los físicos interpretaban las probabilidades como un signo de que los mecanismos de un proceso concreto no se conocían en su totalidad. Pensaban que, tarde o temprano, el mecanismo por el que los electrones realizan las transiciones en los átomos llegaría a comprenderse y se podría formular una nueva versión, desconocida aún, de la mecánica newtoniana. Al final, las cosas se calcularían como siempre y las probabilidades se harían innecesarias. Resultó que éste no era el caso. Aunque la versión modificada de la teoría de Bohr acabó fallando, sirvió de punto de partida a Heisenberg para su nueva y drástica teoría atómica, basada en electrones invisibles y discontinuidades radicales. En sus cimientos se hallaba una matemática extremadamente difícil de aplicar. Ni siquiera el propio Heisenberg, en su primer artículo sobre mecánica cuántica, sabía cómo hacerlo.

Había ido a parar a unos objetos matemáticos denominados matrices, movido por su interés por contabilizar todas las transiciones atómicas posibles entre estados estacionarios. Las matrices proporcionan una manera natural de realizarlo y, además,

suministran herramientas para calcular las características de las líneas espectrales. Para ser algo más preciso: las matrices son disposiciones de números en filas y columnas y en mecánica cuántica cada elemento representa una posible transición atómica, perdiendo o ganando energía. Mediante un procedimiento matemático bien conocido es posible calcular las energías del átomo. Son los denominados valores propios de la matriz y su cálculo suele resultar bastante arduo. A Wolfgang Pauli, uno de los más grandes calculistas de la época, le llevó más de cuarenta páginas deducir los niveles de energía de un simple átomo de hidrógeno a partir de la teoría Heisenberg. A finales de 1925, algunos problemas que aguardaban solución desde hacía mucho tiempo habían quedado resueltos por Heisenberg y sus colaboradores, que habían eludido la teoría de Bohr. La versión «matricial» de la mecánica cuántica parecía prometer grandes avances.

La educación híbrida de Heisenberg fue, sin duda, una de las causas de su desafiante y revolucionaria aproximación a la física atómica. Menos de un año después de entrar en la universidad escribió su primer artículo, en el cual renunciaba a la seguridad de las reglas que trasladaban los resultados de la física de Newton a la física cuántica —el método aceptado entonces— y adoptaba un modelo ya en cierto modo de acuerdo con las ideas cuánticas. Como uno de sus colegas diría después, «una maravillosa combinación de

profunda intuición y virtuosismo formal inspiró a Heisenberg concepciones de una enorme brillantez».

En aquellos momentos, Schrödinger perseguía, como era su costumbre, gran variedad de intereses. Junto a investigaciones sobre la relatividad general, llevaba estudiando desde 1917 la percepción de los colores. Además, se interesaba por ciertos temas relativos al sonido y a los medios elásticos, lo que le condujo a investigar la teoría de ondas, que en breve tiempo le iba a resultar tan útil.

En el terreno personal, Schrödinger vivía en Zurich con Annemarie, su esposa desde hacía cinco años, conocida cariñosamente como Anny. Vivían en una versión suavizada y «a la suiza» de la cultura bohemia de Weimar que tanto escandalizaba a los alemanes conservadores y nacionalistas: el tormentoso, sexual y ambiguo mundo que solemos asociar con Marlene Dietrich y el arte expresionista. La reacción violenta a ese ambiente liberal la personificaban Hitler y su Partido Nazi, cuyo imparable ascenso hizo reconsiderar a Schrödinger, en 1927, la idea de abandonar Zurich y aceptar el puesto de Planck en Berlín. El matrimonio estaba ensombrecido, por una parte, por la incapacidad de Anny para concebir el hijo que tanto deseaba Erwin y, por otra, por las continuas infidelidades del físico. Constituían una singular pareja. Anny había sacrificado sus intereses intelectuales para dedicarse en cuerpo y alma a su marido. Cuando la pasión se enfrió —lo cual tuvo lugar cuando apenas llevaban un año de casados— cada uno

buscó el sexo por su lado, si bien permanecieron unidos tratándose el uno al otro como amigos. Como Anny comentaría años más tarde, «sé que es más fácil vivir con un canario que con un caballo de carreras, pero yo prefiero el caballo». Schrödinger no tuvo un solo amigo íntimo (varón) en toda su vida. Su gusto en el vestir y su intensa vena romántica se alimentaron también de su pasión por el teatro.

A partir de argumentos basados en la teoría de la relatividad, Louis de Broglie sugirió en 1923 que los electrones podían ser también ondas, en contra de la opinión general que los consideraba exclusivamente partículas. Einstein reconoció la importancia de las observaciones de De Broglie y, a su vez, basó en ellas sus investigaciones sobre la teoría de gases. Einstein estaba entusiasmado y escribió a un colega que De Broglie «había desvelado una esquina del gran secreto». Pero De Broglie y Einstein sólo representaron una parte de la inspiración que animó a Schrödinger en su orgía creativa, tal como explicaba en el tercero de los artículos publicados en la primavera de 1926:

«Mi Teoría se ha inspirado en el artículo de L. De Broglie publicado en *Ann. de Physique* (10) 3, pág. 22, 1925 (Thèses, París, 1924) y en comentarios breves e incompletos de A. Einstein, publicados en *Berl. Ber.* (1925) pp. 9 y sig. Hasta donde conozco, no tengo relación genética alguna con Heisenberg. Conozco su teoría, por supuesto, pero producen cierto rechazo en mí (por no decir repulsión) sus

métodos de álgebra trascendental, que me parecen hartamente difíciles, y la ausencia de visualizabilidad». ⁵⁵

El sentido estético al que Schrödinger alude indirectamente aquí es una preferencia por una matemática más asequible y menos desagradable que el «álgebra trascendental» (las matrices) de Heisenberg y que, además, permita visualizar los procesos atómicos. Este punto se clarifica a continuación.

En clave más objetiva, una de las principales críticas de Schrödinger a la mecánica cuántica de Heisenberg es que le parecía «extraordinariamente difícil» abordar procesos tales como los fenómenos de colisión desde la óptica de una «teoría del conocimiento» en la que «se renuncia a la intuición y se opera sólo con conceptos abstractos, tales como probabilidades de transición, niveles de energía y similares». De hecho, en la formulación realizada por Heisenberg durante 1925-1926 solamente era posible calcular niveles de energía del átomo, es decir, trabajar sólo con electrones ligados a átomos. Por otra parte, el concepto de abstracto es relativo: Bohr, Heisenberg y Pauli consideraban que los niveles de energía «y similares» eran completamente concretos. En 1926, Schrödinger admitía la existencia de «cosas» que no pueden ser comprendidas mediante nuestras «formas de pensar», y que pueden no tener una descripción espaciotemporal newtoniana pero, «desde

⁵⁵ E. Schrödinger, «Über das Verhältnis der Heisenberg-Born-Jordanschen Quantenmechanik zu der meinen», *Annalen der Physik*, 1926, 70, págs. 734-756. Este artículo es conocido como la tercera comunicación.

el punto de vista filosófico», estaba seguro de que «la estructura del átomo» no pertenecía a esta categoría.⁵⁶

En cualquier caso, al igual que Heisenberg y otros físicos de la época, Schrödinger se dio cuenta de que los modelos visuales tomados tal cual del mundo de la percepción sensorial no eran adecuados. Prescindiendo por completo de ellos, Heisenberg basó su mecánica cuántica en partículas que no se podían visualizar. Schrödinger buscó un modo de visualizar electrones que fuera distinto del que los científicos se habían acostumbrado a utilizar cuando pensaban en el átomo, es decir, como partículas. Se dio cuenta de que ese nuevo modo podía derivarse de las observaciones de De Broglie y decidió explorarlo. Tal vez era una simple cuestión de estética, pero se trata de una opción más a la hora de construir una teoría. Partiendo de la revolucionaria idea de De Broglie, según la cual los electrones pueden ser ondas a la vez que partículas, Schrödinger aplicó la hipótesis a los electrones del átomo.

La idea básica de Schrödinger fue formular una teoría en la que los electrones atómicos semejaban cuerdas vibrantes sujetas por ambos extremos. El modo en el que vibra la cuerda es un indicador de la energía del electrón. Este tipo de teoría de ondas evita también los saltos cuánticos. Las transiciones atómicas tenían lugar en la forma de ondas que representaban la densidad de carga del electrón rodeando el núcleo y reduciendo o incrementando su radio para pasar de un estado permitido a otro.

⁵⁶ Las citas de este párrafo pertenecen a la segunda comunicación de Schrödinger, «Quantisierung als Eigenwertproblem», *Annalen der Physik*, 1926, 80, págs. 437-490.

Examinemos una particularización de las ideas de Schrödinger para el átomo más simple de todos, el constituido por un único electrón que órbita en torno al núcleo —el átomo de hidrógeno—. A modo de experimento mental, consideremos ese electrón como una cuerda sujeta por ambos extremos, es decir, ligada al átomo de hidrógeno. Cuando la cuerda vibra con la energía mínima como si fuera una onda estacionaria, hay exactamente media longitud de onda entre los extremos. Para el siguiente nivel de energía habrá dos medias longitudes de onda, en el que viene a continuación, tres, y así sucesivamente. La idea es que cada configuración de la cuerda vibrante corresponde a una energía concreta, o valor propio, de la cuerda.

Al aplicar la ecuación de Schrödinger al átomo de hidrógeno obtenemos el mismo tipo de relación entre niveles de energía y funciones de onda permitidas. La ecuación predice los posibles valores (o niveles) de energía que un electrón puede adoptar (representados por E), junto a las denominadas funciones de onda que describen su comportamiento (simbolizadas por la letra griega Ψ , psi). La ecuación dice:

$$\hat{H}\Psi = E\Psi$$

La letra \hat{H} representa la expresión matemática (conocida técnicamente como operador) que simboliza la energía total del átomo. Tras efectuar los cálculos, se obtiene un conjunto de niveles

de energía, a cada uno de los cuales le corresponde al menos una función de onda.⁵⁷

Lo más sorprendente es que esta simple operación matemática predecía exactamente los niveles correctos de energía para el átomo de hidrógeno, reproduciendo el éxito del modelo planetario de Bohr. Pero ¿cómo debemos visualizar los electrones atómicos en el modelo de Schrödinger? Ahí está la dificultad. Schrödinger imaginaba los electrones del átomo como nubes de carga eléctrica, cuya distribución espacial está implícita en la función de onda.

Un importante problema al que Schrödinger tuvo que enfrentarse al escribir su ecuación es que ésta se halla en desacuerdo con la teoría especial de la relatividad. Básicamente, la ecuación es inconsistente con las líneas maestras del principio de relatividad einsteniano, según el cual toda ecuación debía adoptar una forma matemática

⁵⁷ Cuando se trata de determinar los niveles de energía de un electrón atómico, la ecuación de Schrödinger se escribe así:

$$\hat{H}\Psi = E\Psi$$

Donde \hat{H} es el hamiltoniano, la representación matemática de la energía total del sistema (la suma de su energía de movimiento más su energía de posición); ψ (psi) es la función de onda que describe características del sistema tales como su posición en el espacio en cualquier instante y E , uno de los muchos valores posibles para la energía del sistema. Es decir, en el lado izquierdo de la ecuación hay una función matemática que opera sobre ψ , mientras que en el lado derecho tenemos la energía, E , correspondiente a un nivel concreto, multiplicada por la función de onda, ψ , a él asociada.

Nota para los más entusiastas. Esta forma de la ecuación de Schrödinger es en realidad un caso especial, aunque extremadamente importante, y en ella no aparece el tiempo. La forma más general es un poco más compleja:

$$\hat{H}\Psi = \frac{i\hbar}{2\pi} \frac{\partial \Psi}{\partial t}$$

y dice que cuando el hamiltoniano opera sobre la función de onda, el resultado es igual a la tasa de cambio de la función de onda con el tiempo (manteniendo constantes el resto de las variables), multiplicada por la raíz cuadrada de -1 y por la constante de Planck y dividida por 2π .

que le permita incluir medidas realizadas en sistemas que se muevan a altas velocidades, próximas a la de la luz. Lo cierto es que Schrödinger probó inicialmente una aproximación relativista, pero no tuvo éxito. Insertó los resultados de De Broglie en la ecuación relativista que liga energía, masa y momento y, después, particularizó la fórmula para la referencia de todas las teorías cuánticas, el átomo de hidrógeno, con objeto de calcular su espectro de valores energéticos. Fracasó, debido a que su ecuación relativista no incluía el espín del electrón, una propiedad que apenas se comenzaba a entender en aquella época. Por otra parte, Schrödinger encontró que una versión no relativista daba lugar a resultados que concordaban con las observaciones. El problema quedaría superado en 1928, cuando el teórico inglés Paul Dirac propuso una ecuación cuántica para el comportamiento del electrón que era consistente con la relatividad especial. Esta ecuación explicaba de manera natural por qué el electrón tiene espín.⁵⁸ Recordando el episodio, Dirac escribió que Schrödinger debería haber insistido con la ecuación relativista, ya que, en su opinión: «Es más importante que una ecuación sea bella que el hecho de que concuerde con los experimentos».⁵⁹

⁵⁸ Schrödinger nunca publicó sus deliberaciones sobre una ecuación de onda relativista. La ecuación que estuvo considerando sería poco después conocida como ecuación de Klein-Gordon, y fue desempolvada en 1936 por Wolfgang Pauli y Victor Weisskopf en «Über die Quantisierung der skalaren relativistischen Wellengleichung», *Helvetica Physica Acta*, 1934, 7, págs. 709-731. Para un análisis de esta ecuación, véase: A. I. Miller, *Early Quantum Electrodynamics: A Source Book*, Cambridge University Press, 1994.

⁵⁹ Paul Dirac, «The evolution of the physicist's picture of nature», *Scientific American*, 1963, 208, 45-53, pág. 47.

¿Cómo dedujo Schrödinger su ecuación? Las deducciones que Schrödinger presentó en sus artículos no eran, de hecho, deducciones, sino argumentos de plausibilidad: sabía de antemano adonde quería llegar. En realidad, la ecuación de Schrödinger debería ser considerada un axioma, es decir, no deducible: su validez proviene de las soluciones correctas que proporciona a ciertos problemas, tales como el espectro del átomo de hidrógeno, algo que Schrödinger despachaba en pocas páginas —frente a las piruetas matemáticas de Pauli con la mecánica cuántica de Heisenberg.

Schrödinger continuaba demostrando la equivalencia matemática entre la mecánica ondulatoria y la mecánica cuántica, y empleaba el resultado para justificar su desdén hacia esta última: al hablar de teorías atómicas «podía, sin duda alguna, hacerlo en singular».⁶⁰ Para Schrödinger, *sic transit* la mecánica cuántica. Pero ¿qué clase de escenario ofrecía él? Schrödinger mantenía que, frente al equívoco sistema solar en miniatura de Bohr, más valía no ofrecer modelo visual alguno; en este sentido, era preferible la mecánica cuántica, debido precisamente a su «ausencia total de visualización». Pero ello contradecía el punto de vista filosófico de Schrödinger, quien argumentaba que la función de onda de, por ejemplo, el electrón en el átomo de hidrógeno tenía relación con la distribución de la carga eléctrica de ese electrón alrededor del núcleo. La argumentación de Schrödinger, sin embargo, resultó ser

⁶⁰ Nota 2 de la segunda comunicación, pág. 750.

incorrecta, tal como Heisenberg demostró en 1927: en general, las ondas que representan al electrón no permanecen localizadas, es decir, no se mantienen juntas.⁶¹ En cualquier caso, el propio Schrödinger había admitido desde el principio que su representación visual no era aplicable a sistemas que contuvieran más de un electrón. El motivo era que la función de onda que representa a un único electrón podía ser visualizada como una onda en tres dimensiones, ya que dependía de la posición del electrón en un espacio tridimensional, mientras que la función de onda para un sistema con, por ejemplo, dos electrones tenía tres más tres, es decir, seis dimensiones, algo que está fuera de nuestro alcance visualizar.

El estado de la mecánica cuántica en la primera mitad de 1926 podría ser resumido así: No había existido ninguna teoría atómica adecuada hasta junio de 1925, pero en aquel momento había dos y, en apariencia, completamente distintas. Aunque basada en el concepto de partícula, la teoría de Heisenberg renunciaba a toda visualización de la partícula en sí, su aparato matemático resultaba inusual y difícil de aplicar para los físicos y se apoyaba específicamente en discontinuidades. Pero la discontinuidad es anatema en la física de Newton y en la versión del electromagnetismo anterior a los cuantos, en la que todos los

⁶¹ Véase: W. Heisenberg, «Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik», *Zeitschrift für Physik*, 1927, 43, págs. 172-198 (esp. págs. 184-185). Heisenberg señala que Schrödinger (en su segunda comunicación) comete un error matemático elemental al expandir la función de onda del electrón basándose en funciones de onda de un oscilador armónico que tiene la propiedad única de mantener un paquete de ondas localizado. Pero, en general, éste no es el caso.

procesos tienen lugar de forma continua y son visualizados como ondas. En el polo opuesto, la mecánica ondulatoria de Schrödinger contemplaba la *materia* como si se tratara de ondas, ofrecía una representación visual del fenómeno atómico (si bien limitada a un único electrón) y era capaz de explicar las líneas espectrales discretas sin recurrir a los saltos cuánticos. El aparato matemático, más convencional, de la teoría de Schrödinger (basado en ecuaciones diferenciales) creó el marco para un descubrimiento en el ámbito del cálculo, soportado por la demostración de Schrödinger de que ambas teorías eran matemáticamente equivalentes.⁶² La mecánica ondulatoria era del agrado de los físicos que se resistían a incorporar la discontinuidad a su ciencia y preferían una versión de la física atómica basada en una teoría similar a la newtoniana. Aunque las evidencias concluyentes de la dualidad onda-partícula de los electrones no llegarían hasta 1927, experimentos realizados en 1923 confirmaban ya la hipótesis de De Broglie, por lo que muchos físicos la habían empezado a aceptar. Como Einstein escribía a Schrödinger el 26 de abril de 1926: «Estoy convencido de que ha hecho usted un avance decisivo [...], al igual que estoy convencido también de que la vía de Heisenberg [...] es incorrecta». El primer comentario de Heisenberg sobre la mecánica ondulatoria de Schrödinger del que se tiene noticia es una carta del 8 de junio

⁶² Esto, por supuesto, debería haber sido evidente para cualquiera. La razón es que la ecuación de Schrödinger es del tipo Sturm-Liouville, el cual corresponde a un problema de valores propios, y, por lo tanto, es equivalente a un cálculo de valores propios en álgebra de matrices, que es la matemática empleada por Heisenberg en su teoría. Es curioso el hecho de que, hasta la aparición de la ecuación de Schrödinger, a nadie se le había ocurrido aplicar las funciones propias a la mecánica cuántica.

de 1926 a su amigo y colega Pauli, en la que decía enfurecido: «Cuanto más examino los aspectos físicos de la teoría de Schrödinger, más me desagrada ésta. Lo que Schrödinger afirma acerca de la visualizabilidad de su teoría no es, probablemente, del todo correcto. En otras palabras, es mierda».

Durante ese difícil periodo de su vida profesional, Heisenberg fue especialmente franco con Pauli, quien a la sazón se hallaba en la Universidad de Hamburgo. Los intereses de Pauli eran muy amplios e incluso abarcaban temas esotéricos como la numerología y la cábala. Tampoco se privó de explorar el inframundo de la droga y el sexo de Hamburgo. A principios de los años treinta, Pauli se convirtió en seguidor del psicoanalista suizo Carl Jung. Ambos serían coautores de un libro en el que Pauli escribió un memorable análisis jungiano del gran astrónomo Johannes Kepler, un personaje no muy distinto de él en lo que a intereses extracientíficos se refiere.

Además de los ásperos comentarios en sus cartas a Pauli, Heisenberg pronto publicaría también sus opiniones, si bien en tono más mesurado. En un artículo de junio de 1926, escribía que, a pesar de que las interpretaciones físicas de ambas teorías diferían, su equivalencia matemática permitía dejar a un lado esas diferencias; por «conveniencia», utilizaría en sus cálculos la función de onda de Schrödinger, con la advertencia de que no cabía imponer

sobre la teoría cuántica «imágenes intuitivas» del tipo de las de Schrödinger.⁶³

Schrödinger y Heisenberg se encontraron por primera vez en julio de 1926 en Munich, ciudad en la que Arnold Sommerfeld había invitado a Schrödinger a dar dos conferencias sobre su nueva teoría. Sólo había sitio para estar de pie. Tras el segundo discurso, Heisenberg no se pudo contener más e inició un improvisado monólogo en el que atacaba la mecánica ondulatoria de Schrödinger por su aparente incapacidad para explicar el modo en que la radiación interacciona con la materia mediante saltos cuánticos. Ante las voces de protesta del público, el enojado presidente, un eminente físico muniqués, conminó a Heisenberg a que se sentara y guardara silencio. Más tarde, le diría a Heisenberg que su física «y, con ella, todas esas tonterías como los saltos cuánticos, se han acabado». Heisenberg quedó abatido; al parecer no era capaz de convencer a nadie sobre la validez de sus ideas. Pero siguió insistiendo y en agosto del mismo año Schrödinger comenzó a recibir cartas preocupadas de algunos colegas, en las que le preguntaban cómo se podían explicar ciertos efectos cuánticos sin recurrir a las discontinuidades. El propio Schrödinger comenzó a dudar.

La tensión entre la mecánica cuántica y la mecánica ondulatoria aumentó con la publicación, en julio de 1926, de los resultados obtenidos por el mentor de Heisenberg en la Universidad de

⁶³ W. Heisenberg, «Mehrkörperproblem und Resonanz in der Quantenmechanik», *Zeitschrift für Physik*, 1926, 38, págs. 411-426.

Gotinga, Max Born (quien, por cierto, llegaría a ser citado en los libros de música *pop* por ser abuelo de Olivia Newton-John). A sus cuarenta y cinco años, el tímido y reservado Born era el director de una de las tres instituciones en las que Heisenberg estudiaba. (Los otros dos eran Sommerfeld, de la Universidad de Múnich, y Bohr, de la de Copenhague). Heisenberg había descubierto su mecánica cuántica durante una temporada pasada fuera de Múnich, en Gotinga. En el instituto de Born, los físicos estudiaban la naturaleza de los electrones como partículas observando su dispersión al colisionar con el átomo. Los electrones atómicos eran un tipo de problema físico muy diferente. Born estaba interesado en los electrones «libres», es decir, aquellos en los que la fuerza neta que actúa sobre ellos es nula. En aquel momento, ni la mecánica cuántica ni la mecánica ondulatoria se ocupaban de los electrones libres en su movimiento por el espacio.

Procedente del campo matemático, Born había dominado enseguida las sutilezas de las formulaciones de Schrödinger y Heisenberg, además de su contenido físico. Por ello, todos escuchaban con respeto cuando Born hablaba de las deficiencias de ambas teorías a la hora de explicar los experimentos de dispersión de electrones. Finalmente, Born decidió que hacían falta «nuevos conceptos» y que la mecánica ondulatoria sería su vehículo, ya que al menos permitía la posibilidad de algún tipo de metáfora visual.

Born hizo la extraordinaria propuesta de que la función de onda de Schrödinger no representaba ni la distribución de la carga del

electrón alrededor del núcleo, ni un grupo de ondas de carga que se mueven por el espacio. En lugar de ello, la función de onda era una magnitud totalmente abstracta, en el sentido de que no admitía visualización alguna. No se obtenía a partir de ella una densidad de electricidad, sino algo que actuaba *como si* fuese una densidad —la densidad de probabilidad de que el electrón se halle en cierta región del espacio—. Esta sensacional premisa convertía la ecuación de Schrödinger en algo radicalmente nuevo, en un concepto nunca antes contemplado. Mientras las ecuaciones del movimiento newtonianas proporcionan la posición espacial de un sistema en cualquier instante, la de Schrödinger produce una función de onda, a partir de la cual se puede calcular una probabilidad. La ecuación de Schrödinger no nos dice el camino seguido por la partícula, sino el modo en que la probabilidad de detectarla varía con el tiempo. El objetivo de Born era, nada más y nada menos, asociar la función de onda de Schrödinger a la presencia de materia.

En el otoño de 1926, Heisenberg había llegado a odiar a Schrödinger no sólo porque su ecuación era empleada cada vez más —los celos profesionales no son por completo ajenos a las mentes creativas—, sino también debido a otro importante motivo que impactaba de lleno en lo más profundo de su programa de investigación. Como recordaría después, «Schrödinger trataba de retrotraernos a un lenguaje en el que teníamos que describir la naturaleza por “métodos intuitivos”. No lo podía admitir. Ésta era la razón por la que me incomodaban tanto los desarrollos de Schrödinger, a pesar

de sus enormes éxitos.⁶⁴ Después de todo, su ecuación era muchísimo más sencilla de utilizar que la matemática implicada en mi mecánica cuántica». Heisenberg describía esos desarrollos como muy perturbadores para su «estado psicológico en aquella época».

En noviembre de 1926, Heisenberg publicó un artículo que llamó poco la atención, pero que, en sus propias palabras, «fue muy importante para mí».⁶⁵ Estaba escrito por un hombre airado y no citaba en ninguna parte la teoría de la dispersión de Born, pero criticaba duramente a Schrödinger. Heisenberg demostraba que una interpretación en términos de probabilidad sólo puede ser entendida si existen los saltos cuánticos, es decir, las discontinuidades. El principal objeto del artículo de Heisenberg era demostrar que las probabilidades implican la existencia de fenómenos discontinuos, lo que a su vez requiere la presencia de partículas, las cuales, al fin y al cabo, no son sino discontinuidades en la estructura de la naturaleza. Con ello, Heisenberg apostaba firmemente a favor del punto de vista corpuscular y, como consecuencia, en contra de la mecánica ondulatoria de Schrödinger.

En posteriores artículos, Heisenberg subrayaba que los fenómenos que tienen lugar en los diminutos volúmenes del mundo subatómico contradicen nuestra intuición. Con ello pretendía decir que, al contrario de lo que afirmaba Schrödinger, es engañoso aplicar al átomo conceptos extraídos de nuestra percepción diaria, tales como

⁶⁴ Entrevista con W. Heisenberg en *Archive for History of Quantum Physics*, 22 de febrero de 1963, pág. 30.

⁶⁵ *Ibid.*, pág. 3. El artículo de Heisenberg se titula «Schwankungerscheinungen und Quantenmechanik», *Zeitschrift für Physik*, 40, págs. 501-506.

«onda» o «partícula». Fenómenos tales como la dualidad onda-partícula de la luz, enunciada por primera vez por Einstein en 1909, o la dualidad onda-partícula del electrón, propuesta por De Broglie en 1923, son antiintuitivos e inimaginables. ¿Cómo una cosa puede ser continua y discontinua a la vez? Por esta razón, los físicos tardaron en aceptar los cuantos de luz de Einstein. Su principal argumento, establecido por Planck en 1910, era que cuando la luz atravesaba un material dotado de franjas alternativamente opacas y transparentes (conocido como retícula de difracción), se comportaba como las ondas de agua, produciendo un patrón de luces y sombras que variaban suave y continuamente, algo que no era posible explicar si se asumía que la luz estaba formada por partículas. Esta objeción fundamental sólo quedaría resuelta en 1927, cuando Born propuso su interpretación de la función de onda, en la que explicaba esos patrones de difracción en términos de millones de diminutos impactos de partículas individuales de luz. Para muchos físicos, no obstante, la mezcla de elementos corpusculares y ondulatorios en la explicación seguía siendo problemática.

Al igual que el modelo corpuscular para la luz, el modelo ondulatorio del electrón, propuesto por De Broglie en 1923, fue mal acogido al principio. Los físicos aceptaron finalmente la dualidad onda-partícula del electrón gracias a ciertas evidencias experimentales que surgieron aquel mismo año, pero los resultados concluyentes no llegarían hasta 1927. Las primeras evidencias de la existencia de los cuantos de luz aparecieron también en 1923, pero

la persona que realizó los experimentos, Arthur Compton, no se creyó los resultados. Su principal objeción provenía de la relación entre la energía del cuanto de luz (que, al fin y al cabo, era una partícula y, como tal, debía estar localizada) y su longitud de onda (que no lo estaba). ¿Cómo podían verse vinculadas dos magnitudes tan diferentes? ¿No era como tratar de relacionar un pez con una piedra? La naturaleza ondulatoria del electrón, que en 1927 ya era aceptada sin discusión, no perturbó a los físicos tanto como ese ataque a la sacrosanta representación de la luz como una onda.

Tanto para Heisenberg como para Schrödinger, el objetivo último de la teoría cuántica consistía, tal como admitía el primero, en explorar la «clase de realidad» que subyacía en el mundo atómico. La física se convertía así en una rama de la metafísica, debido, ni más ni menos, a que aspiraba a comprender la naturaleza de la realidad física. Heisenberg asumía el hecho en su clásico artículo de 1927, titulado «Sobre la componente intuitiva de la mecánica y la cinemática en la teoría cuántica» y conocido también como el artículo del «principio de incertidumbre».⁶⁶ El término «intuitivo» en el título subraya que este concepto fundamental tiene que ser redefinido a nivel atómico. Heisenberg deja claro enseguida que un aspecto clave a la hora de abordar la mecánica cuántica es el significado de ciertos términos cuando son extrapolados al dominio del átomo: «El presente artículo establece definiciones exactas de los conceptos: posición, velocidad, energía, etc. (p. ej., de un electrón)».

⁶⁶ Véase nota 61.

Heisenberg insiste en que es la propia *interpretación* de la mecánica cuántica lo que está en juego: «Hasta ahora, la interpretación intuitiva de la mecánica cuántica está llena de contradicciones internas que se hacen evidentes cuando se debate sobre continuidad y discontinuidad o sobre ondas y partículas». Una nueva interpretación intuitiva de la teoría atómica, llena de imágenes visuales, se derivaría, según él, de sus ecuaciones y estaría basada en el «principio de incertidumbre». Esto se traduce en que, a diferencia de la física clásica, en el dominio atómico las incertidumbres en las medidas de posición y momento no se pueden reducir a cero a la vez. El producto de ambas es una cantidad extremadamente pequeña, pero no nula. En otras palabras, cuanto más precisa sea la medida de la posición de una partícula, con menos exactitud se podrá determinar su momento, y viceversa.

Heisenberg logró dar a sus ideas una forma matemática precisa. Utilizó el concepto de incertidumbre o «indeterminación», la «imprecisión en el conocimiento» de las medidas simultáneas de posición y momento (para las situaciones que consideraba, el momento $p = \text{masa} \times \text{velocidad}$). Representando la incertidumbre en posición mediante Δx (delta x) y la incertidumbre en momento como Δp (delta p), la relación de Heisenberg dice que el producto $\Delta x \Delta p$ vale al menos $h/(2\pi)$, donde h es la constante de Planck ($6,6 \times 10^{-34}$ julios \times segundo).⁶⁷ Dejando aparte sus quizá poco usuales unidades, lo cierto es que aunque diminuta, la constante de Planck

⁶⁷ El origen de la constante de Planck se relata en el ensayo «Una revolución sin revolucionarios», del presente libro.

no vale cero. Ésta es la razón por la que, de acuerdo con el principio de incertidumbre, cuanto mayor sea la precisión con la que midamos la posición de una partícula en un instante dado, peor conoceremos su momento en ese mismo instante. El hecho es contrario al sentido común y a la idea intuitiva en la física de Newton de que no hay razón que impida conocer en cualquier momento y con la precisión deseada tanto dónde se encuentra una partícula como a qué velocidad se mueve. Según Newton, la exactitud con la que conocemos la posición de una manzana al caer no debería, en principio, tener nada que ver con la precisión con la que sabemos su velocidad al mismo tiempo.

Tras demostrar que las discontinuidades y una representación de las partículas resultaban esenciales en cualquier nueva teoría atómica y que las imágenes que proponía Schrödinger, tomadas de los fenómenos conocidos, eran insuficientes, Heisenberg se ocupaba de responder a las alusiones hechas por aquél en su tercera comunicación de 1926. Lo hacía en una nota a pie de página, como una ocurrencia de última hora. Recordaba el comentario de Schrödinger sobre la versión matricial de la mecánica cuántica en el sentido de que era una teoría «amedrentadora y hasta repulsiva por antiintuitiva y en exceso abstracta», y continuaba haciendo un elogio de doble filo a su rival por haber formulado una teoría que podía no llegar a ser suficientemente estimada debido a que permitía una «penetración matemática de las leyes de la mecánica cuántica». No obstante, seguía Heisenberg, en su «opinión», una

«intuitividad popular» apartaba a los científicos del «camino recto» a la hora de abordar los problemas físicos.

En esa época estaba claro que Schrödinger no tenía intención de polemizar por escrito, pero en privado persistía en la idea de que era posible una imagen ondulatoria de las partículas elementales, sin probabilidades y sin saltos cuánticos. El 4 de octubre de 1927, Schrödinger llegaba al instituto de Bohr en Copenhague para dar una conferencia sobre su teoría. Heisenberg recordaba lo sucedido de este modo:

«Las discusiones entre Bohr y Schrödinger comenzaron en la propia estación de ferrocarril y continuaron todos los días desde primera hora de la mañana hasta avanzada la noche. Schrödinger se alojaba en casa de Bohr, así que no había nada que interrumpiera sus conversaciones. Y aunque Bohr normalmente era muy amable y considerado en su trato con la gente, me sorprendió esta vez que se comportara como un fanático intransigente, alguien incapaz de hacer la menor concesión o de aceptar la remota posibilidad de estar equivocado. Es difícil describir cuán apasionados fueron los diálogos y hasta qué punto arraigaban en lo más profundo de las convicciones de ambos, un hecho que se traslucía en cada una de sus palabras».⁶⁸

Al analizar diversos modos en los que el electrón podía efectuar transiciones atómicas, Schrödinger concluyó: «La idea de los saltos cuánticos es pura fantasía». La respuesta de Bohr fue simplemente:

⁶⁸ W. Heisenberg, *Physics and Beyond: Encounters and Conversations*, Nueva York, Harper & Row, 1971, pág. 73. Traducido por A. J. Pomerans.

«Tiene usted razón. Pero esto no demuestra que los saltos cuánticos no existan. Sólo prueba que no podemos visualizarlos».⁶⁹ Uno de los argumentos finales de Schrödinger fue que «si todos esos condenados saltos cuánticos estuvieran realmente ahí, lamentaría haber tenido alguna vez que ver con la teoría cuántica».⁷⁰ A estas alturas, la tensión había hecho que Schrödinger cayera enfermo y tuviera que guardar reposo. La esposa de Bohr se ocupó amablemente de él; sin embargo, Bohr se sentó junto al lecho e, implacable, continuó: «En cualquier caso, debe admitir que...».⁷¹ Schrödinger se negó a dar su brazo a torcer. Continuaba creyendo que los procesos atómicos podían ser visualizados mediante imágenes tradicionales, convenientemente retocadas. Pero Bohr era de otra opinión y se había interesado cada vez más por el principio de incertidumbre de Heisenberg, el cual indicaba que las ecuaciones de la mecánica cuántica abrirían camino a unos modelos visuales completamente nuevos. Tras cerrar un círculo, la física había regresado al punto de vista platónico, dos mil años atrás, cuando las matemáticas eran la guía hacia lo que constituye la realidad física.

La ecuación de Schrödinger resultó tener una enorme gama de aplicaciones. En el campo de la química, dio pie al nacimiento de una nueva rama, la química cuántica, que estudia los enlaces entre átomos y cuestiones tan complejas como el enlace molecular y la

⁶⁹ *Ibid.*, pág. 74.

⁷⁰ *Ibid.*, pág. 75.

⁷¹ *Ibid.*, pág. 76.

reactividad química. El primer éxito de la ecuación en esta área fue la descripción del enlace de la molécula de hidrógeno realizada en 1927 por Walther Heitler y Fritz London. Este tipo de problemas eran, por supuesto, imposibles de abordar siquiera con la vieja teoría atómica de Bohr. La solución se basaba en otro de los espectaculares hallazgos de Heisenberg, quien en 1926 había deducido el espectro del átomo de helio, una cuestión insoluble en el marco de la teoría de Bohr. El aspecto más notable del descubrimiento era que, según la teoría cuántica, las partículas se podían atraer unas a otras mediante un intercambio extremadamente rápido de sus posiciones. Este fenómeno era la base de la teoría de Heitler y London, y también supondría la clave de la primera teoría sobre la fuerza que mantiene unido el núcleo, que el propio Heisenberg formuló en 1932.

La ecuación de Schrödinger puede ser utilizada también para estudiar el modo en que los elementos químicos reaccionan a nivel molecular; algo que es extremadamente difícil —cuando no imposible— observar experimentalmente con detalle. La función de onda de cada molécula es muy compleja: ha de tener en cuenta tanto las posiciones relativas como las interacciones de todas las partículas que la constituyen. Computar a mano esas funciones a partir de la ecuación de Schrödinger es casi imposible; por este motivo, el cálculo de esas funciones de onda y la comprensión de los procesos químicos a nivel molecular han ido parejos al desarrollo de los ordenadores —imprescindibles en este terreno— desde finales de

los años setenta. Todo lo cual ha obtenido como fruto un avance en casi todas las ramas de la química, desde la producción de nuevos fármacos hasta el estudio de la atmósfera terrestre.

El ámbito de la ecuación de Schrödinger no se limita a los dominios atómico y subatómico. Sirve también para explicar algunos extraordinarios efectos que observamos a mayor escala, por ejemplo, la superconductividad y la superfluidez. Los superconductores son materiales especiales cuya resistencia eléctrica se reduce bruscamente a cero cuando la temperatura desciende por debajo de un valor crítico que, por lo general, es inferior a $-250\text{ }^{\circ}\text{C}$, una temperatura extremadamente baja incluso respecto al más frío de los entornos naturales. Dichos materiales poseen un gran número de cualidades extraordinarias, entre las que cabe destacar la generación de campos magnéticos como consecuencia de la superconducción. El fenómeno de la superfluidez es igualmente enigmático. Sólo tiene lugar en el helio líquido a temperaturas extremadamente bajas, momento en el que se producen cosas muy extrañas: fluye prácticamente sin viscosidad y puede incluso ascender por las paredes del recipiente que lo contiene, desbordándolo. Lo más notable es que tanto la superconductividad como la superfluidez pueden ser analizadas teóricamente aplicando la ecuación de Schrödinger a las moléculas y átomos que constituyen el material.

Además de desempeñar un papel importante en la física y en la química, la ecuación de Schrödinger se ha convertido en objeto de

reflexión filosófica. Consideremos el denominado problema de la medida. Mientras en la física clásica podemos ignorar la interacción entre el aparato de medida y el sistema bajo observación, en la teoría cuántica esto no es así. Pensemos, por ejemplo, en el experimento siguiente. Deseamos medir la posición de una canica que cae, lo cual puede realizarse, entre otros métodos, mediante una fotografía. El proceso implica iluminar la canica y recoger la luz que ésta refleja en una placa fotográfica. El hecho de que la canica sea bombardeada con cuantos de luz no produce efecto apreciable alguno sobre el resultado. En la práctica, la posición, la velocidad e, incluso, el momento de la canica pueden ser determinados simultáneamente con la precisión que se desee.

Pero ¿qué ocurre si la canica es un electrón? Según la mecánica ondulatoria, el electrón que cae puede estar en cualquier sitio, ya que su función de onda se halla esparcida por todo el espacio; por el contrario, la canica se encuentra localizada desde el principio.⁷² Está claro que la pregunta «¿Cuál es la posición del electrón?» no tiene *verdadero* sentido hasta que realmente se realice una medida, una fotografía en este caso. Fotografiar el electrón significa iluminarlo con al menos un cuanto de luz, el cual forma parte del sistema de medida. La interacción de ese único cuanto de luz con el electrón ubica a éste en un instante determinado. Este hecho es conocido como «colapso de la función de onda», porque la interacción entre el sistema de medida (el cuanto de luz) y el

⁷² Esto se debe, más o menos, a que cuanto más masivo es un objeto, menos importantes son sus propiedades ondulatorias. Como consecuencia —y por fortuna—, estamos localizados.

sistema observado (el electrón) reduce la función de onda, previamente extendida, de este último a una región concreta del espacio. En otras palabras, de todas las posiciones posibles que el electrón puede adoptar cuando la función de onda se halla esparcida en el espacio, el proceso de medida selecciona una sola. El estado del electrón, por lo tanto, cambia irreversiblemente de encontrarse en potencia en cualquier parte a hallarse con certeza en un lugar concreto. El principio de incertidumbre nos dice que el coste es una enorme indeterminación en el momento del electrón. Uno de los enigmas aún no resueltos en la teoría cuántica es lo que le sucede a la función de onda de un electrón (o de cualquier otra partícula) durante una medida. Antes de realizar ésta, el electrón es una combinación de varios estados cuánticos, pero, según la tradición cuántica, el mero acto de medir sitúa el electrón en un estado particular. ¿Qué mecanismo subyace detrás de esto? Frente a dicha cuestión trascendental, tanto la de Schrödinger como otras ecuaciones fundamentales guardan un respetuoso silencio.⁷³

Hay una curiosa fotografía de los galardonados con el Premio Nobel en 1933, tomada en la estación de ferrocarril de Estocolmo. Dirac aparece a la derecha de Heisenberg y Schrödinger, a su izquierda. Dirac y Heisenberg visten indumentaria formal y llevan abrigo. En la mayoría de las fotografías, Heisenberg aparece sonriente o en una pose digna y seria, pero aquí semeja apartarse de Schrödinger con

⁷³ Véase: E. Schrödinger, «Die gegenwärtige Situation in der Quantenmechanik», *Die Naturwissenschaften*, 1935, 23, págs. 807-812, 823-828, 844-849. Reimpreso en inglés en *Quantum Theory and Measurement*, en: W. Zurek y J. Wheeler, eds., Princeton University Press, 1983, págs. 152-167.

aire casi de disgusto. Schrödinger es el único de los tres que sonrío y parece estar en su salsa. Viste a la última moda: pantalones bombachos y calcetines largos, una informal cazadora de ancho cuello de piel y su inseparable pajarita. Otra memorable fotografía en la que están presentes ambos adversarios también nos habla de sus estilos tan radicalmente distintos. Corresponde a un encuentro anual de físicos, la Conferencia Solvay celebrada en Bruselas en 1933. Como era habitual en estas fotos, los asistentes de mayor edad aparecían sentados y los más jóvenes, de pie. Por orden, estos últimos van ocupando asientos. En este retrato, Schrödinger está sentado y Heisenberg se queda de pie casi exactamente detrás de él. Aunque muchos físicos consideran que la teoría cuántica es un libro cerrado, aún existen algunos puntos fundamentales que permanecen sin resolver y la mayor parte de ellos provienen de la ecuación de Schrödinger. Éste escribía a Einstein el 23 de marzo de 1936 acerca de su reciente encuentro con Bohr en Londres: «Me pareció bien que trataran de atraerme al punto de vista de Bohr-Heisenberg de una forma tan amigable. [...] Le dije a Bohr que me encantaría que pudieran convencerme y que me quedaría mucho más tranquilo si lo lograban».⁷⁴ Bohr nunca lo consiguió,⁷⁵ y en lugar de ello, optó por dejar de lado a Schrödinger.

⁷⁴ Cita extraída de *Schrödinger: Life and Thought*, de W. Moore (Cambridge University Press, 1989, pág. 314).

⁷⁵ «Are there quantum jumps», incluido en *Speakable and Unspeakable in Quantum Mechanics*, de J. S. Bell (Cambridge University Press, 1987, págs. 201-212), es un magistral estudio sobre los últimos trabajos basados en la mecánica ondulatoria no relativista de Schrödinger.

La frontera en el conflicto entre ondas y partículas quedó rápida y claramente establecida. Las cosas parecieron ir bien para la causa de Schrödinger durante algún tiempo hasta que, en el invierno de 1926, Bohr convocó a Heisenberg a Copenhague para debatir las bases de la física cuántica. Sus deliberaciones ocuparon casi todo el año siguiente. Durante ese tiempo, ambos elaboraron lo que hoy se conoce como *interpretación de Copenhague*, un enfoque que hace énfasis en las probabilidades, las discontinuidades y el colapso de la función de onda, conceptos que eran anatema para Schrödinger. Pero éste no presentó batalla. Schrödinger no se enfrentó a ellos ni por escrito ni en la famosa Conferencia Solvay de 1927, y dejó que fuera Einstein, nada menos, quien mantuviera la bandera en alto. Pero Einstein tampoco se entendió con Bohr y compañía, a pesar de algunas contrapropuestas ingeniosas. Esta *guerra* se prolongó durante un año. Mientras que Schrödinger nunca acabó realizando otro gran descubrimiento antes o después del de la ecuación que lleva su nombre, Heisenberg ya había cosechado varios éxitos notables antes de junio de 1925 y realizaría otros importantes trabajos hasta mediados de los años treinta. Sería siempre alguien a tener en cuenta; en el panteón de los físicos ilustres del siglo XX, Heisenberg sólo tendría por encima a Einstein.

Ironías del destino, aunque Heisenberg venció en la batalla y quedó convencido de haber ganado la guerra, la ecuación de Schrödinger se utiliza con mucha mayor frecuencia que la versión heisenbergiana de la física atómica. Y esto es así a pesar de que la

ecuación de Schrödinger es incompatible con la relatividad, algo que no tiene repercusión en las aplicaciones prácticas (lo cual es chocante, ya que muchas de esas aplicaciones involucran cuantos que viajan a ninguna parte a velocidades próximas a la de la luz). Por su parte, el formalismo matricial de Heisenberg encontró su lugar en áreas profundamente teóricas, tales como la teoría cuántica de campos, dentro de la física de las partículas elementales.

Lo que siempre me ha parecido más fascinante de la disputa entre Heisenberg y Schrödinger es que, en el fondo, se trataba de una elección estética. En principio, ambas versiones de la física atómica concordaban con todos los datos experimentales disponibles sobre el átomo de hidrógeno, y eran fundamentalmente equivalentes en el sentido de que hacían los mismos razonamientos, por ejemplo, para el átomo de helio. Ambos físicos defendían apasionadamente su punto de vista sobre la naturaleza. En este sentido, el gran mérito de Bohr consistió en introducir una tercera estética. Ni Heisenberg ni Schrödinger se planteaban la idea de la dualidad onda-partícula en la luz y en la materia. Bohr aportó la idea de que ondas y partículas debían ser consideradas a la vez, mediante una interpretación adecuada de la ecuación de onda de Schrödinger, algo que ya había realizado Born.

El que existan dos enfoques de la física del átomo no debería sorprendernos, ya que en nuestro mundo de percepciones las cosas siempre se dan por parejas: partículas y ondas, *ying* y *yang*, blanco

y negro, sí y no, amor y odio, luz y oscuridad; no existen los *quizás* intrínsecos, como en el dominio atómico. Sólo mediante la abstracción, a través de la concepción y no de la percepción, podemos ascender a un plano más elevado y apreciar el poder de la ambigüedad. Se trata de algo, por lo general, incómodo para el día a día, en el que sistemáticamente intentamos resolver situaciones ambiguas mediante la decisión, regresando de nuevo al familiar terreno de *esto* o *lo otro*. Como Einstein y Picasso demostraron en la primera década del siglo XX, la ambigüedad es la clave para descubrir representaciones de la naturaleza que están más allá de la superficial apariencia. La vista puede engañar, tal como Einstein descubrió en la física y Picasso en el arte. En la teoría de la relatividad de 1905, el espacio y el tiempo son relativos y cada observador los interpreta de distinta manera. Por ejemplo, dos sucesos que ocurren al mismo tiempo para un observador no serán simultáneos para otro que se halle en movimiento relativo al primero. En el gran óleo de 1907 *Les demoiselles d'Avignon*, que constituiría el germen del cubismo del propio Picasso y del de Georges Braque, el pintor español descubre un modo de representar las figuras en el que en un mismo lienzo se muestran varias perspectivas a la vez.⁷⁶ A su manera, tanto Schrödinger como Heisenberg trasladaron la aventura de la abstracción al dominio atómico.

⁷⁶ Véase: A. I. Miller, *Einstein, Picasso: Space, Time and the Beauty that Causes Havoc*, Nueva York, Basic Books, 2001.

El crítico literario William Empson ha argumentado elocuentemente que las revelaciones de la teoría cuántica podrían también iluminar la literatura.⁷⁷ Antes de pasarse a la literatura en 1928, y siendo estudiante en Cambridge, Empson había estudiado matemáticas y tenía una buena formación en física. Desarrolló nuevas interpretaciones de la obra de Shakespeare, tratando de «unir la noción de probabilidad al objeto natural en vez de a la infalibilidad de la mente humana».⁷⁸ Empson abogaba por renovar el estudio de la literatura desde la perspectiva de una realidad alterada por la teoría cuántica. Así pues, Shakespeare no debía ser analizado en el modo «esto o aquello», sino centrando la atención en las ambigüedades, es decir, en el modo «todas las opciones», lo que podría sacar a la luz significados del texto ocultos hasta la fecha. Es posible que un texto tenga simultáneamente dos significados contradictorios, como en la dualidad onda-partícula. Uno de los ejemplos de Empson es la interpretación de un personaje tan complejo como Falstaff, alguien que debemos contemplar como la suma de dos aparentes opuestos: «la suprema expresión de la parodia y la idealización cómica de la libertad, alguien a la vez infame y trágicamente maltratado».⁷⁹ En opinión de Empson, el lector debería «tener en mente todo el espectro de significados que [Shakespeare] podría haber querido expresar y ponderarlos [...]

⁷⁷ págs. 311-316.

⁷⁸ *ibid.*

⁷⁹ Cita extraída de *ibid.*, pág. 316.

según sus probabilidades»,⁸⁰ al igual que el físico representa el estado de un átomo mediante funciones de onda.

Los conceptos de la teoría cuántica, con sus profundas abstracciones, impregnan en la actualidad cualquier aspecto de nuestras vidas. Han exigido que nos replanteemos una amplia gama de materias y transformemos nuestro conocimiento intuitivo de la naturaleza. Casi todos los físicos emplean la teoría cuántica a diario, aunque muy pocos se hayan parado a pensar sobre las sutilezas de su interpretación. Como una gran obra literaria, la teoría cuántica está abierta a multitud de interpretaciones. La mayor parte de los físicos no son conscientes de ello y asumen que lo que leen en los libros de teoría cuántica es una especie de catecismo. La actitud está tan arraigada que los autores ni se molestan ya en advertir que se basan en la interpretación de Copenhague, establecida en 1926-1927 por Bohr y Heisenberg. Como profesor de historia y filosofía de la física he observado que los alumnos más reflexivos son a quienes más sorprende e incómoda la teoría, pues esperan certezas en la exposición del texto y se topan con ambigüedades en la interpretación. John Bell, el físico que más ha profundizado en los cimientos de la teoría cuántica después de Bohr, Einstein y Heisenberg, decía que la teoría funcionaba bien «a efectos prácticos»,⁸¹ pero recordaba que, no obstante, aún no se entendía en su totalidad la ecuación de

⁸⁰ Cita extraída de *íbid.*, pág. 314.

⁸¹ Véase: J. S. Bell, «Against Measurement», *Sixty-Two Years of Uncertainty: Historical, Philosophical, and Physical Inquiries into the Foundations of Quantum Mechanics*, En: A. I. Miller, ed., Nueva York, Plenum Press, 1990, págs. 17-31.

Schrödinger. Como escribía el intuitivo físico Richard Feynman con su mordaz estilo: «Creo poder afirmar con seguridad que nadie comprende la mecánica cuántica».⁸²

⁸² R. P. Feynman, *The Character of Physical Law*, Londres, Penguin Books, 1992, pág. 129. [Trad, esp.: *El carácter de la física*, Barcelona, Tusquets Editores, col. Metatemas 65, 2000].

Parte 5

Un poco de magia

La ecuación de Dirac

Frank Wilczek

Uno no puede evitar la sensación de que esas fórmulas matemáticas tienen una existencia independiente e inteligencia propia, de que son más sabias que nosotros, más sabias aún que sus descubridores, de que podemos obtener de ellas más de lo que en ellas se puso.

Heinrich Hertz, sobre las ecuaciones del electromagnetismo de Maxwell

Una gran parte de mi trabajo consiste simplemente en jugar con las ecuaciones y ver lo que resulta.

Paul Dirac

Producía exactamente las propiedades que se requerían para un electrón. Fue un auténtico regalo para mí, algo totalmente inesperado.

Paul Dirac, sobre la ecuación de Dirac

Las ecuaciones parecen cosa de magia. Al igual que las escobas del Aprendiz de Brujo, ellas solas cobran vida y producen consecuencias que su creador no esperaba, que no puede controlar

y que, a veces, incluso detesta. Cuando Einstein descubrió $E = mc^2$, como culminación del proceso de consolidación de los fundamentos de la física clásica por parte de la teoría especial de la relatividad, ni las armas de destrucción masiva ni los generadores de energía inextinguible pasaron remotamente por su cabeza.

De todas las ecuaciones de la física, quizá la más *mágica* de todas sea la ecuación de Dirac. Es la que más libremente fue formulada, la menos condicionada por los experimentos y la que tiene las consecuencias más raras y sorprendentes.

A principios de 1928 (la fecha de recepción del artículo original es el 2 de enero), Paul Adrien Maurice Dirac (1902-1984), un ingeniero eléctrico de veinticinco años recién convertido en físico teórico, creaba una notable ecuación que la posteridad bautizaría con su nombre. El propósito de Dirac era muy concreto y en absoluto original. Deseaba producir una ecuación que describiera el comportamiento de los electrones con mayor precisión que las anteriores, unas ecuaciones que, o bien incorporaban la relatividad especial, o bien la mecánica cuántica, pero nunca ambas teorías. En aquel momento había físicos importantes y más experimentados que él que trabajaban en el mismo problema.

A diferencia de esos otros físicos —y de los grandes clásicos, Newton y Maxwell—, Dirac no partió de un minucioso estudio de hechos experimentales; en lugar de ello, basó su investigación en algunas evidencias básicas y en ciertos imperativos teóricos intuitivos (algunos de los cuales sabemos hoy que eran erróneos). Dirac se

propuso plasmar todos esos principios en un esquema compacto y matemáticamente consistente. «Jugando simplemente con las ecuaciones», como él decía, dio con una solución sorprendentemente simple y elegante. Se trata, por supuesto, de la ecuación de Dirac.

Algunas consecuencias de esta ecuación pudieron ser contrastadas con observaciones experimentales previas. La concordancia era excelente y se explicaban así resultados hasta entonces misteriosos. En concreto y tal como se describirá más adelante, la ecuación de Dirac predecía correctamente que los electrones se hallan en perpetua rotación y que actúan como diminutos imanes, y establecía la velocidad de ese giro y la intensidad del campo magnético que generan. Pero otras consecuencias parecían estar en total contradicción con las evidencias. Por ejemplo, la ecuación da lugar a soluciones que semejan describir un comportamiento según el cual los átomos ordinarios podrían evaporarse, en un destello de luz y en una fracción de segundo, de manera espontánea.

Durante años, Dirac y otros físicos se enfrentaron a una extraordinaria paradoja. ¿Cómo era posible que una ecuación fuese *obviamente correcta*, ya que concordaba de forma muy precisa con numerosos resultados experimentales, y, a la vez, propensa a generar soluciones manifiestamente erróneas? La ecuación de Dirac se convirtió en el eje en torno al cual giraba toda la física fundamental. Dando por válida su formulación matemática, los físicos tuvieron que reexaminar el significado de los símbolos que

contenía. Fue en el curso de este confuso e intelectualmente penoso examen —durante el cual Werner Heisenberg escribía a su amigo Wolfgang Pauli: «El capítulo más triste de la física moderna lo constituye la teoría de Dirac» y «Con el fin de no enojarme con Dirac, he decidido hacer un nuevo esfuerzo por cambiar...»— donde tuvo su verdadero origen la física moderna.

Un espectacular resultado fue la predicción de la existencia de la *antimateria*, o más exactamente, que tenía que haber una nueva partícula con la misma masa que el electrón, pero con la carga eléctrica contraria, y que al encontrarse ambas se aniquilarían, dando por resultado pura energía. Un análisis concienzudo de los rayos cósmicos, efectuado por Carl Anderson en 1932, confirmaría poco después la veracidad de la hipótesis.

Pero el resultado más profundo y trascendental fue la completa reelaboración del modo en que describimos la materia. En esta nueva física, las partículas son entes efímeros. Se crean y se destruyen libremente. De hecho, su existencia fugaz y sus intercambios son la fuente de toda interacción. Los verdaderos objetos fundamentales son los campos cuánticos, unos intangibles entes universales y transformativos. Éstos son los conceptos que subyacen bajo nuestra moderna e increíblemente precisa teoría de la materia (el inadecuadamente llamado Modelo Estándar). Y la ecuación de Dirac, reinterpretada a fondo y generalizada, se ha convertido en piedra angular de nuestra comprensión de la naturaleza.

Trataremos de ver qué hay detrás de su magia y, en última instancia, de descubrir el truco. Para empezar, he aquí la ecuación:

$$\left[\gamma^\mu \left(i \frac{\partial}{\partial x^\mu} - e A_\mu(x) \right) + m \right] \psi(x) = 0$$

Los jeroglíficos que contiene están descifrados en el Apéndice del presente ensayo. Por ahora, nos centraremos en la ecuación en sí y trataremos de apreciar sus múltiples facetas.

El problema de Dirac y la unidad de la naturaleza

El antecedente inmediato del descubrimiento de Dirac y el razonamiento que condujo hasta él fue la necesidad de reconciliar dos avanzadas y exitosas teorías físicas que presentaban pequeñas desavenencias. En 1928, la teoría especial de la relatividad de Einstein tenía ya más de veinte años y se hallaba plenamente digerida y totalmente establecida. (La teoría general, que describe la gravitación, no forma parte de nuestra historia; la gravedad es despreciable a escala atómica). Por su parte, la mecánica cuántica de Heisenberg y Schrödinger era una teoría joven, pero proporcionaba un nuevo y brillante enfoque acerca de la estructura del átomo y había explicado un buen número de fenómenos hasta entonces misteriosos. En particular, plasmaba aspectos esenciales de la dinámica de los electrones atómicos. El problema era que las ecuaciones desarrolladas por Schrödinger y Heisenberg no se basaban en la mecánica relativista de Einstein, sino en la vieja mecánica newtoniana. La mecánica de Newton podía constituir una

excelente aproximación en sistemas en los que todas las velocidades son mucho más pequeñas que la de la luz, lo que abarca muchos casos de interés en la química y en la física atómica. Pero los datos experimentales sobre espectros atómicos, que pertenecían al ámbito de la nueva teoría cuántica, eran tan precisos que se podían observar desviaciones respecto a las predicciones de Heisenberg-Schrödinger. Así pues, había una razón *práctica* que aconsejaba buscar una ecuación para el electrón más precisa y basada en la mecánica relativista. El joven Dirac y otros importantes físicos se pusieron manos a la obra.

En realidad, estaban en juego dicotomías mucho más antiguas y fundamentales: la luz frente a la materia o lo continuo frente a lo discreto. Estas dicotomías suponían barreras casi infranqueables a la hora de establecer una descripción unificada de la naturaleza. De las dos teorías que Dirac y sus contemporáneos trataban de conciliar, la relatividad representaba lo continuo y la luz y la teoría cuántica, lo discreto y la materia. Tras la revolución llevada a cabo por Dirac, ambas quedaron reconciliadas bajo una amalgama conceptual de difícil comprensión denominada campo cuántico.

Las dicotomías luz/materia y continuo/discreto debieron ser percibidas ya por los primeros humanos y fueron claramente formuladas y debatidas —sin llegar a conclusión alguna— por los antiguos griegos. Aristóteles consideraba el fuego y la tierra como elementos primarios y argumentaba contra los atomistas, a favor de un *plenum* fundamental («la naturaleza abomina del vacío»).

Esas dicotomías no quedaron resueltas por los grandes logros de la física clásica; por el contrario, se vieron, si cabe, reforzadas por ellos.

La mecánica newtoniana describía el movimiento de cuerpos rígidos a través de un espacio vacío. Aunque el propio Newton especuló en varias ocasiones desde las dos *orillas* de ambas dicotomías, sus seguidores insistían en el carácter «sólido, macizo e impenetrable» de sus átomos, los «ladrillos» con los que estaba edificada toda la naturaleza. Incluso la luz estaba modelada en términos de partículas.

A principios del siglo XIX, una nueva teoría cosechó notables éxitos: la luz está compuesta por ondas. Los físicos asumieron que tenía que existir un éter continuo y omnipresente que impregnara todo el espacio y sirviera de soporte a esas ondas. Los descubrimientos de Faraday y Maxwell, que convertían la luz en un campo electromagnético más —un ente continuo que impregnaba todo el espacio—, matizaron y reforzaron la idea.

No obstante, tanto el propio Maxwell como Ludwig Boltzmann demostraron que las propiedades observadas en los gases, incluyendo muchos detalles sorprendentes, podían ser explicadas suponiendo que aquéllos estaban compuestos por infinidad de partículas diminutas, discretas y perfectamente individualizadas, que se mueven a través de un espacio vacío. Por otra parte, J. J. Thomson en el terreno experimental y Hendrik Lorentz en el teórico establecieron la existencia de los electrones como componentes

fundamentales de la materia. Los electrones aparentaban ser partículas indestructibles, del tipo de las que Newton había soñado. De este modo, a comienzos del siglo XX los físicos disponían de dos conjuntos de teorías completamente distintas que convivían en una armonía no siempre fácil. La electrodinámica de Maxwell era una teoría de lo continuo, de los campos eléctricos y magnéticos y de la luz, que no hablaba para nada de la masa. Por su parte, la mecánica de Newton era una teoría de partículas discretas, cuyas *únicas* propiedades obligadas eran la masa y la carga eléctrica.⁸³

Inicialmente, la teoría cuántica floreció a lo largo de dos ramas, siguiendo la huella de nuestras dicotomías. La primera de ellas, relacionada con la luz, partía de los trabajos de Planck sobre radiación y alcanzaba su cénit en la teoría einsteniana de los fotones. Su resultado fundamental era que la luz estaba compuesta por unidades mínimas indivisibles, denominadas fotones, cuya energía y momento son proporcionales a la frecuencia de dicha luz. Esta idea, obviamente, suponía un enfoque corpuscular de la luz.

La segunda rama se ocupaba de los electrones, y arrancaba de la teoría del átomo de Bohr para alcanzar su clímax en la ecuación de onda de Schrödinger. Establecía que las configuraciones estables del electrón alrededor del núcleo atómico estaban asociadas a patrones regulares de vibraciones ondulatorias. Lo cual generaba un punto de vista ondulatorio sobre la materia.

⁸³ Es decir, para predecir el movimiento de una partícula basta con conocer su carga y su masa; ni más, ni menos. El valor de la carga puede ser cero; en este caso, la partícula sólo estará sometida a interacciones gravitatorias.

Así pues, las viejas dicotomías se habían suavizado: la luz era en parte corpuscular y los electrones se comportaban en parte como ondas. Pero subsistían algunos fuertes contrastes. En concreto, había dos circunstancias que parecían diferenciar claramente la luz de la materia. En primer lugar, si la luz estaba compuesta por partículas, éstas tenían que ser muy especiales y con alguna clase de estructura interna, ya que la luz puede estar polarizada. Para que esta propiedad se manifieste, las partículas de luz deben tener la propiedad equivalente. No es posible realizar una descripción completa de un haz de luz diciendo simplemente que está compuesto por tantos fotones de tales energías; esos parámetros nos indicarían sólo cuán brillante es y cuántos colores contiene, pero no cómo está polarizado. Para obtener una descripción completa, deberíamos incluir información sobre la polarización del haz, lo cual significa que los fotones deberían transportar una especie de flecha que les permitiera conservar la huella de ese parámetro. La idea parece apartarnos del concepto ideal de partícula elemental. Si existe esa flecha, ¿en qué consiste y por qué no puede ser separada de la partícula? En segundo lugar, y más importante aún, los fotones son efímeros. La luz puede ser radiada, como cuando encendemos una linterna, o absorbida, como cuando cubrimos ésta con la mano. De este modo, las partículas de luz pueden ser creadas o destruidas. Esta propiedad básica y familiar de la luz y los fotones nos aparta de la noción tradicional de partícula elemental. La estabilidad de la materia parecía requerir

partículas indestructibles, dotadas de propiedades muy distintas de las de los fugaces fotones.

Observaremos ahora cómo se desvanecieron esas últimas diferencias y cómo fue revelada la unidad de la naturaleza.

Un primer efecto: el espín

Dirac trataba de conciliar la mecánica cuántica del electrón con la relatividad especial. Pensaba —equivocadamente, como ahora sabemos— que la teoría cuántica requería ecuaciones de un tipo particularmente simple, las que los matemáticos denominan «de primer orden». Nunca se cuestionó la idea ni qué implicaba el que las ecuaciones fuesen de esta clase; el hecho es que Dirac buscaba una ecuación que, en un sentido muy preciso, fuera del tipo más simple posible. El problema es que no era fácil encontrar una ecuación simple en ese sentido y, a la vez, consistente con los requisitos de la relatividad especial. Para obtenerla, Dirac tuvo que ampliar el ámbito de su estudio y, finalmente, encontró que no bastaba una única ecuación de primer orden; necesitaba un sistema de cuatro ecuaciones fuertemente interrelacionadas. Este sistema constituye lo que hoy conocemos como *la* ecuación de Dirac.

Dos ecuaciones resultaban aceptables, pero cuatro, al menos inicialmente, representaban un grave problema.

Examinemos primero las *buenas* noticias.

Aunque la teoría de Bohr proporcionaba una versión aproximada de los espectros atómicos, quedaban muchos detalles sin explicar.

Algunas de las discrepancias se referían al número de electrones que podían ocupar cada una de las órbitas; otras tenían que ver con la respuesta de los átomos a los campos magnéticos, tal como se constataba en el desplazamiento de sus líneas espectrales.

Wolfgang Pauli había demostrado, mediante un análisis detallado de las evidencias experimentales, que, aun de forma aproximada, el modelo de Bohr sólo era válido para átomos complejos si se imponía una considerable restricción al número de electrones que podían ocupar una órbita dada. Esto originaba el famoso principio de exclusión de Pauli, que hoy enunciamos como «un estado dado sólo puede ser ocupado por un único electrón». La proposición original de Pauli no era tan clara; el número de electrones que podían ocupar un orbital de Bohr resultaba ser dos y no uno solo, pero Pauli hablaba oscuramente de una «duplicidad no descriptible en términos clásicos», y no daba explicación alguna al respecto.

En 1925, dos licenciados holandeses, Samuel Goudsmit y George Uhlenbeck, idearon una posible explicación para la respuesta al campo magnético. Según ellos, si los electrones fueran en realidad diminutos imanes, desaparecerían las discrepancias. El éxito de su modelo requería que todos los electrones tuviesen la misma fuerza magnética, parámetro que podían calcular. A continuación, proponían un mecanismo para el magnetismo del electrón. Los electrones, por supuesto, son partículas cargadas. Una carga eléctrica en movimiento circular genera un campo magnético. Así pues, si por alguna razón el electrón girara permanentemente

alrededor de su eje, su magnetismo quedaría explicado. Este *espín* intrínseco del electrón presentaba una ventaja adicional. Si la velocidad de giro fuese la mínima permitida por la mecánica cuántica,⁸⁴ la «duplicidad» de Pauli estaría justificada, ya que el espín no tendría la posibilidad de variar su magnitud, sino sólo de apuntar hacia arriba o hacia abajo.

Muchos físicos eminentes se mostraron totalmente escépticos ante la propuesta de Goudsmit y Uhlenbeck. Pauli llegó incluso a tratar de disuadirles de que publicaran su trabajo. Por una parte, el modelo parecía requerir que el electrón girara a una velocidad extremadamente alta, probablemente mayor que la de la luz en su superficie. Por otra, el modelo no explicaba qué es lo que mantenía unido al electrón. Si éste era una distribución de carga eléctrica, toda del mismo signo, debería tender a dispersarse y la rotación, al introducir la fuerza centrífuga, no hacía sino empeorar las cosas. Finalmente, existía una discrepancia cuantitativa entre los requisitos para la intensidad del campo magnético del electrón y la magnitud de su espín. La proporción entre ambas magnitudes viene gobernada por el denominado factor giromagnético, g . La mecánica clásica predecía que $g = 1$, mientras que, para ajustarse a los datos, Goudsmit y Uhlenbeck postulaban que $g = 2$. Sin embargo y a pesar de esas objeciones, su modelo concordaba tercamente con los resultados experimentales.

⁸⁴ En mecánica cuántica sólo están permitidos ciertos valores discretos de espín. Esto tiene mucha relación con la restricción asociada a los orbitales de Bohr.

En ese momento hizo su entrada Dirac. Su sistema de ecuaciones permitía una clase de soluciones para velocidades pequeñas, en la que sólo resultaban relevantes dos de las cuatro funciones que intervenían en aquéllas. Se trataba de duplicidad, pero con una diferencia. En este caso, surgía automáticamente como consecuencia de aplicar unos principios generales y no necesitaba ser introducida. Mejor aún, mediante su ecuación, Dirac podía calcular el magnetismo de los electrones sin recurrir a otras premisas. Y obtuvo que $g = 2$. En el importante artículo de 1928, Dirac era muy conciso. Tras demostrar ese resultado, se limitaba a decir: «El momento magnético es exactamente el asumido por el modelo de electrón giratorio». Y pocas páginas después, tras desarrollar las consecuencias, concluía: «En una primera aproximación, la presente teoría conduce a los mismos niveles de energía que los obtenidos por [C. G.] Darwin, los cuales se hallan en concordancia con los experimentos».

En cualquier caso, los resultados hablaban por ellos mismos y no precisaban más comentarios. Desde entonces, la ecuación de Dirac se convirtió en paso obligado. Cuando surgieron dificultades —que las ha habido de gran calibre—, han sido motivo de un mayor esfuerzo y análisis, pero no de un abandono. Una joya del pensamiento científico como aquélla tenía que ser defendida a toda costa.

Aunque el punto de partida intelectual, como se ha mencionado, fue muy diferente y más abstracto, Dirac comenzaba su artículo

refiriéndose a Goudsmit y Uhlenbeck y al éxito experimental de su modelo. Su aportación personal comenzaba en el segundo párrafo. Lo que en él decía es absolutamente pertinente para los temas que hemos venido examinando.

«La cuestión sigue siendo por qué la naturaleza habría escogido este modelo concreto para el electrón en vez de sentirse satisfecha con una carga puntual. Nos gustaría encontrar alguna deficiencia en los anteriores métodos de aplicar la mecánica cuántica a la carga puntual, de forma que, solventándola, el fenómeno de la duplicidad no requiriera premisa arbitraria alguna».

Así pues, Dirac no estaba proponiendo un nuevo modelo para el electrón. En lugar de ello, definía una nueva propiedad *irreducible* de la materia, inherente a la naturaleza de las cosas —y, en especial, a una implementación consistente de la relatividad y la teoría cuántica— y que se manifestaba incluso en el caso más simple posible de una partícula puntual carente de estructura. El electrón parece ser la encarnación de la forma más simple de materia. Las valiosas propiedades del *espín* de Goudsmit y Uhlenbeck —en particular, su valor constante y sus efectos magnéticos, muy útiles a la hora de describir las observaciones experimentales— han perdurado, asentadas hoy sobre bases mucho más sólidas. Los aspectos arbitrarios o deficientes de su modelo, en cambio, han sido abandonados.

Buscábamos una especie de flecha que formara parte inseparable y necesaria de los entes materiales más elementales, como la polarización en los fotones. Pues bien, hela aquí.

El espín del electrón tiene muchas consecuencias prácticas. Es responsable del fenómeno del ferromagnetismo y de la formación de campos magnéticos en el núcleo de las bobinas eléctricas, las cuales se hallan por doquier en los modernos dispositivos de generación de energía (motores y dinamos). La manipulación activa del espín del electrón permite almacenar y manejar gran cantidad de información en un volumen muy reducido (cintas y discos magnéticos). Incluso el mucho más pequeño e inaccesible espín del núcleo atómico desempeña un importante papel en la tecnología actual. La manipulación de este espín mediante campos magnéticos y ondas de radio y el análisis de la respuesta constituyen la base de las imágenes por resonancia magnética (*Magnetic Resonance Imaging*, MRI), tan útiles en medicina. Esta aplicación, al igual que muchas otras, sería —literalmente— inconcebible sin el exquisito control de la materia al que sólo un conocimiento profundo de ella nos ha permitido acceder.

El espín, en general, y la predicción de Dirac para el momento magnético, en particular, también han desempeñado un trascendente papel en el desarrollo posterior de la física fundamental. En 1940, Polykarp Kusch y sus colaboradores constataron pequeñas desviaciones respecto al $g = 2$ de Dirac. Constituían las primeras evidencias cuantitativas de los efectos de

ciertas partículas virtuales, una importante y característica propiedad de la teoría cuántica de campos. En la década de 1930 se observaron desviaciones muy grandes respecto a $g = 2$ para protones y neutrones. No era más que un primer síntoma de que el protón y el neutrón no son partículas fundamentales en el sentido en que lo es el electrón. Pero nos estamos adelantando a los acontecimientos...

La gran sorpresa: la antimateria

Abordaremos ahora las *malas* noticias.

La ecuación de Dirac consta de cuatro componentes, es decir, contiene cuatro funciones de onda separadas para describir el electrón. Dos de ellas tienen una interpretación atractiva e inmediata y, tal como se ha indicado, describen las dos direcciones posibles del espín. Las otras dos, por el contrario, resultan *a priori* muy problemáticas.

En efecto, las dos ecuaciones adicionales contienen soluciones con energía *negativa* (y cualquier dirección de espín). En física clásica (no cuántica), la existencia de esas soluciones extra podría ser desconcertante, pero no necesariamente catastrófica. En física clásica basta con no utilizar esas soluciones. Al hacerlo, obviamos la cuestión de por qué la *naturaleza* no ha optado por ellas, pero el procedimiento tiene consistencia lógica. En mecánica cuántica no disponemos de esa posibilidad. En la física cuántica, en general, «todo lo que no esté prohibido es obligatorio». En el caso que nos

ocupa podemos ser muy concretos al respecto. Las soluciones de la ecuación de onda del electrón representan todos los posibles comportamientos de éste que pueden darse en las circunstancias adecuadas. Mediante la ecuación de Dirac, y partiendo de un electrón en una de las soluciones de energía positiva, podemos calcular la probabilidad de que emita un fotón y se mueva a una de las soluciones de energía negativa. La energía se tiene que conservar por encima de todo, pero eso aquí no es un problema, simplemente significa que el fotón emitido tiene más energía ¡que el electrón que lo emitió! En cualquier caso, la velocidad resulta ser ridículamente rápida: la transición tendría lugar en una minúscula fracción de segundo. De modo que no cabe ignorar sin más las soluciones de energía negativa. Y como nunca se ha visto que el electrón haga algo tan peculiar como radiar más energía que la que llevaba, la ecuación de Dirac tiene un problema, del cual era plenamente consciente. En su artículo original se limitaba a decir:

«Para esta segunda clase de soluciones, W [la energía] tiene un valor negativo. En teoría clásica superamos esta dificultad excluyendo arbitrariamente las soluciones con W negativa. Pero en teoría cuántica no nos está permitido, ya que, en general, una perturbación causará transiciones desde estados con W positiva a estados con W negativa. [...] Por lo tanto, la teoría resultante es todavía una aproximación, aunque parece ser suficiente para explicar todos los fenómenos de duplicidad sin recurrir a premisas arbitrarias».

Eso era todo. Y ésa era la situación que provocaba los exabruptos de Heisenberg a Pauli, citados anteriormente.

A finales de 1929 —apenas dos años después— Dirac tenía una propuesta. Se basaba en el principio de exclusión de Pauli, por el cual no puede haber dos electrones que respondan a la misma solución de la ecuación de onda. Lo que Dirac proponía era una noción del vacío radicalmente nueva. Según él, lo que consideramos espacio *vacío* está repleto de electrones con energía negativa. En efecto, en palabras de Dirac, *el espacio «vacío» contiene en realidad electrones que obedecen todos ellos a las soluciones de energía negativa*. La gran ventaja de esta propuesta es que resolvía el problema de las transiciones entre soluciones positivas y negativas. Un electrón con energía positiva nunca puede pasar a una solución de energía negativa porque siempre hay otro electrón en ella y el principio de exclusión de Pauli no permite que existan dos.

A primera vista, suena un tanto extravagante que lo que percibimos como espacio vacío se encuentre lleno de cosas. Pero, si meditamos un poco, ¿por qué no? La evolución nos ha conducido a percibir aspectos del universo que son de algún modo útiles para nuestra supervivencia o nuestra capacidad reproductiva. No debería extrañarnos que determinados aspectos invariables, y sobre los que nuestra capacidad de acción es virtualmente nula, sean irrelevantes en ese sentido y escapen por completo a nuestra percepción natural. En cualquier caso, no deberíamos contar con que intuiciones simplistas sobre lo que es raro o infrecuente supongan una guía

fiable a la hora de crear modelos de las estructuras fundamentales del mundo microscópico, ya que esas intuiciones provienen de un universo de fenómenos completamente distinto. Debemos aceptar la propuesta tal cual. La validez de un modelo debe ser juzgada por lo fructíferas y precisas que sean sus consecuencias.

Así pues, a Dirac no le preocupó atentar contra el sentido común. Con buen criterio, se centró en las consecuencias observables de su hipótesis.

En la propuesta de Dirac, el vacío está —paradójicamente— lleno de electrones de energía negativa, lo cual hace que se convierta en un medio con propiedades dinámicas en sí mismo. Los fotones, por ejemplo, pueden interactuar con él. Una de las cosas que pueden suceder es que si iluminamos el vacío, aportando fotones con suficiente energía, un electrón de energía negativa puede absorber uno de ellos y trasladarse a una solución de energía positiva. La solución de energía positiva sería observada como un electrón ordinario, por supuesto. Pero en el estado final existe también un *hueco* en el vacío, ya que la solución originalmente ocupada por el electrón de energía negativa ha quedado *vacante*.

La idea de los huecos, en el contexto de un vacío dinámico, resultó tremendamente original, pero tenía sus precedentes. Dirac se basó en una analogía con los átomos pesados (que contienen muchos electrones). En dichos átomos, algunos de los electrones corresponden a soluciones de la ecuación de onda ubicadas en las proximidades del fuertemente cargado núcleo, por lo que son

atraídos con intensidad. Es necesaria mucha energía para liberar a esos electrones, por lo que en condiciones normales presentan un aspecto invariable del átomo. Pero si uno de esos electrones absorbe un fotón de alta energía (rayos X) y sale despedido del átomo, el cambio en el aspecto normal de éste viene determinado por esta *ausencia*. La ausencia de un electrón, que hubiera aportado una carga negativa, se asemeja, por contraste, a una carga positiva. La carga positiva virtual sigue la órbita del electrón ausente, con lo que tiene las propiedades de una partícula cargada positivamente.

A partir de esa analogía y en otros vagos argumentos —el artículo es muy breve y prácticamente desprovisto de ecuaciones— Dirac proponía que los huecos en el vacío eran partículas con carga positiva. El proceso por el que un fotón excita un electrón con energía negativa en el vacío, haciéndole adoptar energía positiva, se traduce a un fotón dando lugar a un electrón más una partícula cargada positivamente (el hueco). Y viceversa; si existiera previamente un hueco, un electrón de energía positiva podría emitir un fotón y ocupar la solución de energía negativa libre. Esto se interpreta como la aniquilación de un electrón y un hueco, transformándose ambos en pura energía. Hemos hablado de la emisión de un fotón, pero se trata sólo de una posibilidad más. Podrían ser emitidos varios fotones o cualquier otra forma de radiación que materialice la energía liberada.

El primer artículo de Dirac sobre los huecos se titulaba «Una teoría sobre electrones y protones». El protón, que constituye el núcleo del

átomo de hidrógeno y es un componente básico de otros núcleos más complejos, era por aquel entonces la única partícula cargada positivamente que se conocía. Era lógico tratar de identificar con él los hipotéticos huecos. Pero enseguida aparecieron notables dificultades al respecto. Por ejemplo, los dos tipos de proceso que acabamos de describir —la producción de un par electrón-protón y la aniquilación de un par electrón-protón— no habían sido observados nunca. El segundo es especialmente inquietante, ya que, mediante él, los átomos de hidrógeno se autodestruirían espontáneamente en apenas microsegundos, cosa que, afortunadamente, no sucede.

Existía también un problema lógico a la hora de identificar los huecos con los protones. Basándose en la simetría de las ecuaciones, se podía demostrar que un hueco tenía que poseer la misma masa que un electrón. Un protón, por descontado, tiene una masa muy superior.

En 1931, Dirac renunció a su identificación inicial entre huecos y protones, y aceptó la consecuencia lógica de su propia ecuación y del vacío dinámico que requería: «De existir, un hueco sería una nueva clase de partícula elemental, desconocida en la física experimental y con la misma masa que el electrón, pero con la carga opuesta».

El 2 de agosto de 1932, Carl Anderson, un experimentalista norteamericano que estudiaba las trazas dejadas por los rayos cósmicos en una cámara de niebla, observó que algunas perdían

energía según lo esperado para los electrones, pero resultaban desviadas en sentido opuesto por el campo magnético. Anderson interpretó de inmediato el fenómeno como indicativo de la existencia de una nueva partícula con la masa del electrón y la carga contraria. Ironías del destino, desconocía por completo la predicción de Dirac.

A miles de kilómetros del St. John's College de Cambridge, los huecos de Dirac —el producto de su intuición teórica— acababan de aparecer, cayendo del cielo de Pasadena.

De este modo, las *malas* noticias acabaron siendo, a la larga, noticias *aún mejores*. La rana se convirtió en príncipe: la magia en estado puro.

En la actualidad, los huecos de Dirac, conocidos hoy como positrones, ya no son un prodigio, sino una herramienta. Una de sus notables aplicaciones es la tomografía por emisión de positrones (*Positron-Electron Tomography*, PET), que sirve para fotografiar el cerebro en acción. ¿Cómo llegan hasta nuestro cerebro los positrones, dejando aparte este artículo, que los introduce de manera conceptual y no física? Mediante una inyección de moléculas que contienen átomos cuyos núcleos son radiactivos y se desintegran generando positrones, entre otras partículas. Estos positrones no llegan muy lejos antes de aniquilar a uno de los electrones vecinos, dando lugar normalmente a dos fotones que atraviesan nuestro cráneo y pueden ser detectados. A partir de ello es posible reconstruir la procedencia original de las moléculas y

trazar un mapa de nuestro metabolismo o estudiar la pérdida de energía de esos fotones en su camino hacia el exterior, y obtener un perfil de densidad y, en definitiva, una imagen del tejido cerebral.

Otra importante aplicación pertenece al campo de la física fundamental. Electrones y positrones pueden ser acelerados hasta hacerlos adquirir una elevada energía. Al confluir los haces de ambos, se aniquilan y se produce una forma altamente concentrada de «energía pura». Muchos de los progresos registrados en física fundamental en los últimos cincuenta años se han basado en estudios de este tipo, llevados a cabo en los grandes aceleradores de partículas repartidos por todo el mundo, el mayor y más moderno de los cuales es el gran colisionador electrón-positrón (*Large Electron-Positron Collider*, LEP) ubicado en el CERN, a las afueras de Ginebra. Hablaremos de uno de los hitos de esta rama de la física un poco más adelante.

Las ideas físicas de la teoría de huecos de Dirac, que, como hemos mencionado, se basaban en parte en estudios anteriores de los átomos pesados, desembocaron en la denominada física del estado sólido. En los sólidos existe una configuración básica o de referencia, dotada de la mínima energía posible, en la que los electrones ocupan todos los estados disponibles hasta un determinado nivel. Esta configuración es el equivalente al vacío en la teoría de huecos. Existen también configuraciones de energía más alta, en las que algunos de los estados de baja energía no están ocupados por electrón alguno. En esas configuraciones hay

vacantes o *huecos* —que es como técnicamente se denominan— en los que por regla general debería hallarse un electrón. Estos huecos se comportan en muchos aspectos como partículas con carga positiva. Los diodos y transistores de estado sólido se basan en una hábil manipulación de las densidades de electrones y huecos en las uniones entre distintos materiales. Una de las posibilidades más llamativas es conducir electrones y huecos a un lugar donde puedan recombinarse (aniquilarse mutuamente). Esto permite obtener una fuente de fotones que es posible controlar de manera muy precisa, dando lugar a dispositivos tan conocidos como los diodos emisores de luz (*Light-Emitting Diodes*, LED) y los láseres de estado sólido.

Desde 1932 se han observado muchos otros tipos de antipartículas. De hecho, para cada una de las partículas descubiertas hasta el día de hoy se ha encontrado también la antipartícula correspondiente. Hay antineutrones, antiprotones, antimuones (el muón es una partícula muy similar al electrón, aunque más pesada), antiquarks de varias clases, antineutrinos, antimesones π , antimesones K, etc.⁸⁵ Muchas de esas partículas no obedecen a la ecuación de Dirac y algunas tampoco lo hacen al principio de exclusión de Pauli. De lo cual se deduce que la razón física de la existencia de antimateria ha de ser muy general —mucho más de lo que fueron los argumentos que en su día condujeron a Dirac a predecir la existencia de positrones.

⁸⁵ Un caso interesante lo constituye el fotón, que es su propia antipartícula. Esto no es posible para una partícula cargada, pero el fotón es eléctricamente neutro.

Efectivamente, hay un argumento muy general por el cual, si aplicamos a la vez la mecánica cuántica y la relatividad especial, toda partícula tiene que tener su correspondiente antipartícula. La adecuada presentación del argumento requiere una fuerte formación matemática o unas dosis elevadas de paciencia, nada de lo cual quisiera exigirle al lector. Por ello nos contentaremos con una versión aproximada, que señala por qué la antimateria es una consecuencia plausible de aplicar conjuntamente la relatividad y la mecánica cuántica, sin entrar en mucho detalle.

Consideremos una partícula, que denominaremos *shmoo* (por darle algún nombre, subrayando que podría ser *cualquiera*), que se mueve hacia el este a una velocidad muy próxima a la de la luz. Según la mecánica cuántica, existe realmente cierta incertidumbre en su posición. Es decir, si medimos ésta, existe cierta probabilidad de que hallemos que, en el instante inicial, el *shmoo* está un poco más al oeste que su posición media esperada y, en un instante posterior, algo más al este que la citada posición. La partícula, pues, habría viajado un trayecto más largo del esperado para ese intervalo, lo cual significa que se habría desplazado más deprisa. Pero habíamos postulado que la velocidad esperada era prácticamente la de la luz, por lo que una velocidad mayor violaría la relatividad especial, para la cual la velocidad de la luz es la máxima posible. Por consiguiente, se trata de una paradoja.

Veamos cómo es posible resolver la paradoja por medio de las antipartículas: el *shmoo* que observamos ¿no es necesariamente el

shmoo original! También, es posible que, en el instante posterior, existan dos *shmoos*, el original y uno nuevo. En ese caso, habría también un *antishmoo*, para equilibrar la carga y conservar otras magnitudes potencialmente asociadas al *shmoo* adicional. Como casi siempre en teoría cuántica, y para evitar contradicciones, debemos ser muy específicos a la hora de pensar en lo que significa medir algo. Una forma de medir la posición del *shmoo* sería proyectando luz sobre él. Aunque para medir con exactitud la posición de un *shmoo* que se mueve rápidamente, necesitaríamos usar fotones de alta energía, por lo que entonces existe la posibilidad de que dichos fotones creen parejas *shmoo-antishmoo*. Y cuando registrásemos los resultados de la medida podríamos estar hablando del *shmoo* equivocado.

Los significados más profundos: la teoría cuántica de campos

La teoría de huecos de Dirac es verdaderamente brillante, pero la naturaleza aún va más lejos. Aunque la teoría goce de una consistencia interna y dé cobertura a una amplia gama de aplicaciones, ciertas consideraciones importantes nos obligan a ir más allá.

En primer lugar, hay partículas que no poseen espín y no obedecen a la ecuación de Dirac y, aun así, tienen antipartículas. La existencia de antipartículas es una consecuencia general de combinar la mecánica cuántica con la relatividad especial, como ya hemos visto. En concreto, por ejemplo, los mesones de carga positiva π^+ (descubiertos en 1947) y los bosones W^+ (descubiertos en

1983) desempeñan un importante papel en la física de las partículas elementales y poseen las respectivas antipartículas π^- y W^- . Pero no cabe hacer uso de la teoría de huecos para explicar el comportamiento de esas antipartículas, ya que π^+ y W^+ no respetan el principio de exclusión de Pauli. No es posible, por lo tanto, interpretar esas antipartículas como huecos en un mar de soluciones de energía negativa. Dada una solución de energía negativa, sea cual sea la ecuación que satisfaga,⁸⁶ el que esté ocupada por una partícula no impide que otra pueda adoptar el mismo estado. De modo que las transiciones catastróficas hacia estados de energía negativa, que la teoría de huecos de Dirac prohíbe a los electrones, tienen que ser evitadas de otra manera.

En segundo lugar, existen procesos en los que el número de electrones menos el de positrones cambia. Un ejemplo representativo es la desintegración de un neutrón para dar un protón, un electrón y un antineutrino. En teoría de huecos, la excitación de un electrón con energía negativa hacia un estado de energía positiva es interpretada como la creación de un par electrón-positrón y la desexcitación de un electrón con energía positiva hacia un estado de energía negativa vacante, como la aniquilación de un par electrón-positrón. En ninguno de los dos casos se altera la *diferencia* entre el número de electrones y el número de positrones. La teoría de huecos no puede explicar los cambios que se produzcan en esta diferencia. Esto significa que hay

⁸⁶ De hecho, esas partículas responden a ecuaciones de onda que presentan soluciones de energía negativa.

procesos naturales importantes, relacionados incluso con los electrones, que no encajan fácilmente en la teoría de Dirac.

La tercera y última razón nos retrotrae a la discusión inicial. Tratábamos de resolver las grandes dicotomías luz/materia y continuo/discreto. La relatividad y la mecánica cuántica, por separado, nos acercaban a la meta, y la ecuación de Dirac, con el concepto de espín, nos aproximaba todavía más. Pero aún no hemos llegado a meta. Los fotones son efímeros y en cuanto a los electrones... Estos últimos son, asimismo, efímeros —al menos, eso dicen los experimentos— y esta circunstancia no encaja del todo en nuestra argumentación teórica. En teoría de huecos, los electrones pueden aparecer y desaparecer, pero sólo cuando los positrones desaparecen o aparecen respectivamente.

Más que como contradicciones, debemos interpretar lo anterior como indicios de algo que, hoy por hoy, se nos escapa. Debe haber alguna alternativa a la teoría de huecos que dé cobertura a toda forma de materia y que trate la creación y destrucción de partículas como un fenómeno primordial.

Paradójicamente, el propio Dirac había elaborado con anterioridad el prototipo de una teoría así. En 1927, había aplicado los principios de la nueva mecánica cuántica a las ecuaciones de la electrodinámica clásica de Maxwell. Demostró que el revolucionario postulado einsteniano de que la luz está formada por partículas (fotones) era una consecuencia lógica de la aplicación de dichos principios y que las propiedades de esas partículas se derivaban

correctamente de ellos. Pocos fenómenos son tan comunes como la creación de luz a partir de la oscuridad (al encender, por ejemplo, una linterna) o su aniquilación (cuando, por poner el caso, un gato negro atraviesa el haz). Pero trasladado al lenguaje de los fotones, esto significa que la teoría cuántica de las ecuaciones de Maxwell es una teoría sobre la creación y destrucción de partículas (fotones). De hecho, en la teoría de Dirac el campo electromagnético aparece básicamente como un agente de creación y destrucción. Las partículas —los fotones— que observamos resultan de la acción de ese campo, que constituye el objeto primario. Los fotones vienen y se van, pero el campo permanece. La potencia de este planteamiento pareció habersele escapado a Dirac y a sus contemporáneos durante algún tiempo, quizá debido precisamente al aparente carácter especial de la luz (la dicotomía, una vez más). Pero se trata de un razonamiento general, que también cabe aplicar al objeto que aparece en la ecuación de Dirac, el campo del electrón.

El resultado de una aplicación lógica de los principios de la mecánica cuántica a la ecuación de Dirac es un objeto similar al que el físico británico encontró para las ecuaciones de Maxwell. Se trata de un objeto que destruye electrones y crea positrones.⁸⁷ Ambos constituyen ejemplos de *campos cuánticos*. Cuando el objeto que aparece en la ecuación de Dirac es interpretado como un campo cuántico, las soluciones de energía negativa adoptan un significado completamente distinto y sin aspectos problemáticos. Las

⁸⁷ Existe también otro objeto, muy relacionado con el anterior y conocido como *conjugado hermitiano*, que crea electrones y destruye positrones.

soluciones de energía positiva multiplican operadores de destrucción de electrones, mientras que las de energía negativa multiplican operadores de creación de positrones. En este nuevo marco, la diferencia entre los dos tipos de solución consiste en que la energía negativa representa la energía que necesitamos tomar prestada para crear un positrón, mientras que la energía positiva es la que obtenemos al destruir un electrón. Los números negativos no son aquí más paradójicos que en nuestra cuenta bancaria.

Con el desarrollo de la teoría cuántica de campos, las posibilidades que la ecuación de Dirac y la teoría de huecos pusieron en evidencia, pero no llegaron a materializar, se alcanzaron por fin. La descripción de la luz y la de la materia tuvieron finalmente una raíz común. Dirac decía, con comprensible satisfacción, que con la emergencia de la electrodinámica cuántica, los físicos habían logrado obtener las ecuaciones de partida que servían para describir «toda la química y la mayor parte de la física».

En 1932, Enrico Fermi elaboró una teoría sobre la desintegración radiactiva (desintegración beta), que incluía la desintegración del neutrón mencionada anteriormente, llevando los conceptos de la teoría cuántica de campos muy lejos de su origen. Dado que estos procesos implican la creación y destrucción de protones —el paradigma de materia *estable*—, las viejas dicotomías habían quedado claramente atrás. Tanto las partículas como la luz son epifenómenos, manifestaciones superficiales de unas realidades perdurables y más profundas, los campos cuánticos. Esos campos

Llenan por completo el espacio y, en este sentido, son continuos. Pero las excitaciones a que dan lugar son discretas, tanto si las observamos como partículas de materia o como partículas de luz.

En la teoría de huecos, nuestra idea del vacío era la de un mar de electrones con energía negativa. En la teoría cuántica de campos, la imagen es totalmente diferente. Pero aquí no hay vuelta atrás. El nuevo escenario difiere de forma aún más radical de la inocente idea de *espacio vacío*. La incertidumbre cuántica, combinada con la posibilidad de que existan procesos de creación y destrucción, implica un vacío rebotante de actividad. Pares de partículas y antipartículas que nacen continuamente para desaparecer una fracción de segundo después. En 1987 escribí un soneto, titulado «Partículas virtuales», que describe el escenario:

*No pienses que no hay nada ahí
Por más que vacíes; por mucho que hagas
dejarás detrás un frenesí incansable
de fútiles clones innumerables.
Nacen en un parpadeo y danzan alrededor;
todo lo que tocan comienza a dudar:
¿qué hago aquí? ¿Cuánto he de pesar?
Dichos pensamientos suelen conducir a la desintegración.
¡No temáis! La terminología es equívoca;
la desintegración es la reproducción de las partículas
y el frenesí, aunque inconsciente, puede servir a nobles
fines,*

la materia, intercambiada, sirve para crear lazos de amistad.

*¿Ser o no ser? La elección parece obvia,
pero Hamlet vaciló. La materia también lo hace.*

Consecuencias

Con la génesis de la teoría cuántica de campos, llegamos a la frontera intelectual natural en nuestro recorrido en torno a la ecuación de Dirac. A mediados de la década de 1930, las paradojas inmediatas a las que la ecuación dio lugar habían quedado resueltas y sus objetivos iniciales estaban plenamente cubiertos. Dirac recibió el Premio Nobel en 1933 y el Premio Anderson en 1935.

Los años posteriores fueron de profundización en la teoría cuántica de campos y de extensión de sus aplicaciones. Por medio de ella, los físicos han elaborado y establecido, con asombroso grado de rigor y por encima de cualquier duda razonable, lo que durante muchos años —o tal vez para siempre— será la teoría fundamental sobre la materia. El modo en que los acontecimientos han tenido lugar y la naturaleza de la teoría constituyen una historia épica que abarca muchas otras ideas y en la que la ecuación de Dirac ha desempeñado un papel relevante, aunque no protagonista. No obstante, algunos desarrollos posteriores están tan fuertemente vinculados a nuestro tema principal y son tan atractivos por ellos mismos que merecen ser mencionados en este ensayo.

La génesis de la teoría cuántica de campos marca un límite natural también en otro sentido. Es el límite que el propio Dirac no se atrevió a cruzar. Al igual que Einstein, Dirac siguió un camino distinto en sus últimos años. Ignoró los trabajos de la mayoría de los físicos y discrepó del resto. En los maravillosos desarrollos a los que dio pie su obra, Dirac desempeñó un papel marginal.

La EDC y los momentos magnéticos

La interacción con el omnipresente vacío dinámico de la teoría cuántica de campos modifica las propiedades observadas en las partículas. En lugar de ver las propiedades teóricas de las partículas *desnudas*, contemplamos partículas físicas *ataviadas* con su interacción con las fluctuaciones cuánticas del vacío dinámico.

En particular, el electrón físico no es el electrón *desnudo*, por lo que no satisface por completo la condición $g = 2$ de Dirac. Cuando, en 1947, Polykarp Kusch realizó medidas muy precisas de este parámetro, encontró que g era mayor que 2 en un factor de 1,00119. Cuantitativamente, no era una desviación muy grande, pero supuso un gran estímulo para la física teórica ya que representaba un reto muy concreto. En aquella época, había tantos cabos sueltos en la física fundamental que era difícil saber por dónde empezar. (Toda una constelación de partículas recién descubiertas —entre ellas, el muón y el mesón π —, ninguna teoría que explicara satisfactoriamente qué fuerza mantiene unido el núcleo atómico, resultados fragmentarios e injustificados en

relación con la desintegración radiactiva, anomalías en los rayos cósmicos de alta energía...). De hecho, existía un conflicto filosófico básico sobre la estrategia a seguir.

Los más veteranos —los fundadores de la teoría cuántica, incluyendo a Einstein, Schrödinger, Bohr, Heisenberg y Pauli— se preparaban para otra revolución. Pensaban que resultaba inútil emplear tiempo en realizar cálculos más precisos en electrodinámica cuántica o EDC (*Quantum Electrodynamics*, QED), ya que esta teoría era, seguramente, incompleta y, probablemente, errónea. Tampoco ayudaba mucho el que los cálculos requeridos para lograr esa precisión fueran muy complejos y el que a veces revelaran resultados sin sentido (infinitos). Por ello, los viejos maestros trataban de encontrar una teoría de otra clase, aunque, por desgracia, sin adoptar una línea de trabajo clara.

Ironías del destino, fue una joven generación de teóricos —Julian Schwinger, Richard Feynman y Freeman Dyson en Estados Unidos y Sinitiro Tomonaga en Japón— la que asumió el papel conservador.⁸⁸ Hallaron un modo de efectuar esos cálculos más exactos y obtener resultados válidos sin cuestionar la teoría fundamental. La teoría, en efecto, era exactamente la que Dirac había elaborado en las décadas de 1920 y 1930. En 1947, Schwinger realizó un nuevo cálculo que incluía los efectos del vacío dinámico. El resultado fue una corrección a $g = 2$ que concordaba espectacularmente con las medidas de Kusch. Aún habría otros

⁸⁸ También fueron importantes las contribuciones hechas por los teóricos, de algo más edad, Kramers y Bethe y por Lamb, un teórico convertido en experimentalista.

éxitos. Kusch recibió el Premio Nobel en 1955 y Schwinger, Feynman y Tomonaga lo obtendrían conjuntamente diez años después —es difícil entender este retraso.

Para sorpresa de todos, Dirac no aprobó los nuevos procedimientos. Tal vez la cautela estuviera justificada al principio, cuando los métodos matemáticos resultaban poco habituales, no estaban bien definidos e implicaban ciertas dosis de improvisación. Pero todas esas dificultades técnicas habían sido superadas ya. Aunque la EDC presenta problemas de base, contemplada —de manera optimista— como una teoría cerrada, dichos problemas son de una índole muy diferente de la que preocupaba a Dirac y quedan plausiblemente resueltos si se enmarca la EDC en una teoría más amplia y asintóticamente libre (ver más adelante). Por otra parte, el hecho tiene pocas repercusiones prácticas en la mayoría de sus predicciones.

Feynman decía de la EDC que era «la joya de la física, nuestro máspreciado bien». Pero, en 1951, Dirac escribiría: «Los últimos trabajos de Lamb, Schwinger, Feynman y otros han obtenido un gran éxito [...], pero la teoría resultante es antiestética e incompleta». Y en su último artículo, en 1984, diría: «Esas reglas de renormalización hacen que los resultados concuerden sorprendente y excesivamente bien con los experimentos. La mayor parte de los físicos concluyen que estas reglas son, por lo tanto, correctas. Yo creo que no es un argumento válido. El hecho de que los resultados concuerden con los experimentos no demuestra que una teoría sea correcta».

El lector percibirá, sin duda, un cierto contraste con la actitud del joven Dirac, quien se aferraba a ultranza a su ecuación, precisamente porque explicaba los resultados experimentales.

En la actualidad, el valor del momento magnético del electrón (es decir, su fuerza magnética), determinado experimentalmente, es:

$$(g/2)_{\text{experimental}} = 1,001\ 159\ 652\ 188\ 4(43)$$

mientras que la predicción teórica, basada en la EDC y calculada con gran precisión, es:

$$(g/2)_{\text{teórico}} = 1,001\ 159\ 652\ 187\ 9(88)$$

(Se indica la incertidumbre en los dos últimos dígitos). Se trata quizá de la confrontación más precisa y difícil de toda la historia de la ciencia entre un cálculo teórico especialmente complejo —aunque muy bien definido— y un fascinante experimento. Es lo que Feynman quería decir al hablar de «nuestro máspreciado bien».

Al tiempo de redactar el presente artículo, determinar de forma aún más exacta el momento magnético del electrón y el de su pariente el muón sigue constituyendo un reto para la física experimental. Con las precisiones alcanzables hoy, los resultados serán sensibles a los efectos de las fluctuaciones cuánticas debidas a potenciales nuevas partículas pesadas, en particular, a las que se prevé estén asociadas a la supersimetría.

La CDC y la teoría sobre la materia

El momento magnético del protón no sólo no satisface la condición de Dirac ($g = 2$), sino que vale aproximadamente 5,6. El caso de los neutrones es aún peor. El neutrón es eléctricamente neutro; la ecuación original de Dirac no predice momento magnético alguno para él. El hecho es que el neutrón posee un momento magnético que es aproximadamente dos tercios del correspondiente al protón y presenta la orientación opuesta respecto al espín, lo cual corresponde a un valor infinito de g . Las discrepancias halladas en estos momentos magnéticos fueron el primer síntoma de que tanto protones como neutrones son objetos más complejos que el electrón. En los estudios realizados con posterioridad surgieron muchas más complicaciones. Las fuerzas entre los protones y los neutrones resultaron ser muy complejas, dependientes no sólo de las distancias entre ellos, sino también de sus velocidades, orientaciones de espín y todo tipo de caprichosas combinaciones entre estos parámetros. En realidad, pronto se vio que no eran *fuerzas* en el sentido tradicional del término. La existencia de una fuerza entre dos protones, en el sentido tradicional, significaría que el movimiento de uno de ellos podría verse afectado por la presencia del otro; es decir, si lanzásemos un protón contra otro, éste se desviaría. Lo que en la práctica se observa es que, cuando un protón colisiona con otro, suele surgir toda una pléyade de partículas, la mayoría de las cuales son muy inestables. Aparecen mesones π , mesones K , mesones ρ , bariones Λ y Σ , sus

antipartículas correspondientes y muchas más. Todas esas partículas interactúan fuertemente entre ellas. De este modo, el problema de las fuerzas nucleares, la frontera de la física desde la década de 1930, se convirtió en el reto de comprender todo un nuevo y amplio mundo de partículas y las reacciones que se producían entre ellas, reacciones que parecían ser las más potentes de la naturaleza. Hasta la terminología tuvo que cambiar. Los físicos dejaron de hablar de fuerzas nucleares y comenzaron a hablar de la interacción fuerte.

En la actualidad sabemos que es posible describir todas las complejidades de la interacción fuerte, a nivel fundamental, mediante una teoría denominada cromodinámica cuántica o CDC (*Quantum chromodynamics*, QCD), que es una amplia generalización de la EDC. Los componentes elementales de la CDC son los quarks y los gluones. Hay seis tipos o *sabores* de quarks: *u*, *d*, *s*, *c*, *b*, *t* (*up*-arriba, *down*-abajo, *strange*-extraño, *charm*-encanto, *bottom*-base, *top*-techo). Reminiscencia de los leptones cargados,⁸⁹ los quarks son muy parecidos unos a otros, difiriendo básicamente en su masa. En la materia ordinaria sólo se hallan presentes los dos más ligeros, *u* y *d*. Haciendo una analogía con los componentes básicos de la EDC, los quarks desempeñan más o menos el papel de los electrones y los gluones, el de los fotones. La gran diferencia es que, mientras en la EDC sólo hay un tipo de carga y un fotón, en la CDC tenemos tres

⁸⁹ «Leptón» es sólo un término genérico que abarca electrones, muones, las denominadas partículas tau y sus correspondientes antipartículas. Todas estas partículas tienen propiedades muy similares, incluyendo igual carga y espín, y difieren en la masa.

tipos de carga —los denominados *colores*— y ocho gluones. Algunos gluones responden a cambios de color, de manera similar a como los fotones responden a la carga eléctrica. Otros sirven de intermediarios en las transiciones entre un color y otro. Por ejemplo, un quark u con carga azul puede radiar un gluón y convertirse en un quark u con carga verde. Como la carga global debe conservarse, ese gluón ha de tener carga azul +1 y carga verde -1. Dado que los gluones transportan carga de color no equilibrada, en CDC existen procesos elementales en los que los gluones radian otros gluones. No hay nada equivalente en la EDC. Los fotones son eléctricamente neutros y, en términos generales, no interaccionan unos con otros. Gran parte de la riqueza y complejidad de la CDC proviene de esta nueva característica.

Presentada de este modo, sin profundizar en los conceptos o en los fenómenos, la CDC puede parecer extravagante y arbitraria. En realidad, la CDC es una teoría de gran simetría y belleza matemática. Desgraciadamente, no disponemos de espacio para hacerle justicia aquí. Pero le debo al lector algunas explicaciones.

¿Cómo hemos llegado a una teoría de este tipo? Y ¿cómo sabemos que es correcta? En el caso de la CDC, se trata de dos cuestiones muy distintas. El recorrido histórico hacia su descubrimiento fue tortuoso, plagado de pistas falsas y callejones sin salida. Pero, contemplado retrospectivamente, no tuvo por qué ser así. De haber existido antes los aceleradores de ultra-alta energía, la CDC se

hubiera materializado ella sola ante nuestros asombrados ojos.⁹⁰ La narración que sigue reúne la mayoría de las ideas que hemos examinado hasta ahora y me parece un epílogo adecuado para esta parte del artículo.

Cuando aceleramos electrones y positrones hasta imprimirles una energía muy elevada y los hacemos colisionar, observamos dos tipos de sucesos. En uno de ellos, las partículas del estado final son leptones y fotones. En esta clase de suceso, el estado final suele consistir simplemente en un leptón y su antileptón; pero en alrededor de un 1 por ciento de los casos existe también un fotón y en un 0,01 por ciento, el número de fotones es dos. La probabilidad de esos sucesos, y la de que las diversas partículas adopten distintos ángulos y energías, pueden ser computadas mediante EDC y los resultados son muy satisfactorios. Por el contrario, si no supiésemos nada de EDC, podríamos deducir las reglas básicas de la interacción fundamental de la EDC —es decir, la emisión de un fotón por parte de un electrón— sólo con analizar los citados sucesos. La interacción fundamental de la luz con la materia se halla ante nuestras propias narices.

En el otro tipo de suceso observamos algo radicalmente distinto. En lugar de dos o, como mucho, media docena de partículas, surgen muchas y muy diferentes. Aparecen cosas como mesones π , mesones K, protones, neutrones y sus antipartículas (partículas todas ellas que, a diferencia de los fotones y los leptones, producen

⁹⁰ Hasta llegar a un par de problemas profundos, pero bien definidos y resolubles, como veremos enseguida.

interacciones fuertes). La distribución angular de esas partículas es muy específica. En vez de surgir de manera independiente, cada una por su lado, emergen en contadas direcciones, formando delgados *chorros*. En un 90 por ciento de los casos sólo se producen dos chorros, saliendo en sentidos opuestos. En alrededor del 10 por ciento, los chorros son tres, y en un 1 por ciento, cuatro —imagine el lector su distribución espacial.

Ahora bien, si no aplicamos la lupa en busca de partículas individuales y nos limitamos a seguir el flujo de energía y momento, los dos tipos de suceso —el que da lugar a *partículas EDC* y el de los chorros de partículas con interacciones fuertes— ¡parecen el mismo! En esta historia imaginaria sería difícil no sucumbir a la tentación de tratar los chorros como si fuesen partículas y establecer reglas para la probabilidad de los diferentes patrones de radiación —con los correspondientes números, ángulos y energías para las partículas-chorro— análogas a las empleadas con éxito en EDC. Y la cosa funcionaría muy bien, ya que las reglas que realmente describen las observaciones son muy parecidas a las de la EDC. Por supuesto, las reglas adecuadas son justamente las de la CDC, que incluyen los nuevos procesos por los que un gluón puede generar otro. Todas esas reglas —los cimientos de la teoría— podrían haber sido deducidas directamente a partir de los datos. *Quark y gluón* serían conceptos con una definición operativa directa y precisa, en términos de chorros. Como decíamos, todo se halla ante nuestros ojos (... una vez hemos comprendido hacia dónde hay que mirar).

Aun así, nos enfrentaríamos a dos grandes rompecabezas conceptuales. ¿Por qué los experimentos muestran *quarks* y *gluones* en lugar de quarks y gluones simplemente —es decir, chorros en lugar de simples partículas—? Y ¿cómo relacionamos esos conceptos teóricos que describen directa y exitosamente los sucesos de alta energía con el resto de los fenómenos de la interacción fuerte? La conexión entre la supuesta teoría fundamental y las observaciones diarias en absoluto resulta obvia. Por ejemplo, si deseáramos construir protones a partir de los *quarks* y los *gluones* que aparecen en la teoría fundamental, no sería posible, ya que los chorros que sirven para definir operativamente esas *partículas* a menudo contienen ellos mismos protones, entre otras cosas.

Existe una solución elegante para ambos problemas. Se trata del fenómeno de la *libertad asintótica* en CDC. Conforme a la libertad asintótica, los sucesos de radiación que impliquen grandes cambios en el flujo de energía y momento son infrecuentes, mientras que los que sólo supongan pequeños cambios son muy comunes. La libertad asintótica no es una premisa aparte, sino una consecuencia matemática profunda de la estructura de la CDC.

La libertad asintótica explica claramente por qué surgen los chorros en las aniquilaciones electrón-positrón a altas energías, el tipo de suceso asociado a partículas con interacciones fuertes. Inmediatamente después de que el electrón y el positrón se aniquilen mutuamente, aparecen un quark y un antiquark que se mueven a gran velocidad y en sentidos opuestos. Ambos radian

gluones enseguida y éstos radian, a su vez, dando origen a una compleja cascada integrada por multitud de partículas. Pero a pesar de toda esa conmoción, el flujo global de energía y momento no se ve perturbado significativamente. Las radiaciones que perturban el citado flujo son poco habituales, según la libertad asintótica. Así pues, tenemos un puñado de partículas que se mueven todas ellas en la misma dirección, la establecida inicialmente por el quark y el antiquark. En definitiva, se ha producido un chorro. Cuando tiene lugar uno de esos raros sucesos de radiación que perturban el flujo de energía y momento, el gluón radiado origina un chorro por sí mismo, con lo que tendríamos un suceso de tres chorros, y así sucesivamente.

La libertad asintótica indica también por qué los protones (y otras partículas con interacciones fuertes), que se manifiestan como entidades individuales estables o cuasi estables, son en realidad objetos complejos. Estas partículas son, por definición, configuraciones de quarks, antiquarks y gluones dotadas de un grado de estabilidad razonable. Ni qué decir tiene que, como la probabilidad de que quarks, antiquarks y gluones radien es muy alta, no será válida cualquier configuración. Para que una configuración dada sea estable deberá existir un equilibrio dinámico, en el que la emisión de radiación en una parte del sistema quede equilibrada mediante su absorción en otra.

Lo cierto es que la libertad asintótica fue descubierta en el campo teórico (por David Gross y por mí e, independientemente, por David

Politzer) y la CDC, propuesta en 1973 (por Gross y por mí) como candidata a teoría de la interacción fuerte, basándonos en evidencias mucho menos directas. La existencia de los chorros y sus propiedades fueron una *predicción* teórica, antes de ser constatados experimentalmente. En la actualidad y gracias a la evidencia experimental acumulada, la CDC ha sido aceptada como teoría fundamental de la interacción fuerte, junto con la EDC para la interacción electromagnética.

Mediante la CDC se ha registrado un enorme progreso en la descripción de las propiedades del protón, el neutrón y las demás partículas con interacciones fuertes. Esto ha obligado a realizar complejos y tediosos cálculos numéricos en los ordenadores más potentes, pero el resultado ha merecido la pena. Un hecho a destacar es que es posible calcular las masas del protón y el neutrón a partir de principios fundamentales y sin emplear parámetros libres importantes. Como ya se ha indicado, desde un punto de vista fundamental, las partículas citadas constituyen un complejísimo equilibrio dinámico entre quarks, antiquarks y gluones. La mayor parte de su masa —y, por lo tanto, la de toda la materia conocida, incluyendo nuestros propios cuerpos— proviene de la pura energía de esos objetos —que, esencialmente, ellos mismos carecen de masa— al moverse, de acuerdo con $m = E/c^2$. A ese nivel, al menos, somos criaturas etéreas.

Dirac decía que la EDC describía «toda la química y la mayor parte de la física» y, en efecto, es la teoría fundamental para la estructura

externa del átomo (y para mucho más). En el mismo sentido, la CDC es la teoría fundamental de los núcleos atómicos (y de mucho más). Juntas, conforman una teoría de la materia notablemente completa, compacta, fructífera y contrastada.

La fecundidad de la razón

Acabamos de ver cómo el «jugar con las ecuaciones» condujo a Dirac hasta una ecuación cargada de significados que él jamás previó, y a los que en muchos casos se opuso, pero que demostró ser correcta y enormemente fructífera. ¿Cómo pudo suceder algo así? ¿Pueden ser las matemáticas verdaderamente creativas? ¿Cabe realmente llegar, mediante cálculos o procesos lógicos, a percepciones esencialmente nuevas, a obtener, en definitiva, más de lo que se invirtió?

La cuestión es de plena actualidad, pues incide en el corazón del debate relativo a la inteligencia de las máquinas. (En si éstas podrían llegar a desarrollar una especie de mente semejante a la humana o incluso superior).

A primera vista, los argumentos en contra parecen irrefutables. El de más peso, al menos desde el punto de vista psicológico, es el de la introspección. Al reflexionar sobre nuestros propios procesos mentales, no podemos evitar sentir que no consisten exclusivamente —o, al menos, principalmente— en manipulaciones de símbolos basadas en reglas preestablecidas. Simplemente, no percibimos que sea así. Por lo general, pensamos en términos de imágenes y emociones, no de meros símbolos. Y el flujo de nuestros

pensamientos se ve constantemente estimulado y redirigido por interacciones con el mundo exterior y por impulsos internos de una forma que difiere claramente del modo en que se desenvuelve un algoritmo matemático.

Otro argumento proviene de nuestra experiencia con los modernos ordenadores digitales, los cuales son, en cierto sentido, unos matemáticos ideales. Son infinitamente más rápidos, infalibles e infatigables que cualquier ser humano a la hora de aplicar unas reglas (axiomas) precisas. Y en muchas tareas especializadas, tales como gestionar los vuelos de una compañía aérea o las rutas de distribución de carburantes para maximizar beneficios, sobrepasan con mucho la capacidad humana. Y, aun así, respecto al patrón humano, hasta el ordenador más potente resulta frágil, limitado y bastante estúpido. Un pequeño error de programación, unos *bytes* de código malicioso o un fallo en la memoria pueden bloquear la máquina más moderna o hacerla adoptar un comportamiento autodestructivo. La comunicación se limita a formatos rígidamente establecidos, ajenos a la riqueza del lenguaje natural. Los resultados absurdos pueden —y suelen— emerger sin ningún tipo de pudor ni censura.

Sin embargo, un análisis más profundo de estos argumentos hace aflorar algunas dudas. Aunque la naturaleza del vínculo entre los patrones de las señales eléctricas en las células nerviosas con los procesos mentales asociados sigue siendo misteriosa en muchos aspectos, hoy se conoce una pequeña parte en relación, sobre todo,

con las primeras etapas del proceso sensorial. Nada en lo descubierto hasta ahora sugiere la presencia de algo más exótico que simples señales eléctricas y químicas que siguen leyes físicas perfectamente conocidas. La inmensa mayoría de los científicos acepta como hipótesis de trabajo la existencia de un nexo entre los patrones de las señales eléctricas y el pensamiento. El patrón formado por los fotones que inciden en nuestra retina es descompuesto y analizado en unidades elementales, conducido a través de un sinnúmero de canales diferentes, procesado y —de algún modo— reensamblado después para proporcionarnos una «imagen del mundo» engañosamente simple, organizada en forma de objetos en el espacio y que tendemos a considerar obvia. El hecho es que no tenemos ni la más remota idea de cómo realizamos la mayor parte de lo que hacemos, incluso —o, más bien, en particular— nuestros procesos mentales más simples. Quienes han tratado de construir máquinas capaces de reconocer los objetos que muestra una fotografía o de pasearse y explorar el entorno como un bebé, han sufrido experiencias frustrantes, a pesar de que, ellos mismos, puedan hacer todo eso sin ninguna dificultad. Pero no pueden enseñar a otros, sencillamente porque no saben *cómo* lo hacen. En definitiva, la introspección no es una guía fiable a la hora de comprender la estructura última del pensamiento, tanto en lo que creemos conocido, como en lo desconocido.

En cuanto a los ordenadores, todo veredicto negativo es seguramente prematuro, ya que evolucionan muy deprisa. Un hito

reciente es la victoria de *Deep Blue* sobre el campeón mundial de ajedrez Gary Kasparov en un torneo rápido. Nadie se negaría a aceptar que un juego de ese nivel habría sido considerado profundamente creativo de haber sido llevado a cabo por un jugador humano. Sin embargo, un éxito así en un campo limitado no hace sino exacerbar la cuestión: ¿Qué es lo que falta, qué es lo que impide la emergencia de creatividad a partir del puro cálculo en un frente más amplio? Creo que ciertos casos-ejemplo pueden resultar de gran valor a la hora de responder esta trascendental pregunta.

En la física moderna, y tal vez en toda la historia de la ciencia, ningún episodio ilustra mejor la naturaleza profundamente creativa del razonamiento matemático que la historia de la ecuación de Dirac. Sabemos *a posteriori* que lo que Dirac trataba de hacer es estrictamente imposible. Las leyes de la mecánica cuántica, tal como se conocían en 1928, no pueden ser consistentes con la relatividad especial. Y, aun así, partiendo de supuestos inconsistentes, Dirac llegó a una ecuación que hoy sigue siendo una de las piedras angulares de la física.

Nos hallamos, pues, ante un ejemplo concreto, significativo y bien documentado de cómo un razonamiento matemático sobre el mundo físico, culminando en una ecuación específica, condujo a resultados que supusieron toda una sorpresa para el propio pensador. En aparente desafío a algunas leyes de conservación, obtuvo mucho más de lo que invirtió. ¿Qué hizo posible ese salto? ¿Cómo lo logró Dirac, en particular? ¿Qué llevó a Dirac y a sus contemporáneos a

aferrarse insistentemente a la ecuación cuando ésta los arrastraba «mar adentro»?⁹¹

En los comentarios de Dirac hallamos algunas pistas. En su breve ensayo «Mi vida como físico», se enorgullece de su formación como ingeniero: «Los estudios de ingeniería ejercieron una fuerte influencia en mí. [...] Aprendí que, al describir la naturaleza, se deben aceptar las aproximaciones y que incluso el trabajo basado en ellas puede ser interesante y, en ocasiones, hermoso». En este sentido, una de las causas de la fe inicial de Dirac (y de la de otros) en su ecuación, lo que les movía a tolerar sus aparentes fallos, era simplemente que podían encontrar soluciones aproximadas que concordaban brillantemente con los datos experimentales para el espectro del hidrógeno. En sus primeros artículos, Dirac se complacía en mencionar —sin que pareciera preocuparle demasiado— que existían otras soluciones, matemáticamente válidas también, pero carentes de toda interpretación física razonable.

En lo que a primera vista puede parecer un enfoque completamente distinto, Dirac subrayaba a menudo el poder heurístico de la belleza matemática: «Al tratar de expresar en forma matemática las leyes fundamentales de la naturaleza, el investigador debería perseguir sobre todo la belleza matemática». Éste fue otro de los pilares que al

⁹¹ Mucho después, en la década de 1960, Heisenberg recordaba: «Hasta aquel momento [1928] tenía la impresión de que, en teoría cuántica, habíamos arribado a puerto. El artículo de Dirac nos arrojó al mar de nuevo».

principio sostuvieron su fe en la ecuación: era (y es) extraordinariamente hermosa.

Por desgracia, es difícil precisar y casi imposible transmitir al lector profano la naturaleza de la belleza matemática. Pero cabe hacer analogías con otras clases de belleza. Una faceta que puede hacer que sea hermosa una pieza musical, una novela o una obra de teatro es la acumulación de tensión entre temas trascendentes y bien desarrollados, tensión que luego es resuelta de manera convincente y sorprendente. Un aspecto que puede convertir en bella una escultura o una obra arquitectónica es la simetría —el equilibrio en las proporciones y la complejidad dirigida a un fin—. La ecuación de Dirac posee todas esas características en grado sumo.

Recordemos que Dirac trataba de conciliar la mecánica cuántica de los electrones con la relatividad especial. Es hermoso contemplar cómo la tensión entre los requisitos de simplicidad y relatividad puede ser resuelta y hallar que, esencialmente, sólo hay un modo de hacerlo. Éste es uno de los aspectos de la belleza matemática de la ecuación de Dirac. Otro de ellos, su equilibrio y simetría, es casi sensual. El espacio y el tiempo, el momento y la energía, aparecen en escena. Los diversos términos del sistema de ecuaciones han de ser coreografiados siguiendo la música de la relatividad y los patrones de 0s y 1s (e *is*) danzan ante nuestros ojos.

Las líneas convergen cuando la física conduce a la belleza matemática, o en los raros y mágicos momentos en los que las

matemáticas nos llevan hasta la verdad física. Dirac buscaba una ecuación que satisficiera hipótesis motivadas por la realidad física. Encontró que, para obtenerla, eran precisas cuatro componentes (un sistema de ecuaciones). Fue la primera sorpresa. Dos de las componentes eran bienvenidas: representaban con claridad las dos direcciones posibles del espín del electrón. Pero las otras dos, al principio, no tenían una interpretación física convincente y cuestionaban incluso el significado de la ecuación misma. Sin embargo, la ecuación pareció cobrar vida propia y trascender las ideas que la habían alumbrado. Y así, poco tiempo después, las dos componentes adicionales presagiaban el positrón, como ya vimos.

Con esta convergencia llegamos al corazón del método seguido por Dirac en su búsqueda de la ecuación que lleva su nombre, similar al empleado por James Clerk Maxwell al deducir las suyas y por el propio Einstein cuando creó sus teorías especial y general de la relatividad. Todos ellos procedieron mediante *lógica experimental*, un término contradictorio sólo en apariencia. En lógica experimental se formulan hipótesis en forma de ecuaciones y se experimenta con esas ecuaciones. Es decir, se intenta perfeccionar las ecuaciones desde el punto de vista de la belleza y la consistencia, y seguidamente se verifica si esas ecuaciones *mejoradas* explican algún aspecto de la naturaleza. Los matemáticos utilizan la técnica de la «reducción al absurdo»: para demostrar A , se asume el opuesto de A y se llega a una contradicción. La lógica experimental es una «validación por fecundidad»: para validar A , se asume A y se

demuestra que conduce a resultados útiles. Frente al *modus operandi* de la lógica deductiva, la lógica experimental se inspira en la máxima: «Más vale pedir perdón que solicitar permiso». De hecho, como ya hemos visto, la lógica experimental no contempla la inconsistencia como una catástrofe irremediable. Si una línea de investigación tiene cierto éxito y es fructífera, no debería ser abandonada según su inconsistencia o su carácter aproximado. Por el contrario, se ha de buscar el modo de convertirla en correcta.

Considerando todo lo anterior, volvamos a la cuestión de la creatividad en el razonamiento matemático. Decíamos que, en cierto sentido, los modernos ordenadores son como matemáticos ideales. En el seno de cualquier rama de las matemáticas dotada de axiomas razonables y precisos, sabemos cómo programar un ordenador de forma que demuestre sistemáticamente todos los teoremas válidos.⁹² Ejecutando un programa, una máquina de esta clase produciría teoremas válidos con mucha más velocidad y precisión que cualquier matemático humano. Pero intentar crear matemática avanzada mediante un programa así no sería mucho mejor que poner a una manada de monos a aporrear máquinas de escribir y esperar que en alguna de ellas aparezca una obra de Shakespeare.

⁹² Se trata de una consecuencia del teorema de la completitud de Gödel para la lógica de primer orden. El lector informado se preguntará cómo el que todos los teoremas válidos puedan ser demostrados de manera mecánica es consistente con el famoso teorema de la incompletitud, del propio Gödel. (No es una errata de imprenta: Gödel demostró tanto el teorema de la completitud como el teorema de la incompletitud. ¡Un tipo astuto!). Para decirlo en pocas palabras, el teorema de la incompletitud de Gödel dice que en cualquier sistema matemático cabe formular proposiciones válidas en las que ni dicha proposición ni su negación constituyen un teorema. Esta «incompletitud» no excluye la posibilidad de enumerar sistemáticamente todos los teoremas.

Obtendríamos miles de teoremas ciertos, pero la inmensa mayoría de ellos serían triviales, con el grano irremisiblemente entremezclado con la paja. En la práctica, si el lector examina con detenimiento revistas de matemáticas o de física matemática —y no digamos revistas literarias— no encontrará muchos trabajos propuestos por ordenadores. Los intentos de enseñar a los ordenadores a hacer matemática creativa *auténtica* —como los de enseñarles a reconocer objetos o a navegar por un escenario real— han cosechado éxitos muy limitados. En el fondo, se trata de problemas con muchos puntos en común. La matemática y la física creativas no se basan en la estricta lógica, sino en la lógica experimental. La lógica experimental implica reconocer patrones, jugar con ellos, hacer suposiciones para explicarlos y, especialmente, apreciar la belleza. Y la física creativa requiere más: la capacidad de construir patrones del mundo que nos rodea y de valorar no sólo su consistencia lógica, sino también su fidelidad (¡aproximada!) a la realidad.

¿Puede ser creativo el razonamiento puramente matemático? Indudablemente, si es a la manera de Dirac, unido a la capacidad de tolerar aproximaciones, de apreciar la belleza y de aprender mediante la interacción con el mundo real. Cada uno de esos factores ha desempeñado un papel en todos los grandes progresos registrados en la física. El reto consiste, pues, en cómo plasmar esas capacidades en mecanismos concretos.

Apéndice

Habíamos presentado la ecuación de Dirac en la forma:

$$\left[\gamma^\mu \left(i \frac{\partial}{\partial x^\mu} - e A_\mu(x) \right) + m \right] \psi(x) = 0$$

Identifiquemos ahora sus términos. La función de onda $\psi(x)$ es el objeto cuyo comportamiento describe la ecuación. Posee cuatro componentes: $\psi_{e\uparrow}(x)$, $\psi_{e\downarrow}(x)$, $\psi_{p\uparrow}(x)$, $\psi_{p\downarrow}(x)$. Cada una de ellas es una función cuyo valor depende del espacio y el tiempo, tal como indica el argumento (x) . Para Dirac, esos valores eran números complejos, cuyo cuadrado viene a representar (en términos genéricos) la probabilidad de encontrar el correspondiente tipo de partícula: un electrón con el espín hacia arriba, un electrón con el espín hacia abajo, un positrón con el espín hacia arriba o un positrón con el espín hacia abajo. En la interpretación moderna, los valores son operadores que crean electrones o destruyen positrones.

Como es habitual en las teorías relativistas, se asume la convención de suma de Einstein. Se supone que los subíndices y superíndices μ toman los valores 0, 1, 2, 3, representando el tiempo y las tres direcciones espaciales, y engloban las contribuciones de esos cuatro valores. El operador derivada $\partial/\partial x^0$ mide cuán rápido cambia la función de onda a lo largo del tiempo y los otros operadores de este tipo, la tasa de cambio de dicha función a lo largo de las tres coordenadas espaciales. Los campos $A(x)$, con varios subíndices, son los potenciales electromagnéticos. Especifican los campos magnéticos y eléctricos que experimenta el electrón. La carga del

electrón es $-e$. Representa la intensidad de su respuesta a esos campos. La masa del electrón es m .

La innovación técnica más característica de Dirac fue la introducción de las matrices γ :

$$\gamma^0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \gamma^1 = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$$\gamma^2 = \begin{pmatrix} 0 & 0 & 0 & i \\ 0 & 0 & -i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix} \quad \gamma^3 = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}$$

Los demás elementos de su ecuación —función de onda, derivadas, potenciales electromagnéticos, carga y masa— aparecían ya en la ecuación de Schrödinger. Las matrices γ eran una novedad absoluta. Permitían a Dirac formular una ecuación en la que tiempo y espacio aparecían al mismo nivel y, a la vez, le obligaban a introducir una función de onda con cuatro componentes.

Expresada con detalle, la ecuación de Dirac dice así:

$$\begin{pmatrix} i\partial_0 - eA_0 + m & 0 & -i(\partial_1 + \partial_3) + e(A_1 + A_3) & \partial_2 + ieA_2 \\ 0 & i\partial_0 - eA_0 + m & -\partial_2 - ieA_2 & -i(\partial_1 - \partial_3) + e(A_1 - A_3) \\ i(\partial_1 - \partial_3) - e(A_1 - A_3) & -\partial_2 - ieA_2 & -i\partial_0 - eA_0 + m & 0 \\ \partial_2 + ieA_2 & i(\partial_1 - \partial_3) - e(A_1 - A_3) & 0 & -i\partial_0 + eA_0 + m \end{pmatrix} \begin{pmatrix} \psi_{e+}(x) \\ \psi_{e-}(x) \\ \psi_{p+}(x) \\ \psi_{p-}(x) \end{pmatrix} = 0$$

Agradecimientos

El presente trabajo ha sido financiado en parte por el Departamento de Energía de Estados Unidos a través del acuerdo de investigación #DF-FC02-94ER40818. Marty Stock colaboró con LaTeX.

Lecturas recomendadas

Como material de lectura sobre física atómica y teoría cuántica, incluyendo extractos de importantes fuentes originales, recomiendo especialmente *The World of the Atom*, de H. Boorse y L. Motz (Basic Books, 1996). Los capítulos más trascendentes parecen recién escritos.

El clásico de Dirac es *The Principles of Quantum Mechanics*, 4.^a ed., Cambridge University Press, 1958.

R. P. Feynman, *QED: The Strange Theory of Light and Matter*, Princeton University Press, 1985. Se trata de un exigente, aunque bello y honesto tratamiento de los principios de la electrodinámica cuántica que no requiere conocimientos matemáticos previos.

Como breve panorámica de la QED, de fácil comprensión tras el libro de Feynman y que tampoco requiere una formación matemática previa, recomiendo «QED made simple», de F. Wilczek, publicado en *Physics Today* 2000, vol. 53N8, págs. 22-28.

Actualmente estoy elaborando un informe completo, que se titulará simplemente *QED* (Princeton).

Para un repaso conceptual de la teoría cuántica de campos, véase mi artículo «Quantum field theory», publicado en el volumen dedicado al Centenario de la Sociedad Americana de Física de la *Review of Modern Physics*, 1999, vol. 71, págs. S85-S95. El volumen está publicado también como *More Things in Heaven and Earth — A Celebration of Physics at the Millennium*, Ed: B. Bederson, ed., Nueva York, Springer-Verlag, 1999, y contiene otros profundos artículos que tocan muchos de nuestros temas.

Parte 6

Comprender la información, bit a bit

Las ecuaciones de Shannon

Igor Aleksander

La información es, en la actualidad, un bien de consumo, igual que el metal o el petróleo, y un servicio público similar al agua o la electricidad. Los políticos, los expertos en bolsa, los analistas, los sociólogos y, sin ir más lejos, cualquiera de nosotros hemos afirmado alguna vez que algo era un signo de la «era de la información» en la que vivimos. Aunque casi nadie sabe en realidad en qué consiste esa era de la información, hay signos de ella por todas partes: trituradores de basura llenos de faxes, correos electrónicos emitidos desde un avión y teléfonos móviles hasta en las playas nudistas.

Es opinable el hecho de si ahora tenemos muchas más cosas que decirnos los unos a los otros, ya que no es la información en sí lo que caracteriza esta era de la información. La marca de esta revolucionaria etapa es más bien la extraordinaria posibilidad de conectarnos con otros o de acceder a ordenadores desde virtualmente cualquier parte. Hace cincuenta años sólo disponíamos del teléfono y de las estaciones inalámbricas. Hoy existen ordenadores conectados a una red global (Internet), teléfonos móviles digitales y cables de fibra óptica. Hasta los productos de entretenimiento han cambiado de manera radical.

Hemos pasado de los vinilos de 78 rpm al videodisco digital y de la vieja *Brownie* a las cámaras digitales que hoy abarrotan los estantes de las tiendas.

Todo lo anterior apunta hacia los enormes progresos registrados en las tecnologías asociadas a la transmisión de información. Pero ¿qué es la información? ¿Qué es lo que dificulta su transmisión? ¿Por qué todos esos progresos han requerido masivas inversiones industriales? ¿Por qué el término *digital* (que significa que algo está representado mediante símbolos separados, tales como los números) aparece tan a menudo en los nombres de esas tecnologías?

Mi objetivo no es proporcionar una descripción detallada de cómo funcionan las tecnologías citadas, sino reivindicar a un héroe de la era de la información, un hombre sin cuya aguda intuición no hubieran sido posibles muchas de ellas. Claude Shannon fue a la vez matemático e ingeniero y conjugó ambas disciplinas de una forma que hizo cambiar el mundo para siempre.

El nombre de Shannon está unido a dos ecuaciones que constituyen la base de la teoría de la información. Su formulación intimida un poco:

$$I = -p \log_2 p$$

y

$$C = W \log_2 (1 + S/N)$$

La primera de ellas nos dice que la cantidad de información contenida en un mensaje dado puede ser representada como una magnitud denominada I , cuya unidad de medida es el *bit*. Aunque los bits y el término *digital* aparezcan a menudo en esas descripciones, ambas ecuaciones son continuas —no digitales—, lo cual significa que son aplicables tanto a las tradicionales líneas telefónicas analógicas como a sus versiones modernas digitales. La primera ecuación establece que la cantidad de información I depende del nivel de *sorpresa* que encierra el mensaje. Matemáticamente, el modo de expresar esa sorpresa es la probabilidad p ; cuanto menos probable sea un suceso, más inesperado será y, por lo tanto, más información aportará. Veremos más tarde de dónde viene ese \log_2 . Baste decir por ahora que, sin esa ecuación, el mundo carecería de una unidad de medida fundamental, tan importante como el galón, el litro, el watio o el kilómetro.

La segunda de las ecuaciones de Shannon es un indicador de la *calidad* de un medio de transmisión, como una línea telefónica o el cable de antena de un televisor. Nos dice que C (en bits por segundo), la cantidad de información que puede ser transmitida a través de una línea u otro medio, depende básicamente de dos factores: W , el ancho de banda (la gama de frecuencias que puede pasar a través de él) y S/N , la relación señal/ruido. Tenemos cierta

noción de esto último cuando, en medio de una fiesta, no nos queda más remedio que gritar (o sea, incrementar S , la señal, para superar el ruido, N). También debemos elevar la voz cuando hablamos con alguien con sordera parcial (alguien cuyo W es limitado). Usando una analogía con el consumo de combustible, C (en bits por segundo) es un factor de calidad de la misma manera que los kilómetros por litro lo es para el motor de un vehículo. Se trata de leyes muy generales: son válidas tanto para una simple conexión telefónica que transmite señales de voz convertidas en voltajes eléctricos como para la televisión de alta definición, en la que las imágenes se transforman en cadenas de números.

El pensamiento y la obra de Shannon trascienden las ecuaciones mismas: éstas son meramente los símbolos que resumen una percepción excepcional de la naturaleza y del empleo de la información.

El propio anonimato de Shannon es una evidencia de su éxito. Todo usuario habitual de Internet se limita a sentarse ante el ordenador, encenderlo y esperar a que el texto y las imágenes aparezcan en pantalla. Supongamos que mi amiga Jill ha prometido enviarme su último autorretrato digital. *Pinchará* sobre la foto en su ordenador, la *adjuntará* a un mensaje de correo electrónico y, finalmente, pulsará (de manera virtual) el botón de *enviar*. Ahora bien, ¿cómo están interconectados nuestros ordenadores? Si es a través de la línea telefónica convencional, tendría que esperar más de treinta minutos para que la imagen de Jill apareciera en la pantalla.

Aunque las demoras de Internet se nos hagan interminables, el hecho de que una transmisión así dure tan sólo minuto y medio, como mucho, se debe a los descubrimientos de Shannon. Incluso el que hayamos estimado *a priori* esos treinta minutos a partir de nuestros datos sobre las líneas telefónicas, lo debemos a que Shannon nos enseñó a evaluar tanto la imagen como la línea con el fin de hacer pasar la una a través de la otra de la mejor manera posible. Han sido los desarrollos realizados a partir de esos cálculos los que permiten hoy que intercambiemos a través de la red las imágenes de nuestros seres queridos.

Internet es un sistema de interconexiones entre cientos de millones de ordenadores esparcidos por todo el mundo. Suministra información a nuestros ordenadores de una forma similar a como el grifo llena de agua la bañera. Al igual que compramos tuberías de media o un tercio de pulgada para nuestro cuarto de baño, adquirimos el enlace a Internet más adecuado para nuestro ordenador. Nuestros cálculos basados en las ideas de Shannon nos dicen que las líneas telefónicas convencionales no tienen suficiente capacidad para manejar la gran cantidad de información contenida, por ejemplo, en una fotografía. Dejando para más adelante la cuestión del bit y en qué consiste, esa capacidad se mide en bits por segundo; cuanto mayor sea el valor, más deprisa se llenará de información nuestro ordenador. La foto de Jill contiene, digamos,

veinte millones de esos bits y una línea telefónica tiene una capacidad de diez mil bits por segundo.⁹³

$$200.000/10.000 \text{ segundos} = 2.000 \text{ segundos} = 33,3 \text{ minutos}$$

Y, a pesar de todo, lo que conecta nuestro ordenador a Internet no es sino una línea telefónica. Imaginemos un mundo en el que no pudiésemos medir los litros de agua que consumimos o los kilowatios de electricidad suministrados por la compañía local. Ése hubiera sido el mundo de la información sin la aportación de Shannon: un mundo sin bits. Shannon nos proporcionó la unidad de medida de la información, pero también puso los cimientos de la teoría de la información misma, algo que todo aquel que diseña una red de comunicaciones necesita conocer.

Shannon nació en Gay lord, Michigan, en 1916. Hijo de un hombre de negocios y una maestra de escuela, pronto demostró una notable aptitud para la ingeniería y las matemáticas. Como a muchos jóvenes de su tiempo, le encantaba cacharrear con la radio, la tecnología más novedosa en la época. Llegó incluso a ganar algún dinero reparando aparatos de radio para una tienda local.

Cuando tenía dieciséis años, ingresó en la Universidad de Michigan para estudiar matemáticas e ingeniería. Cuatro años más tarde era

⁹³ Por razones que no vamos a analizar aquí, el ancho de banda de un canal puede ser especificado en hertzios (Hz) (es decir, en número de ciclos de onda por segundo) o en bits por segundo. Cuando se usa esta última unidad, el valor se duplica, ya que un ciclo de onda (una semionda arriba más una semionda abajo) equivale a dos bits. De este modo, una línea que transmita un máximo de diez mil bits por segundo transmitirá también un máximo de 5.000 Hz.

investigador ayudante en el Instituto Tecnológico de Massachusetts (MIT) en Cambridge, donde trabajó en los primeros proyectos de ordenadores con el carismático Vannevar Bush, un personaje que se convertiría en el asesor científico del presidente Roosevelt y, según muchos, en uno de los fundadores de Internet.

Cuando Shannon llegó al MIT, poco antes de estallar la segunda guerra mundial, los ordenadores apenas existían todavía. El término *ordenador* se empleaba raramente; las máquinas de cálculo eran dispositivos en su mayor parte mecánicos, llenos de engranajes, muelles y similares. Algunos laboratorios trataban de implementar máquinas electrónicas o dispositivos mixtos electromecánicos. Vannevar Bush era uno de los pocos que confiaban en materializar el sueño de Charles Babbage, enunciado un siglo antes, de una máquina que liberara al ser humano del penoso trabajo de hacer cálculos repetitivos, algo que en aquel entonces había sido considerado una fantasía. Sin embargo, tanto Bush como, después, John von Neumann —el *abuelo* del concepto actual de ordenador y una de las mentes más preclaras de Princeton, según Einstein— eran estrategias muy apreciados por las agencias gubernamentales de Estados Unidos, por lo que lograron convencer al gobierno de la importancia de la computación mecanizada y obtener fondos para los primeros desarrollos. Sin Bush y sin Von Neumann, los ordenadores electrónicos no hubieran alcanzado el grado de avance que presentan hoy.

En lamentable contraste, cuando el pionero británico de los ordenadores Maurice Wilkes trató, a finales de los años cuarenta, de obtener financiación en el Reino Unido para construir una máquina computadora, obtuvo una desabrida respuesta de lo que entonces era el Departamento de Investigación Industrial y Científica. Los burócratas llegaron a sugerir que, si Wilkes y sus colegas se sentaban con unas cuantas calculadoras mecánicas, podrían resolver todos los problemas computacionales del mundo sin necesidad de construir una fantástica máquina computadora.

Así pues, Shannon tuvo la suerte de estar al lado de un gran visionario, algo que contribuyó, sin duda, a su valentía a la hora de enfrentarse a ambiciosos retos de ingeniería. Pero el MIT era ya entonces, igual que ahora, un lugar caro incluso para quien dispusiera de una beca que cubriese la matrícula. Vannevar había inventado un dispositivo de cálculo denominado «anализador diferencial» (AD). La máquina almacenaba números en cilindros dentados giratorios, de un aspecto parecido al de los tambores de un cuentakilómetros mecánico. Para ayudar al joven Shannon a ganar algunos dólares, Bush le ofreció trabajar a tiempo parcial en el AD, cosa que el primero aceptó gustosamente. El AD era el sueño de un experimentalista: una enorme colección de cilindros giratorios, engranajes e interruptores eléctricos. Su principal cometido era encontrar soluciones de ecuaciones matemáticas, las cuales se introducían interconectando apropiadamente las partes del dispositivo. La respuesta aparecía en el mecanismo tipo

cuental kilómetros. Podía llevar días enteros ajustar la máquina para trabajar en un único problema y había que desmantelarla y repetir el proceso para abordar el siguiente. De este modo, Shannon se convirtió en uno de los primeros programadores de la historia, encargado de preparar el AD para satisfacer las necesidades de diversos científicos.

Fueron años de formación para el joven Shannon, en los que se haría consciente de la necesidad de entender las dos principales facetas de la información. En primer lugar, los cálculos del AD generaban una *cantidad* de información y, en segundo lugar, en los indicadores de salida existía una *velocidad límite* a la hora de presentar la información calculada. Cantidad de información y velocidad de transmisión iban a ser los pilares de la teoría de la información de Shannon y las dos magnitudes en juego en sus famosas ecuaciones.

Otro importante factor que ejerció influencia en Shannon fue su fascinación por los conmutadores eléctricos y los complejos sistemas para encaminar la electricidad que podían ser contruidos con sólo un puñado de dichos conmutadores (pensemos simplemente en las combinaciones en el encendido y apagado de una luz a que dan lugar dos conmutadores situados en una habitación). Había estudiado las leyes de la lógica establecidas un siglo antes por el británico George Boole, quien en el Queen's College de Cork las había presentado como «las leyes del pensamiento». Por ejemplo, si afirmamos «Alvin y Bob no estaban

juntos en la fiesta», es lo mismo que decir «O Alvin no estaba en la fiesta o Bob no estaba en la fiesta». Boole propuso una notación (conocida hoy como álgebra de Boole) en la que las afirmaciones de arriba pueden ser convertidas en una regla que siempre es cierta:

$$\text{No (A y B)} = (\text{No A}) \text{ o } (\text{No B})$$

donde A y B son proposiciones que pueden ser verdaderas o falsas. El álgebra booleana posee muchas reglas de este tipo.

Hemos mencionado lo anterior porque los conmutadores constituyen la base tanto del direccionamiento como del almacenamiento de la información, y porque para Shannon fue el punto de partida de un atrevido salto conceptual. Un conmutador cerrado es como una proposición *verdadera* en lógica y un conmutador abierto, como una proposición *falsa*. De este modo, si A y B fuesen conmutadores en vez de proposiciones, el álgebra booleana sería aplicable al modo en que se puede conmutar una red para interconectar comunicadores y los conmutadores podrían ser organizados de forma que almacenasen mensajes. De hecho, hoy día la conmutación masiva que tiene lugar en el interior de un ordenador se diseña o analiza mediante el álgebra de Boole. Antes de concluir su primer año en el MIT, Shannon había escrito su tesis magistral sobre la aplicación del álgebra booleana a los circuitos de conmutación, un artículo publicado en 1938 y titulado «Análisis simbólico de los circuitos de retransmisión y conmutación». El

artículo se convertiría en un clásico de la literatura sobre ordenadores y el álgebra booleana, en materia regular de estudio del primer año de ingeniería como método estándar de diseño de circuitos, tanto en ordenadores como en sistemas de telecomunicación. Toda una hazaña para un estudiante de veintidós años, al que impulsaba su deseo de conjugar la naturaleza de la conmutación con la naturaleza de la información para comprender los límites en la velocidad de transmisión de esta última entre dos puntos geográficos.

Shannon dejó el MIT en 1940 con un título de ingeniero y otro de doctor en matemáticas bajo el brazo. En la actualidad, el MIT festeja el hecho mediante un *Shannon Day* anual, en el que se analizan los últimos avances en telecomunicaciones. Tras pasar un año en el prestigioso Instituto de Estudios Avanzados de Princeton, Shannon ingresó en la más importante entidad de investigación industrial de Estados Unidos: los Laboratorios de la Bell Telephone en Murray Hill, Nueva Jersey: los famosos *Bell Labs*. Allí, animado por sus colegas, publicó en 1948 sus informes internos sobre una teoría estadística de la comunicación. Se trataba de la celebrada «Teoría matemática de la comunicación». A modo de ejemplo de la lógica que condujo a Shannon a cuantificar la comunicación, en 1950 escribió el primer programa jugador de ajedrez de la historia, el cual incorporaba un astuto método para reducir el número de posiciones que la máquina debía analizar para encontrar la jugada óptima. Este *algoritmo* fue utilizado en la programación del *Deep Blue* de

IBM que batió al gran maestro Gary Kasparov en 1997 (la primera vez que un campeón mundial era vencido por una máquina).

Hacia 1957, Shannon era considerado uno de los más importantes científicos estadounidenses. La revista *Time* lo incluía entre los nueve líderes de la ciencia estadounidense en un artículo especial, publicado seis semanas después de que la Unión Soviética lanzara con éxito el primer satélite artificial, el *Sputnik*, hecho que había llevado a temer que Estados Unidos quedara detrás de sus rivales en la guerra fría. Por ese artículo sabemos que Shannon era aficionado al jazz y amante de la ciencia ficción y que «como muchos científicos, prefería trabajar por la noche, con el café y los cigarrillos a mano».

En las películas sobre la segunda guerra mundial, como *The Cruel Sea* (*El mar cruel*), el radiotelegrafista acciona el manipulador Morse —di-di-di, da-da-da, di-di-di— para enviar al mundo a través del radiotransmisor un SOS, la señal de que el barco tiene problemas y puede que sean los últimos minutos que permanezca a flote. Pero ¿por qué el operador no toma un micrófono y, de viva voz y usando el mismo radiotransmisor, le explica a quien pueda oírle lo que está pasando? La respuesta es que ese simple patrón de *dis* y *das* tiene muchas más probabilidades de atravesar la crepitación y el siseo del radiotransmisor que la palabra hablada, cuyos variadísimos y sutiles tonos se perderían entre lo que los ingenieros denominan ruido electrónico. Es más fácil ver a lo lejos una débil luz que apreciar a la misma distancia los detalles de una fotografía. Una

linterna que se enciende y se apaga siguiendo el código adecuado puede ser más expresiva que una fotografía con pobre visibilidad. Pero estas vagas nociones necesitan una teoría, y ésta fue la teoría en la que Shannon trabajó durante sus primeros años en los *Bell Labs*.

Todos los sistemas de comunicación son susceptibles al ruido: suena como una especie de chisporroteo en un teléfono y se ve como si fuera nieve en la pantalla del televisor. El ruido distorsiona de manera impredecible la información que el emisor trata de hacer llegar al receptor. Puede llegar a hacer incluso que la información recibida sea ininteligible y, por lo tanto, inútil. Existe otra limitación, algo que los especialistas denominan ancho de banda. La mayoría de los aficionados a la alta fidelidad conocen el término. Suelen preguntar en primer lugar cuál es la respuesta en *graves* — es decir, cuál es la frecuencia más baja (el profundo sonido de la tuba) que el equipo es capaz de reproducir— y, a continuación, cuál es la respuesta en *agudos* o altas frecuencias (las notas más agudas del violín). Restando el límite inferior (digamos, 25 ciclos por segundo, o sea, 25 hertzios) del superior (pongamos, 5.000 hertzios) obtenemos el *ancho de banda* del equipo (4.975 hertzios en nuestro ejemplo). En otras palabras, un mal equipo con poco ancho de banda no permitirá que el oyente disfrute en su totalidad de los matices de una orquesta sinfónica. En términos técnicos, al igual que el ruido, el ancho de banda hace que la información recibida sea, de alguna manera, *menos* que la información transmitida. Toda

comunicación entre un transmisor y un receptor viene caracterizada por cierto valor de ancho de banda, y Shannon deseaba predecir con exactitud y mediante cálculo la magnitud de esa pérdida de información.

Shannon resumió la situación de un modo que se convirtió en la base de la propia teoría de la información. En su planteamiento, todo enlace entre la fuente de información y el destino de sus mensajes consta de cinco elementos fundamentales. En primer lugar, está la fuente en sí. En el caso de alguien que desea transmitir una imagen digital a través de Internet, la fuente es un ordenador en el que esta imagen se halla almacenada en forma de veinte millones de estados 0 o 1 (bits) de una memoria.

El segundo elemento es un codificador, entendiendo como tal el conjunto de dispositivos que preparan la imagen para que sea transmitida en un tiempo razonable a través, por ejemplo, de una línea telefónica de bajo ancho de banda. Como primer paso en la codificación, un moderno ordenador dispone de un programa que *comprime* la imagen. Esta reside en la memoria de la máquina como una secuencia de números, cada uno de los cuales representa el color y la luminosidad de un punto de esa imagen. La compresión elimina las redundancias de esa secuencia de números.⁹⁴ La

⁹⁴ Una de las muchas formas de lograrlo es detectar grandes áreas que tengan el mismo color y, en lugar de transmitir una y otra vez la misma señal, enviar solamente dos números: el color e intensidad del primer punto del área y la cantidad de puntos consecutivos que tienen el mismo valor. Supongamos que en mil puntos consecutivos de la imagen el color vale 144 (sobre un máximo de 256, por ejemplo) y la intensidad, 186 (también sobre 256). Anticipando el significado de *bit* como un dígito tipo encendido/apagado y asumiendo que un conjunto de ocho bits puede adoptar 256 valores (combinaciones) distintos, en lugar de transmitir los dieciséis

siguiente fase de la codificación consiste en transformar los números que representan la imagen en *tonos* que puedan ser transmitidos por la línea telefónica. Este proceso recibe el nombre de *modulación* y es necesario debido a que el canal telefónico está diseñado para transportar señales audibles, es decir, la voz humana. La mayoría de los teléfonos disponen hoy del «marcado por tonos», en el que al pulsar cada tecla se emite un tono identificativo del número. Se trata de un ejemplo de modulación.

El tercer elemento de un sistema de comunicación es el propio cable telefónico, con su ruido y su ancho de banda restringido.

El cuarto es un decodificador, que devuelve la información recibida a un estado lo más próximo posible al de la transmitida. En el caso de una imagen, el decodificador convierte los tonos otra vez en números y reinterpreta éstos de forma que el quinto elemento del sistema, el receptor —la pantalla del ordenador en nuestro ejemplo—, pueda reconstruir la imagen. Todo aquel que haya comprado un módem para su ordenador, ha adquirido en realidad un dispositivo que contiene tanto un codificador (*modulador*) como un decodificador (*demodulador*).

Aunque lo anterior se trate de una tosca imagen de un sistema de comunicaciones, merece la pena intentar construir un boceto de

bits mil veces, el ordenador se limita a enviar esos dieciséis bits seguidos de un número binario que representa las repeticiones. Como con diez bits se pueden codificar 1.024 estados, el ordenador transmitirá sólo veintiséis bits ($16 + 10$), en vez de dieciséis mil. El llamado ratio de compresión resulta, en este caso, de dieciséis mil a veintiséis (o sea, de 615 a 1). Por supuesto, si la imagen fuese muy variada no seríamos tan afortunados. En la práctica, lo habitual es obtener ratios en torno a veinte a uno.

teoría que lo explique y analizar los efectos de las ecuaciones resultantes.

En primer lugar, necesitamos una unidad de medida.⁹⁵ Ya sabemos que la información se mide en bits, cada uno de los cuales puede tener dos valores: 0 y 1. El bit (acrónimo del término inglés «*binary unit*», unidad binaria) fue una de las propuestas clave de Shannon. Volvamos atrás y veamos el porqué de su importancia.

Supongamos que Jill quiere transmitir su autorretrato y, además, una foto de su padre, otra de su madre y otra de cada uno de sus dos hermanos, más una foto del perro, otra del gato y otra de la casa familiar: ocho imágenes en total. El novio de Jill las ha almacenado una única vez en su ordenador. Tras numerarlas del 1 al 8, si ella desea hacerle ver una de esas imágenes, lo único que necesita es transmitirle el número apropiado. El ordenador se limita a presentar en pantalla la imagen sin que la foto en sí tenga que ser transmitida. Ahora bien, un bit puede representar dos números: 0 y 1. Con dos bits, obtenemos cuatro (00, 01, 10 y 11). Tres bits representan ocho (000, 001, 010, 011, 100, 101, 110 y 111). He aquí uno de los principales hallazgos de Shannon: la información es proporcional a lo que desconocemos. Cuando el novio de Jill no tenía información sobre esas fotos, necesitaba veinte millones de bits para describir cada una de ellas. Una vez almacenadas, basta con tres bits para referirse a cualquiera. Éste es el modo en el que la probabilidad *se cuele* en la primera de las ecuaciones de Shannon:

⁹⁵ En esta sección y en la siguiente no seremos demasiado estrictos con el rigor matemático.

a priori, y suponiendo que nunca antes hubiera visto a su novia, la probabilidad de imaginar su aspecto físico con exactitud sería bajísima. Cuanto más inesperado sea un suceso, más información aporta su ocurrencia. La primera ecuación de Shannon relaciona la información con el «logaritmo en base 2» (\log_2) de la probabilidad. Esta aparentemente extraña jerga no es, sin embargo, difícil de comprender. Tomando algunos ejemplos sencillos, el logaritmo en base 2 de 8 ($2 \times 2 \times 2$, o sea, 2^3) vale 3; el de 16 ($2 \times 2 \times 2 \times 2$, o sea, 2^4) vale 4; el de 32 ($2 \times 2 \times 2 \times 2 \times 2$, o sea, 2^5) vale 5. Es decir, el logaritmo en base 2 de un número es, simplemente, el exponente al que hay que elevar 2 para obtener dicho número. Si acudimos de nuevo a la ecuación:

$$I = -p \log_2 p \text{ (medida en bits, por las razones que veremos)}$$

observamos que dice: «*La cantidad de información involucrada en el conocimiento de un suceso depende de la probabilidad p de que este suceso ocurra*». El concepto nos conduce inexorablemente a la definición del bit. Si lanzamos una moneda al aire, el resultado será uno de estos dos sucesos: cara o cruz. Cada suceso tiene una probabilidad de que ocurra; en este caso y si la moneda no está trucada, la probabilidad de cualquiera de ellos es de la mitad. Para obtener la información total asociada al suceso de arrojar la moneda al aire, sumamos el contenido en información de cada una de las posibilidades:

$$I = [-(1/2) \log_2 (1/2)] + [-(1/2) \log_2 (1/2)]$$

El resultado es exactamente 1. No es casualidad y para un matemático está claro por qué Shannon utilizó el \log_2 en la ecuación. La unidad de información, el bit, está asociada a un interruptor encendido o apagado, a un número 0 o 1, a una moneda que resulta cara o cruz; el \log_2 asegura (como veremos) que *cualquier* otra cantidad de información puede ser medida en bits. La fórmula cubre también el caso de certidumbre. Si la ocurrencia de un suceso o su no ocurrencia son ciertas, la ecuación nos dice que el suceso proporciona 0 bits de información. Dos bits son como dos monedas, que dan lugar a cuatro mensajes; por ello, el contenido en información de cuatro mensajes equiprobables es dos bits. La ecuación puede ser aplicada a cualquier número de mensajes. Por ejemplo, la transmisión de todas las letras mayúsculas del alfabeto inglés representa veintiséis mensajes, lo que requeriría cinco bits (ya que $2^5 = 32$, es decir, algo más de 26). La primera ecuación de Shannon nos permite, pues, medir inequívocamente cuánta información está contenida en algo que tratamos de comunicar a alguien.

Las palabras *crepitación* y *siseo* han aparecido ya en el presente ensayo. Allí donde se utilice la electricidad o un medio inalámbrico para transmitir información, la naturaleza hace que el *ruido* sea un compañero inevitable. Si interconectamos dos ordenadores

mediante un cable telefónico, al receptor le llegará un cierto nivel de energía eléctrica que no ha sido emitido por el transmisor. La información que nos interesa es enviada a través del cable codificada en forma de una secuencia de voltajes eléctricos, cada uno de los cuales representa, por ejemplo, un punto de imagen de la foto de Jill.⁹⁶ Pero los electrones del cable tienen la costumbre de andar danzando por ahí. Esa actividad aleatoria modifica el voltaje transmitido, de modo que al receptor puede llegarle un valor modificado aleatoriamente. El fenómeno no sólo se da en la transmisión por cable; el espacio libre utilizado para la transmisión vía radio está plagado de partículas cargadas en movimiento que inducen variaciones significativas en las señales transmitidas. Esas variaciones aparecerán en la imagen recibida como puntos anómalos, en forma de *nieve*, o como un siseo alrededor de los mensajes en código Morse que transmitía el operador de la nave siniestrada de la película antes mencionada.

Aquí es donde entra en escena la segunda ecuación. Las imperfecciones de un canal en cuanto a limitación de frecuencias (ancho de banda W) y ruido (N) afectan a la capacidad C de un medio de transmisión para transmitir una señal de intensidad S , de esta manera:

$$C = W \log_2 (1 + S/N) \text{ en bits por segundo}$$

⁹⁶ El voltaje o nivel de tensión mide la energía transferida a cada carga eléctrica. El codificador convierte los números que representan el color y la intensidad de los puntos de imagen en niveles de tensión que son transmitidos al otro extremo del cable.

Para entender la fórmula usaremos de nuevo el ejemplo de la imagen transmitida a través de Internet. Haremos, en primer lugar, el supuesto aproximado de que el ancho de banda W es una frecuencia *máxima* de transmisión y de que el límite inferior del ancho de banda es cero. En segunda aproximación, supondremos que el concepto equivale al «máximo número de paquetes de bits que podemos transmitir por segundo». Un paquete de bits representa una gama de números (3 bits representan 8 números, 4 bits representan 16, y así sucesivamente). Ahora bien, esta gama de números depende de cuánto ruido haya en el sistema. El término $(1 + S/N)$ nos dice cuán a menudo el ruido hará cambiar el número del paquete. Si no existe ruido, N es 0 y $(1 + S/N)$ vale infinito ($+\infty$), lo que significa que los paquetes pueden ser tan grandes como queramos. En este caso, la imagen entera podría constituir un único paquete que podríamos transmitir W veces por segundo. Con valores de W del orden de diez mil incluso para líneas de baja calidad, la comunicación sería prodigiosamente rápida en una línea sin ruido. Por desgracia, el ruido siempre está presente y si, por ejemplo, tuviera un nivel que fuera una séptima parte del de la señal, el término $(1 + 7)$ nos dice que, de cada ocho números transmitidos, habría uno afectado por el ruido. En este caso, $\log_2(8) = 3$, lo que significa que ahora sólo podremos transmitir tres bits por paquete, W veces por segundo. Con $W = 10.000$, incluso la

versión comprimida de la imagen —digamos un millón de bits— llevaría transmitirla:

$$1.000.000/(3 \times 10.000) \text{ segundos} = 33,3 \text{ segundos}$$

(La versión sin comprimir tardaría alrededor de 10 min.).

Si el nivel del ruido igualara al de la señal, $(1 + S/M)$ valdría 2 y sólo se podría enviar un bit por paquete, con lo que la transmisión comprimida tardaría en torno a un minuto y cuarenta segundos. A medida que el ruido se hace mayor, $(1 + S/M)$ tiende a valer 1, lo que significa que no se puede transmitir ningún bit por paquete, ya que $\log_2(1) = 0$.

Para el radiotelegrafista del barco que se hunde, es el alto nivel de ruido lo que no le permite transmitir la voz (que requiere unos ocho mil bits por segundo), pero sí los tres o cuatro bits por segundo de los *dis* y *das* que, mediante el código Morse, sintetizan lo esencial del mensaje.

El cable telefónico ordinario, con sus diez mil bits por segundo, no es el único medio utilizado para la transmisión de información. Hay todo tipo de cables y otros medios con anchos de banda muy superiores. El *coaxial*, por ejemplo, consta de un hilo central, rodeado de un aislante plástico y enfundado el conjunto en una malla metálica. El ancho de banda de este cable puede alcanzar los doscientos millones de hertzios (es decir, 200 megahertzios, lo cual equivale a cuatrocientos millones de bits por segundo). Obviamente,

permite una comunicación mucho más rápida, aunque también es más caro. Para anchos de banda aún mayores se utiliza la fibra óptica, que en vez de transmitir impulsos eléctricos transmite impulsos de luz (generados habitualmente por láser). Las emisoras de radio de frecuencia modulada (FM) trabajan en el entorno de los 100 megahertzios, lo que significa que el espacio libre, a través del cual la información viaja en forma de ondas electromagnéticas, presenta un ancho de banda muy elevado.

Todo eso está muy bien, pero hasta la más exquisita transmisión de música clásica requiere solamente un ancho de banda de 30.000 hertzios. ¿Para qué sirven esos anchos de banda tan gigantescos? Regresemos al concepto de codificador o modulador de Shannon. Para entenderlo mejor, usaremos de nuevo el ejemplo de la imagen digital. Anteriormente señalamos (v. nota 2) que para representar el color de cada punto de imagen se requieren 256 valores, es decir, 8 bits (simplemente, porque $\log_2(256) = 8$). Supongamos que disponemos de una combinación ancho de banda/ruido muy favorable y que deseamos transmitir ocho veces ese número en el mismo tiempo (transmitir ocho imágenes simultáneamente). ¿Es posible hacer uso del ancho de banda como si fueran ocho canales separados en lugar de uno solo? La solución es sencilla. A todo lo que vaya hacia el canal 1 se le asigna el número 1 como prefijo. La información destinada al canal 2, recibe un 2, y así sucesivamente. De este modo, en cada periodo transmitimos un grupo de ocho números, precedido cada uno del identificador del canal. En el

extremo receptor, el decodificador debe ser capaz de detectar esos prefijos y separar los puntos de imagen. Los prefijos son, pues, los *portadores* de la información relativa a un canal.

Sucede algo similar cuando sintonizamos una emisora de radio. Lo que captamos es la *portadora* de un canal concreto y el aparato decodifica el contenido de ese canal. De esta forma, el ancho de banda del espacio libre —pongamos, trescientos millones de hertzios— puede alojar diez mil emisoras de radio, o incluso más (ya que no todas requieren esos 30.000 hertzios de ancho de banda).

Esta clase de codificación adopta una curiosa forma en Internet. Los identificadores son del tipo de *jack@toc.ac.uk*, lo cual podría ser la dirección de correo electrónico de Jack a la que Jill envía sus fotos, de modo que sea Jack y sólo Jack quien las reciba. En el caso del correo electrónico, esta forma de codificar hace que el mensaje de Jill y su identificador de canal deambulen a través de las inmensidades de la red buscando el ordenador de Jack, cuya dirección es la destinataria del mensaje. El ordenador de destino decodifica entonces las imágenes de Jill y las presenta en pantalla. El proceso establece un canal único entre Jack y Jill a través de la jungla de cables, enlaces vía satélite y transmisiones de radio que componen Internet.

Hay que subrayar que la *compresión* constituye una parte importante del proceso de codificación y decodificación contemplado por Shannon en su esquema de cinco etapas: «fuente-codificador-canal-decodificador-destino». A quienes utilizan Internet y

descargan fotografías o películas les resultan familiares los formatos jpeg (para imágenes fijas) y mpeg (para imágenes en movimiento). Se trata de protocolos de codificación y decodificación que evitan a los usuarios muchas horas de tiempo de descarga. Miremos hacia donde miremos en el amplio mundo actual de las telecomunicaciones, constataremos que el modelo de Shannon sobre la naturaleza de la información es de gran ayuda en el diseño de sistemas capaces de proporcionar una comunicación de alta velocidad.

Un consecuencia inesperada de la definición de bit hecha por Shannon es que ese bit no sólo es la unidad de información transmitida, sino que también se convierte en la unidad de almacenamiento de información o de *memoria*. Un único interruptor puede estar abierto o cerrado, según la definición por la que un bit es portador de sólo dos mensajes. Un interruptor almacena o memoriza, por lo tanto, un bit de información. Dos interruptores pueden adoptar cuatro combinaciones de aperturas y cierres, y así sucesivamente. El término \log_2 de la ecuación de Shannon sirve para calcular el número de interruptores que se requieren para almacenar, por ejemplo, un millón de mensajes: el $\log_2 (1.000.000)$ resulta ser un número relativamente pequeño: alrededor de veinte interruptores. Esta relación es la que hace que los ordenadores tengan esas impresionantes capacidades de memoria. Quien disponga de un ordenador personal razonablemente reciente sabe que existen al menos dos tipos de memoria: el disco duro y la

memoria de acceso aleatorio (*Random Access Memory*; RAM). En la actualidad, es normal que un disco duro almacene al menos cinco gigabytes. Por razones que no vienen al caso, un grupo de ocho bits se denomina byte, por lo que esos cinco gigabytes se convierten en cuarenta mil millones bits. El disco duro es un disco de metal giratorio en el que un bit es almacenado magnetizando un tramo de pista circular por medio de una *cabeza* (la cual se transforma en un imán al ser alimentada por una corriente eléctrica). El tramo de pista está magnetizado o no y, en este sentido, equivale a un interruptor: almacena un bit. El disco giratorio contiene muchas de esas pistas formadas por «interruptores abiertos o cerrados», que pueden ser leídas por la misma cabeza, ya que los tramos magnetizados inducen en ella una corriente. La razón por la que un ordenador incorpora también memoria RAM es que el disco duro es relativamente lento, debido a las inercias de sus partes móviles. Puede llevar hasta una centésima de segundo acceder a un punto concreto del disco de metal. La RAM es mucho más rápida: sus tiempos de acceso se hallan en el orden de las milmillonésimas de segundo. El estado de los diminutos interruptores de silicio que la forman puede ser examinado de manera similar a la que usamos para consultar un archivador. Necesitamos sólo la etiqueta, al igual que *impuestos* o *hipotecas*, que identifica la carpeta deseada. Mediante un simple vistazo, localizamos el documento. De manera parecida, cada interruptor de silicio ostenta una etiqueta denominada *dirección* que, al aplicarla al conjunto de los

interruptores, extrae el contenido del que la posee. La RAM es, pues, más rápida, pero más pequeña que el disco duro.

Usando estas ideas podemos construir ahora un escenario más completo de lo que sucede cuando descargamos una imagen de un ordenador a otro. Si la imagen fue adquirida con una cámara digital, la captó en primer lugar un dispositivo electrónico sensible a la luz, siendo luego almacenada en la propia RAM de la cámara. Transmitida más tarde (mediante el *software* y los cables adecuados) al disco duro del ordenador de Jill, ocupa veinte millones de los cuarenta mil millones de interruptores magnéticos que contiene éste. Si Jill quiere ver el resultado en la pantalla, tendrá que mover la información a la memoria RAM del ordenador. Esta operación corre a cargo de ciertos programas que hacen que la energía eléctrica de los bits se convierta de nuevo en patrones luminosos. Cuando el destinatario, Jack, solicita la imagen para su descarga, se produce la transmisión a través de Internet y la imagen queda almacenada en la RAM de su ordenador, que puede visualizarla en su pantalla. Para guardarla de forma permanente, Jack la transfiere a su disco duro.

La capacidad de almacenamiento de los ordenadores ha crecido en los últimos años de manera prodigiosa. Tomando como referencia el área de almacenamiento que requiere la imagen de nuestro ejemplo y aplicando el \log_2 de Shannon es posible calcular cuántas imágenes diferentes cabría representar en esa área. La respuesta está en la expresión:

$$20.000 = \log_2(x)$$

El valor de x resulta ser más o menos un 10 seguido de 7 millones de ceros, una cifra verdaderamente astronómica.

El invento de Shannon, el bit, no sólo nos ha permitido cuantificar la información, sino que se ha convertido en la moneda de cambio de la computación. A pesar de los inmensos avances tecnológicos registrados en este campo, los conceptos y formulaciones de Shannon permanecen inalterados. La enorme versatilidad de los ordenadores y la inconcebible potencia de esos millones de ordenadores conectados a través de Internet han dado lugar a entes cuya complejidad empieza a estar más allá de lo imaginable. Es ese \log_2 lo que está detrás.

¿Por qué todo se está volviendo digital? Existen hoy teléfonos digitales donde antes sólo había teléfonos ordinarios (analógicos), han surgido la radio y la televisión digitales y preferimos la música grabada de manera digital en los discos compactos a los viejos registros sobre vinilo. Todos los bienes de consumo relacionados con las comunicaciones se están convirtiendo en digitales. Los propios gobiernos apoyan esta tendencia, aunque en muchos casos no sepan dar razones convincentes para ello. Shannon presagiaba esta ola de cambios al establecer la capacidad de un canal en la segunda de sus ecuaciones y cuando definía una estructura

estándar codificador-canal-decodificador para cualquier sistema electrónico de comunicaciones.

Digital significa simplemente que los datos son transmitidos como símbolos discretos. Para aclarar la noción de *digital* tal vez ayude considerar algo que no lo sea. La voz humana, el método de comunicación por excelencia, no lo es. Cuando hablamos, generamos ondas de presión en el aire mediante el movimiento de nuestras cuerdas vocales, la forma de la cavidad que hay tras nuestra boca y la configuración de los labios y la lengua. Esas ondas llegan al tímpano de nuestro interlocutor y hacen que su cóclea (el pequeño órgano en forma de caracol del oído) convierta la presión en señales nerviosas que, transmitidas al cerebro, se traducen en la sensación subjetiva de *audición*. Pero desde el momento en que la comunicación está asistida por la electrónica, aparece la opción de transformar esas ondas en secuencias de números. El sistema pasa a ser digital: ya no se transmiten ondas, sino la información que contienen, codificada en forma de bits.

Aunque la teoría de Shannon es aplicable tanto a sistemas digitales como analógicos, está implícito en ella el que los primeros son más eficientes y el que los dígitos binarios son la forma óptima de transmitir información. El razonamiento está basado en el coste y puede ser esquematizado de la manera siguiente. Supongamos que la representación de un punto de imagen requiere doscientos cincuenta y seis valores. Podemos imaginar aquí el canal como una especie de caja que ha de tener el tamaño adecuado para alojar

todos los números comprendidos entre 0 y 255 en la forma, por ejemplo, de pequeños cubos. Definimos entonces el coste del canal no como el coste de todos esos cubos, sino el de la caja necesaria para transportarlos. Es razonable pensar que cuantos más cubos (números) tenga que alojar la caja, más alto será su coste. Nuestro canal costaría doscientas cincuenta y seis unidades de cierta moneda.

¿Cuál sería el coste si utilizásemos dos canales más pequeños para transportar la misma información? Las cajas sólo tendrían que transportar dieciséis cubos, ya que al tomar un número de cada caja obtendríamos 16×16 combinaciones, o sea doscientos cincuenta y seis números. Sin embargo, el coste total de esos dos canales sería de $16 + 16 = 32$ unidades. Si seguimos avanzando en la misma dirección, llegaríamos hasta el punto en el que un canal aloja sólo dos números; con ocho de esos canales obtendríamos de nuevo $2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 = 256$ y el coste de la configuración sería tan sólo $2 + 2 + 2 + 2 + 2 + 2 + 2 + 2 = 16$ unidades. Según Norbert Wiener (el *abuelo de la cibernética*), el hecho demuestra la gran perspicacia de Shannon al definir el bit. El sistema de coste mínimo está constituido por canales binarios que transportan un solo bit, el modo más económico de transmitir información.

Es justamente la eficiencia económica de la codificación binaria lo que está haciendo que el mundo se transforme en digital. Y es, una vez más, el \log_2 de Shannon —que aparece en ambas ecuaciones—

lo que está detrás, ya que en toda aplicación que emplee ondas analógicas (con amplitudes que representan números enteros de manera precisa), la teoría de Shannon dice que sería mucho más barato usar bits para transmitir la misma información. Probablemente, el ejemplo más claro de ello es la rápida evolución que ha tenido lugar en la música grabada desde los discos de vinilo, que registraban el sonido en surcos (y ocupaban una gran superficie para almacenar apenas veinte minutos de música), hasta los modernos discos versátiles digitales (*Digital Versatile Disc*, DVD), pasando por los discos compactos. Mediante técnicas digitales, un DVD puede almacenar hasta cuatro horas de música en una vigésimo quinta parte del espacio que ocupaba un vinilo. Pero la idea es aplicable a muchas otras cosas. Los teléfonos móviles e inalámbricos funcionan mucho mejor en su versión digital. Y todo ello se debe al \log_2 .

Así pues, estas dos ecuaciones:

$$I = -p \log_2 p$$

y

$$C = W \log_2(1 + S/N)$$

han transformado el mundo de las comunicaciones. Y a pesar de su imponente aspecto, su verdadero poder reside en la relación que subyace en ambas:

Información en bits = \log_2 (lo que se quiere comunicar).

El modelo de Shannon (fuente-codificador-canal-decodificador-destino) es válido tanto si enviamos imágenes digitales mediante la última tecnología de Internet, como si charlamos a través de nuestro teléfono móvil o si intentamos que el camarero de un bar musical entienda nuestra demanda. La primera ecuación nace en el seno de este modelo: se trata de una definición muy general de la información, basada en la sorpresa y en la probabilidad. Pero el mensaje importante de esta ecuación consiste en que, si la probabilidad de un suceso es del cincuenta por ciento, éste contiene exactamente un bit de información. Generalizando esta idea, toda transacción real puede ser convertida en una cadena de bits del tamaño apropiado.

La segunda ecuación se concentra en la naturaleza del canal: la línea telefónica, el espacio libre o el ambiente ruidoso del bar. Shannon demostró que existe un límite para el número de bits por segundo que pueden ser transmitidos a través de un medio dado, límite que viene determinado por el ancho de banda y el ruido del canal. El modo de aprovechar este límite con la máxima economía es mediante la codificación digital. La clave está en diseñar

codificadores cada vez más perfectos que conviertan la información de partida en cadenas de bits codificadas de forma óptima. Hay un buen número de empresas que trabajan en el tema de la codificación desde hace bastantes años, tanto en el campo de la telefonía móvil como en el de la música y el vídeo de entretenimiento.

Las ideas de Shannon no están restringidas al mundo de las comunicaciones. Disponemos de ecuaciones similares en otros campos de la ciencia, asociadas al concepto de *entropía*, el grado de desorden de un sistema físico. El concepto equivale al *grado de sorpresa* en teoría de la información. En cualquier caso, la formulación de Shannon demuestra que la información responde a las mismas leyes que gobiernan la física, la termodinámica y la química física y que son bien conocidas por los matemáticos. La teoría de la información era un área científica de la que sólo los diseñadores de equipos electrónicos tenían noción —y era una noción muy vaga— antes de 1950. Shannon hizo ver que se trata de una materia tan importante como puedan ser las partículas elementales y que posee unas leyes equivalentes a las que gobiernan a esas partículas. En mi trabajo relativo al modelado de la intrincada arquitectura del cerebro, el lenguaje de la teoría de la información está siempre presente. La capacidad de almacenamiento de las células cerebrales puede ser medida en bits y la anatomía de las interconexiones entre muchas áreas

funcionales del cerebro puede ser estudiada mediante la noción de capacidad de un canal.

El modesto personaje al que debemos estas ideas, Claude Shannon, es uno de los gigantes tecnológicos del siglo XX. Aunque, quizás, el término *tecnología* sea inadecuado aquí, pues realmente Shannon hizo una contribución *intelectual* de primera magnitud al mundo contemporáneo. A Shannon le atraían las cosas complejas. Sus ecuaciones no se refieren a la naturaleza, sino a sistemas diseñados y desarrollados por ingenieros. Son ecuaciones que plasman de forma elegante la complejidad de la información y la problemática de los medios utilizados para almacenarla o transmitirla. La contribución de Shannon consistió en caracterizar esos medios de un modo que fuera útil para los ingenieros. Shannon ocupa un lugar similar al de otros grandes innovadores, como su héroe de juventud Thomas Edison (quien resultó ser pariente lejano de Shannon, para alegría de éste) o Johannes Gutenberg. Al igual que la imprenta, Internet es un monumento al lenguaje, la habilidad más característica del ser humano. Y al igual que el giro de una prensa para hacer vino estimuló la imaginación de Gutenberg, la imaginación de Shannon despertó ante el *click* de los conmutadores de un analizador diferencial.

Tras una brillante carrera académica, Shannon dejó el MIT en 1978, para convertirse en profesor emérito y uno de los más respetados decanos de la ciencia estadounidense. En 1985 recibió el Premio Kyoto, equivalente al Premio Nobel en el campo de los ordenadores.

Tras su jubilación, continuó trabajando en una variada gama de temas, entre ellos una teoría matemática para los juegos malabares, el diseño de un *pogo stick*⁹⁷ motorizado y el desarrollo de un método de juego en bolsa basado en la teoría de las probabilidades. El final de su vida quedó trágicamente ensombrecido por la enfermedad de Alzheimer, la cual le impidió asistir a la inauguración de una estatua en su honor, erigida en su ciudad natal de Gaylord, Michigan, en otoño de 2000.

Claude Shannon falleció el 24 de febrero de 2001 en una clínica de Massachusetts. Su figura fue cortésmente elogiada esos días, pero resulta lamentable que la mayoría de los medios de comunicación — demasiado ocupados en participar en la revolución de la información— hayan ignorado hasta la fecha el trascendente papel de Shannon en la configuración actual del mundo.

En la actualidad, utilizamos el término *intelectual* para referirnos a quien hace alguna contribución notable a la filosofía, la política o las humanidades. No siempre fue así. El ideal platónico o aristotélico incluía también el ingenio y la abstracción matemática. Shannon cambió el mundo siendo un maestro en ambos campos.

Lecturas recomendadas

C. E. Shannon, «A mathematical theory of communication», *Bell System Technical Journal*, 1948, vol. 27, págs. 379-423 y 623-656.

⁹⁷ Saltador. Juguete que consiste en un bastón en forma de «T», dotado de un muelle o un cilindro neumático, con el que se avanza a saltos (*N. del T.*)

S. Román, *Introduction to Coding and Information Theory*, Dortmund, Springer-Verlag, 1996.

G. Boole, *An Investigation of the Laws of Thought* (Londres, Dover Publications, 1995).

C. E. Shannon, «A symbolic analysis of relay and switching circuits», *Transactions of the American Institute of Electrical Engineering*, 1938, vol. 57: págs. 713-732.

N. Wiener, *Cybernetics*, Cambridge, MIT Press, 1948. [Trad. esp.: *Cibernética*, Barcelona, Tusquets Editores, col. Metatemas 8, 1985].

I. Aleksander, *Impossible Minds: My Neurons My Consciousness*, Londres, Imperial College Press, 1996.

—, *How to Build a Mind*, Londres, Weidenfeld and Nicholson, 2000.

Parte 7

Simetrías ocultas

La ecuación de Yang-Mills

Christine Sutton

Verano en Nueva York, húmedo y caluroso, como en las películas. Año 1953. Stalin acaba de morir, Isabel II es la nueva reina de Inglaterra y un joven senador, llamado John Fitzgerald Kennedy, está a punto de contraer matrimonio con Jacqueline Lee Bouvier. Los caminos de dos jóvenes se cruzan en el despacho del Laboratorio Brookhaven de Long Island que comparten. Como una rara alineación de planetas, pasan fugazmente a través de la misma región del espacio y el tiempo. La yuxtaposición da lugar a una ecuación que podría constituir la base del Santo Grial de la física: la «teoría del todo».

Robert Lawrence Mills y Cheng Ning Yang nacieron en dos extremos del mundo, pero siempre les unió su pasión por la física teórica. Yang, que cumplía treinta y un años en septiembre de 1953, había llegado a Estados Unidos procedente de China y obtuvo un doctorado en la Universidad de Chicago antes de ingresar en el Instituto de Estudios Avanzados de Princeton, Nueva Jersey. Mills, de veintiséis, acababa de entrar en el Laboratorio Brookhaven tras estudiar en las universidades de Columbia y Cambridge. En el verano de 1953, Yang pasaba una temporada en Brookhaven y le hicieron hueco en el despacho que ocupaba Mills. Sus trayectorias

divergieron enseguida, pero la ecuación de Yang-Mills hizo que, tras ese breve encuentro, sus nombres fueran inseparables.

En la década de 1950, la ecuación de Yang-Mills parecía el resultado de una interesante idea con poca conexión con la realidad, pero a finales del siglo XX el panorama cambió. La ecuación está detrás de los Premios Nobel de Física de 1979 y 1999, y tiene la suficiente importancia matemática como para haber sido incluida por el Instituto Clay de Matemáticas entre los siete «Problemas del milenio», cuya resolución rigurosa supone para quien la obtenga un premio de un millón de dólares.

¿Por qué todo ese interés? ¿Qué hace que la ecuación de Yang-Mills sea tan importante? ¿En qué consiste dicha ecuación? Para poder contestar estas preguntas, primero debemos echar un vistazo a los conceptos que los físicos usan a diario para interpretar los fenómenos del mundo que nos rodea.

Las fuerzas de la naturaleza

La historia de la ecuación de Yang-Mills comienza en el siglo XVII, cuando, según la leyenda, la caída de una manzana condujo a Newton a formular su ecuación de la gravedad. En la actualidad, ponemos satélites en órbita alrededor de la Tierra y enviamos sondas espaciales a lejanos planetas gracias a trayectorias calculadas según sus métodos. Entre las grandes obras newtonianas, hay que destacar los *Philosophiae naturalis principia mathematica*, publicados en 1687 y conocidos simplemente como

los *Principia*. En esta magna obra, Newton se propuso explicar en términos matemáticos todo lo que conocía del mundo físico, desde las trayectorias de los planetas hasta los ciclos de las mareas. Sus principales herramientas eran ecuaciones que relacionaban movimientos y fuerzas —las mismas ecuaciones que siguen siendo la base de la mecánica y la dinámica que se enseñan en todas las escuelas y universidades del mundo—. No obstante, Newton era consciente de que su obra abarcaba sólo ciertos aspectos del mundo físico. En el prefacio de los *Principia*, escribe:

«Me gustaría que pudiésemos explicar los demás fenómenos de la naturaleza mediante el mismo tipo de razonamiento que el empleado a partir de los principios mecánicos, pues muchas razones me inducen a pensar que todos ellos dependen de ciertas fuerzas, debido a las cuales las partículas de los cuerpos, por causas hasta ahora desconocidas, resultan, o bien atraídas mutuamente para adherirse en figuras regulares, o bien repelidas y alejadas unas de otras; unas fuerzas desconocidas cuya naturaleza, hasta la fecha, han tratado los filósofos de desvelar en vano».

Trescientos años después, el deseo de Newton va camino de cumplirse, a medida que las investigaciones de los modernos filósofos naturales —los físicos— revelan la estructura de esas fuerzas misteriosas. La ecuación de Yang-Mills parece articular

matemáticamente el principio básico que subyace bajo esas fuerzas, tal como Newton deseaba. En cierto sentido, es un equivalente moderno de las ecuaciones newtonianas del movimiento, una fórmula que pone en evidencia la belleza de las interrelaciones naturales, algo cuya potencialidad es tan apreciada hoy como en los tiempos de Newton.

Actualmente, cuando los estudiantes de física dominan la mecánica newtoniana —y son capaces de aplicar sus ecuaciones a bolas de billar que colisionan, cohetes espaciales y similares—, aprenden que la materia que contemplamos a nuestro alrededor y toda la que contiene el universo en general está formada por partículas controladas por fuerzas. Esas fuerzas son conocidas hoy día en su mayor parte, tal como profetizaban las sorprendentes palabras de Newton; dan forma y estructura a ese universo de partículas. Constituyen, en suma, el esqueleto invisible del universo.

Pero ¿qué queremos decir al hablar de fuerzas? Las diversas fuerzas gobiernan las interacciones entre las partículas y las agrupan en estructuras a todas las escalas, desde los diminutos átomos hasta las colosales galaxias. Las fuerzas actúan de manera invisible, unas veces atrayendo las partículas, como cuando la gente se arracima para escuchar a un músico callejero, y otras lanzándolas lejos unas de otras, como el timbre que señala el fin de la jornada escolar. Sin las fuerzas no existiría más que un gas de partículas, sin que éstas tuvieran modo de interactuar unas con otras o de revelar siquiera su existencia.

La más conocida de las fuerzas que han construido el universo a partir de sus componentes fundamentales es la gravedad, la fuerza que dominaron las matemáticas de Newton hace tres siglos. Algo menos conocida es la electromagnética, la fuerza que hay detrás de las muchas facetas de la electricidad y el magnetismo, desde los fenómenos naturales de los rayos y los imanes hasta la moderna brujería de la televisión y la radio. Otras dos fuerzas, las denominadas fuerte y débil, son mucho menos conocidas, pero sus efectos determinan en gran medida por qué la materia es como es.

Las fuerzas débil y fuerte actúan en el seno del núcleo que constituye el corazón de cada átomo en todo tipo de materia. Ambas compiten ahí con la fuerza electromagnética por el control último de ese átomo. En ocasiones, la fuerza fuerte es la vencedora y logra reunir a los constituyentes del núcleo (las partículas denominadas protones y neutrones) para formar un ente estable. Más a menudo, la fuerza débil o la electromagnética son las victoriosas, produciendo una enorme variedad de núcleos inestables y radiactivos, algunos de los cuales —como el de uranio— se dan hoy día de forma natural en la Tierra.

El descubrimiento de las fuerzas fuerte y débil, escondidas hasta ahora en el dominio submicroscópico del átomo, ha permitido que los físicos de comienzos del siglo XXI se planteen deducir todos los fenómenos naturales a partir de las interacciones de fuerzas fundamentales. Por supuesto, los físicos no pueden *deducir* una vaca o la hierba que ésta come a partir de esos principios, pero sí

las propiedades de los materiales. Pueden calcular las interacciones electromagnéticas colectivas de los electrones que revolotean alrededor del átomo en un sólido y utilizar esos cálculos para diseñar nuevos materiales. Pueden usar su conocimiento de las fuerzas débil y fuerte en el núcleo atómico para calcular de qué modo ciertos elementos vitales, tales como el carbono, el oxígeno o el hierro, son fabricados en el corazón de las estrellas. En cualquier caso, lo más atractivo para ellos es el convencimiento de que están cada vez más cerca de una teoría unificada, un conjunto único de ecuaciones interrelacionadas capaz de describir todas las fuerzas, y la ecuación de Yang-Mills es una pieza clave en esta *unificación*.

En la década de 1930, la naturaleza de la materia parecía haber sido reducida a unos pocos componentes fundamentales. Por aquel entonces ya se sabía que los elementos químicos, desde el hidrógeno y el helio hasta el uranio, pasando por el carbono, el oxígeno y el hierro, estaban constituidos cada uno por un tipo distinto de átomo. Sin embargo, esa gran variedad de átomos estaba construida, a su vez, a partir de sólo tres componentes básicos: electrones con carga negativa, protones con carga positiva y neutrones carentes de carga eléctrica. Los protones y los neutrones residían juntos en el núcleo, mientras que los electrones orbitaban alrededor de éste a distancias relativamente grandes, determinando el tamaño del átomo y dando a la materia su forma.

La propia existencia del núcleo atómico es, a primera vista, paradójica, ya que los protones, dotados todos de la misma carga

eléctrica, deberían repelerse entre ellos. Las fuerzas eléctricas presentes en el núcleo deberían desintegrarlo sin más. Como claramente no es así en la materia estable que nos rodea, tiene que haber una fuerza más potente que la fuerza eléctrica. Pero debe actuar sólo a distancias comparables al tamaño del núcleo; en caso contrario, los átomos adyacentes tenderían a aproximarse y la materia sería mucho más densa de lo que es. Esa fuerza que mantiene unidos los protones y los neutrones del núcleo se denomina, simplemente, fuerza fuerte. Pero ¿cuál es su origen? ¿Puede ser comprendida del mismo modo que los físicos han llegado a dominar el electromagnetismo, la fuerza mejor conocida de todas? Éste fue el reto que condujo a la ecuación de Yang-Mills.

Las claves del electromagnetismo

La esencia del electromagnetismo es la carga eléctrica, la cual puede ser positiva o negativa. Las corrientes eléctricas son cargas en movimiento. Las ondas de radio que nos llegan de una emisora provienen del movimiento sincronizado de cargas eléctricas en el seno de ciertos dispositivos oportunamente denominados osciladores. Los portadores de esas cargas eléctricas son (básicamente) electrones, los componentes de los átomos dotados de carga negativa. Sin embargo, los principios fundamentales del electromagnetismo eran ya conocidos mucho antes de que la estructura del átomo y su contenido estuvieran claros. El motivo de que así fuera es doble. En primer lugar, es el concepto de carga

eléctrica y no el de átomo o electrón lo que resulta fundamental a la hora de estudiar la fuerza electromagnética; en segundo lugar, la fuerza electromagnética es de largo alcance y se extiende mucho más allá de los límites de un único átomo, por lo que sus efectos pudieron ser medidos sin dificultad hace más de dos siglos.

Pero ¿qué es la carga eléctrica? Según una definición que parece peligrosamente circular, la carga eléctrica es la fuente del campo electromagnético. Ese campo es la región de influencia alrededor de una carga que determina la fuerza que otra carga percibe a cierta distancia. La fuerza es la materialización del campo (en cierto sentido, es el efecto real y mensurable de ciertos tentáculos invisibles que se extienden desde la carga). Dos cargas del mismo tipo —dos positivas, por ejemplo— se repelen entre ellas: la fuerza que existe entre ellas tiende a separarlas. Esa fuerza decrece rápidamente a medida que las cargas se distancian, hasta que la influencia mutua se hace insignificante. El campo electromagnético que rodea a cada carga determina la fuerza que experimentan las demás.

Las cargas eléctricas estacionarias producen campos eléctricos, pero las cargas en movimiento crean también campos *magnéticos*. Los electrones que giran en los átomos de un pequeño imán crean los campos magnéticos que atraen clavos o crean patrones característicos en las limaduras de hierro de los experimentos escolares. Sin embargo, las *cargas magnéticas* no existen, según parece. Serían polos magnéticos aislados; pero los imanes siempre

tienen un número par de polos, típicamente dos, denominados norte y sur.

Para calcular la fuerza electromagnética debida a una carga eléctrica, o a un conjunto de cargas, se requieren ecuaciones que describan los campos eléctricos y magnéticos básicos. En la década de 1860, el físico escocés James Clerk Maxwell consiguió sintetizar todo el saber que existía sobre la electricidad y el magnetismo en un bello y conciso conjunto autoconsistente de ecuaciones. Al igual que las ecuaciones del movimiento de Newton, las ecuaciones de Maxwell se siguen utilizando hoy día. Constituyen la receta para calcular los campos eléctricos creados por cargas eléctricas o campos magnéticos y para calcular los campos magnéticos producidos por corrientes eléctricas. Las ecuaciones incorporan también un importante aspecto del electromagnetismo: la conservación de la carga eléctrica.

La conservación de la carga significa simplemente que la carga ni se crea ni se destruye. Cuando *cargamos* algo —al peinarnos en un ambiente seco o al recargar la batería del coche—, nos limitamos a redistribuir cargas ya existentes: básicamente, electrones atómicos. En la naturaleza existen procesos que *pueden* crear una partícula cargada, como un electrón, pero siempre dan lugar también a otra partícula con la carga opuesta. La fuerza electromagnética puede crear un electrón negativo junto con una partícula similar con carga positiva denominada antielectrón o positrón. El electrón y el positrón surgen siempre en el mismo lugar, lo cual tiene una

consecuencia importante. La conservación de la carga es más que una propiedad global de un sistema, por la que una carga positiva creada *aquí* se compensa con la aparición de otra carga negativa *allá*. Se trata de una propiedad *local* aplicable a todos y cada uno de los puntos del universo y en cualquier instante del tiempo. Una de las facetas más bellas de las ecuaciones de Maxwell es que garantizan esta *conservación local* de la carga eléctrica y lo hacen a través de la simetría inherente al comportamiento de la fuerza electromagnética.

Casi un siglo después de Maxwell, Chen Ning Yang se planteó recorrer el camino opuesto, tratando de comprender la fuerza fuerte entre las partículas. ¿Se podría partir de una magnitud a conservar y hacer uso de la simetría para descubrir las ecuaciones de la fuerza fuerte?

La importancia de la simetría

En matemáticas se entiende por simetría el que algo resulte inalterado tras realizar una acción concreta, como, por ejemplo, un cuadrado que gire 90 grados o un círculo que gire un ángulo cualquiera. En 1918, la joven matemática alemana Emmy Noether descubrió una relación profunda entre la simetría y la conservación de magnitudes físicas, tales como la carga eléctrica. Constató que cada magnitud que se conservaba llevaba aparejada una simetría, y viceversa.

En un sistema dinámico en el que ciertos objetos se mueven por efecto de unas fuerzas, se conservan tanto la energía como el momento; dicho de otra manera, el saldo neto de estas magnitudes permanece constante. Cuando un cohete espacial parte hacia la Luna, posee un momento que no tenía cuando estaba inmóvil en la rampa de lanzamiento. Para compensarlo, el momento terrestre cambia, si bien de manera imperceptible debido a la enorme masa de la Tierra; ambos son iguales en magnitud pero de sentidos opuestos, con lo que la suma vale cero, como antes del lanzamiento; el momento se conserva. Pero ¿qué clase de simetría está involucrada en la conservación del momento? Se trata de la simetría de las ecuaciones del movimiento en distintos puntos del espacio. El movimiento desde la rampa de lanzamiento en la Tierra hasta un punto de la trayectoria hacia la Luna no altera las ecuaciones básicas: ésta es la simetría. La conservación del momento garantiza esta simetría, y viceversa.

El teorema de Emmy Noether dice que, dado que la carga eléctrica se conserva en todo momento, en la fuerza electromagnética debe existir una simetría relacionada. En efecto, la hay y se refiere a algo conocido como *potencial*, una forma de caracterizar el campo debido a una fuerza, ya sea eléctrica, gravitatoria o de otra clase.

El potencial constituye un modo más *compacto* de describir un campo, al igual que un mapa hipsométrico bidimensional es una representación abreviada de un paisaje en tres dimensiones. Las curvas de nivel del mapa unen puntos de la misma altitud; cuanto

más juntas estén, más escarpado es el terreno. El mapa hipsométrico contiene toda la información que una persona entrenada necesita para recorrer las montañas. De manera similar, el potencial eléctrico de un conjunto de cargas, por ejemplo, contiene toda la información que un físico necesita para calcular el campo eléctrico y, por lo tanto, las fuerzas eléctricas que hay en juego en el sistema.

Muchos de nosotros conocemos el potencial eléctrico a través de su sinónimo, el *voltaje* (o *tensión*). Un pájaro posado en un cable de alta tensión puede cantar tan tranquilamente como si estuviera en la rama de un árbol, gracias a que los campos eléctricos que dan lugar a las fuerzas eléctricas dependen de las *diferencias* de voltaje o potencial eléctrico. Si el potencial de todo el planeta se elevara 1.000 voltios, nuestras centrales y aparatos eléctricos funcionarían exactamente igual que ahora. Lo que importa es la diferencia entre el *activo* y la *masa* (o *tierra*) y no sus valores absolutos. Esta *invariancia* es un ejemplo de simetría *global*: el campo eléctrico no varía si se agrega (o se sustrae) el mismo potencial a todo punto del espacio y el tiempo. Del mismo modo, las ecuaciones de Maxwell no se ven afectadas por un cambio global del potencial eléctrico, ya que se refieren básicamente a campos y no a potenciales.

Sin embargo, las ecuaciones de Maxwell contienen también una invariancia o simetría local más rigurosa. Aunque el potencial eléctrico varíe en diferentes cantidades en distintos puntos del espacio y el tiempo, las ecuaciones de Maxwell siguen siendo las

mismas. Se trata de una invariancia local y se debe a que las cargas eléctricas llevan aparejados campos magnéticos además de campos eléctricos. Los cambios locales en el potencial eléctrico dan lugar a variaciones también locales en otro potencial, conocido como potencial magnético. El efecto neto de los cambios en ambos potenciales hace que los campos eléctricos y magnéticos descritos por las ecuaciones de Maxwell permanezcan inalterados aunque esos cambios sean locales. Existe una simetría local en las ecuaciones de Maxwell y es esa simetría la que parece estar vinculada con la conservación de la carga eléctrica.

Ondas y partículas

La existencia de una fuerza descrita mediante ecuaciones —la fuerza electromagnética— tras la que se revela un principio de simetría produce una especial sensación de belleza. El hecho plantea la posibilidad de que los procesos físicos que observamos — en otras palabras, las propias interconexiones que constatamos entre la electricidad y el magnetismo— provengan de simetrías locales. Y esto nos lleva de nuevo hasta Yang y Mills, quienes trataban de deducir las ecuaciones de las interacciones fuertes entre partículas a partir del principio de invariancia local.

En los cien años transcurridos entre Maxwell y Yang y Mills, se había producido una gran revolución en la física con el advenimiento de la mecánica cuántica, la cual sustituye a la mecánica newtoniana cuando el sistema a estudiar es muy

pequeño. A escala atómica, no es posible saber exactamente a la vez dónde se halla una partícula y a qué velocidad se mueve, ya que el mero acto de observar perturba a la partícula. Podemos medir la velocidad de un coche mediante un radar que haga rebotar ondas de radio en él al pasar por un punto determinado. La energía de esas ondas es tan pequeña que no afecta al movimiento del vehículo. Pero si en lugar de un coche se trata de una molécula, las ondas tendrán la suficiente energía como para desviar la molécula. La mecánica cuántica se enfrenta al problema básico de no conocer simultáneamente la posición y la velocidad (o, en términos más rigurosos, el momento) tratando las partículas como *ondas* y describiendo matemáticamente una partícula mediante la denominada función de onda, que está relacionada con la probabilidad de encontrar dicha partícula en un estado concreto. Al igual que los voltajes pueden ser incrementados o disminuidos sin que cambie el campo eléctrico entre ellos, una onda también puede ser modificada de un modo que no altera sus efectos globales. La propiedad de la onda que podemos hacer variar en este sentido es la denominada fase, la cual viene a ser una medida del punto del patrón ondulatorio en el que se halla la onda en un instante dado. El valor de la fase en una posición fija va variando a medida que la onda sube y baja. Un cambio de fase aplicado a la onda en su totalidad desplaza únicamente el patrón y no afecta a propiedades importantes, tales como la intensidad o la longitud de onda.

De la misma manera, si aplicamos un desfase constante a la función de onda que describe a una partícula, el comportamiento observable de ésta no varía. He aquí otro ejemplo de simetría global. ¿Existe también una simetría local, como en las ecuaciones de Maxwell? Supongamos que el desfase es local y que varía en distintos puntos del espacio y el tiempo. ¿Resultan en este caso afectadas las ecuaciones de la mecánica cuántica que describen la partícula?

La respuesta es, en principio, afirmativa, con lo que parece que deberíamos abandonar esta línea de razonamiento y nuestra obsesión por la simetría local. Sin embargo, si tratamos de modificar las ecuaciones de la partícula de un modo que no les afecte ese cambio de fase local, nos topamos con un notable descubrimiento. Las ecuaciones son invariantes suponiendo que la partícula se mueve bajo la influencia de algún campo de fuerzas. La situación reproduce la conexión existente en el electromagnetismo entre los cambios locales en los potenciales eléctrico y magnético, sólo que ahora las variaciones locales en la fase de la partícula están relacionadas con cambios locales en el campo en cuyo seno se mueve la partícula. El descubrimiento es todavía más notable cuando constatamos que el campo electromagnético proporciona justamente las modificaciones requeridas para las ecuaciones de la mecánica cuántica —suponiendo que hacemos depender el desfase de la partícula de la carga eléctrica de ésta—. Según parece, el

principio de invariancia local revela la naturaleza de las interacciones electromagnéticas de las partículas cargadas.

El matemático alemán Hermann Weyl fue el primero en darse cuenta de la profunda conexión entre la invariancia local en la función de onda de una partícula y la teoría electromagnética. Llamó a esa invariancia «invariancia *gauge*»,⁹⁸ ya que al principio pensó en cambios de escala o de *calibre* en vez de variaciones de fase. En su ya clásico artículo publicado en 1929, decía: «Creo que este nuevo principio de invariancia *gauge*, que no surge de la especulación, sino de los experimentos, nos dice que el campo electromagnético es un fenómeno que necesariamente acompaña [...] a los campos de onda del material [...]».

Así pues, Weyl se atrevía a proponer que la invariancia *gauge* —una simetría básica— podía ser utilizada como principio a partir del cual cabría deducir la teoría electromagnética. En el caso del electromagnetismo, se trataba de una idea elegante pero que no aportaba nada nuevo, ya que la fuerza electromagnética era bien conocida y estaba perfectamente caracterizada por las ecuaciones de Maxwell. La propuesta de Weyl resultaba de mucho más interés para una fuerza, como la fuerza fuerte, en la que lo equivalente a las ecuaciones de Maxwell no existía todavía. ¿Sería posible encontrar unas ecuaciones similares partiendo del principio de simetría adecuado? Cuando Weyl publicó su artículo, aún no se conocía del todo la composición del núcleo atómico y no existía la noción de

⁹⁸ Habitualmente, el término *gauge* («calibre» o «medida» en inglés) se emplea en traducir en este contexto. (N. del T.)

fuerza fuerte. A la aplicación del principio de Weyl a un nuevo campo aún no le había llegado su hora.

Una nueva clase de simetría

Veinte años más tarde, esas profundas ideas que vinculaban la simetría con el electromagnetismo subyugaron a un joven físico chino que empezaba sus estudios en la Universidad de Chicago. Chen Ning («Frank») Yang era hijo de un profesor de matemáticas y había llegado a Estados Unidos en 1945. Había adoptado el nombre de Franklin —de ahí su diminutivo— en honor de Benjamín Franklin, cuya autobiografía había leído en China. En la Universidad de Kunming, en la provincia de Yunnan, y después en Chicago, Yang estudió minuciosamente los artículos que sobre la teoría de campos escribiera Wolfgang Pauli, uno de los físicos teóricos más sobresalientes de la época. Yang confesaba después que le había «impresionado mucho la idea de que la conservación de la carga tenía relación con la invariancia de la teoría relativa a los cambios de fase e incluso más aún el hecho de que la invariancia *gauge determinara* todas las interacciones electromagnéticas».

Al principio, Yang no sabía que esas ideas se debían a Weyl y seguía sin saberlo cuando ambos coincidieron en el Instituto de Estudios Avanzados de Princeton y hasta se reunían de vez en cuando. Weyl había abandonado Alemania en 1933 y obtenido un puesto en Princeton, convirtiéndose en ciudadano estadounidense en 1939, mientras que Yang llegó al Instituto en 1949. Al parecer, Weyl, que

falleció en 1955, nunca supo del trascendental artículo que Yang y Mills escribieron, en el que se demostraba por primera vez cómo la simetría de la invariancia *gauge* podía, en efecto, determinar el comportamiento de una fuerza fundamental.

Durante su estancia en la Universidad de Chicago, Yang había empezado a aplicar esas ideas a otra propiedad de las partículas que, como la carga eléctrica, se conserva en sus interacciones. Su objetivo era hallar las ecuaciones que describen el campo asociado a la invariancia *gauge* de esa propiedad, denominada con el no muy afortunado nombre de «espín isotópico» o «isospín». El isospín es una especie de etiqueta que acompaña a las partículas dotadas de distinta carga eléctrica pero que, de no ser por este hecho, parecerían la misma. Imaginemos a un par de gemelos idénticos, Peter y Paul, vestidos de la misma manera, salvo que uno de ellos lleva abrigo. Si eliminamos el abrigo, no podremos distinguir a uno de otro, aunque tengan nombres diferentes. Lo mismo sucede con partículas tales como el protón y el neutrón; el protón lleva un *abrigo* de carga positiva, mientras que el neutrón está *desnudo*, es decir, sin carga. Los estudios de los núcleos atómicos desarrollados en la década de 1930 revelaban que, dejando aparte las diferencias debidas a la distinta carga eléctrica —una vez retirado al protón su imaginario *abrigo* de carga—, neutrones y protones, neutrones y protones y protones interactuaban todos de la misma manera. Dicho en otras palabras, la *otra* fuerza existente entre las partículas —la fuerza fuerte— no apreciaba diferencias entre ellas.

El protón y el neutrón, que poseen masas muy semejantes, son, para la fuerza fuerte, dos estados de una misma partícula, el *nucleón*; como los nombres de los gemelos, el valor del isospín es lo único que los diferencia. La situación reproduce el hecho por el que una partícula puede aparecer en distintos estados de la propiedad cuántica denominada *espín*; de hecho, la matemática que describe los estados de espín de una partícula sirve también para describir los estados de isospín.

En términos matemáticos, podemos *rotar* el isospín de un protón, convirtiéndolo en un neutrón, y los efectos de la fuerza fuerte sobre la partícula no cambiarán. Existe una simetría en la fuerza y, según el teorema de Emmy Noether, algo tiene que conservarse; ese algo es el isospín. Ahora ya disponemos de todos los elementos para abordar la ecuación de Yang-Mills.

Un nuevo tipo de campo

Desde 1949, Yang trató en varias ocasiones de aplicar al isospín los procedimientos de la invariancia *gauge* utilizados en el electromagnetismo. Sus intentos, en palabras del propio Yang, le llevaban siempre «a un callejón sin salida», a tropezar una y otra vez en el mismo punto de los cálculos, cuando trataba de definir la intensidad del campo asociado. Pero no se dio por vencido. Tal como explicaba en sus *Selected Papers*: «Este tipo de fallo sistemático en una idea aparentemente buena es, desde luego, un lugar común para cualquier investigador. La mayoría de esas ideas son

finalmente abandonadas o aparcadas. Pero algunas sobreviven y llegan a convertirse incluso en una obsesión. Y, muy de vez en cuando, una de esas obsesiones resulta ser algo bueno». A este respecto, los experimentos desvelaban muchos tipos de partículas de corta vida, sugiriendo otras tantas ideas sobre las fuerzas que había tras sus interacciones. Para Yang, «la necesidad de un principio que permitiera explicar dichas interacciones se hizo aún más obvia».

Yang retomó esas ideas en el Laboratorio Brookhaven, en el verano de 1953, contagiando su obsesión a Robert Mills, el joven físico con el que compartía despacho. Juntos lograron superar el escollo que siempre había detenido a Yang y descubrieron las ecuaciones del campo asociado a la simetría *gauge* del isospín.

Si ignoramos los efectos electromagnéticos, la elección entre lo que llamamos protón y lo que denominamos neutrón se convierte en arbitraria; cambiando todos los neutrones por protones y viceversa, las reacciones nucleares seguirían siendo las mismas. Esto equivale a un cambio global de los estados de isospín: habríamos *rotado* el isospín en todos los puntos del espacio y el tiempo una misma cantidad, con lo que todos los protones se transformarían en neutrones y todos los neutrones, en protones. Pero ¿qué sucedería, se preguntaron Yang y Mills, si hiciésemos cambios diferentes en distintos puntos espaciotemporales? Supongamos que la *rotación* entre dos estados de isospín es completamente arbitraria o «carente de sentido físico», como decían Yang y Mills en su artículo. Sería

como el desfase arbitrario en la función de onda de una partícula cargada, que se ve compensado por una alteración del campo electromagnético. ¿Existe algún campo que compense de manera similar los cambios locales de isospín y asegure que las reacciones nucleares parezcan siempre las mismas?

La teoría asociada al isospín resulta ser más compleja que la electromagnética por un motivo fundamental. El campo compensador debe ser capaz de ajustar las variaciones locales o «rotaciones» de isospín de modo que se mantenga la identidad de un protón o un neutrón en cualquier parte. Para que así sea, el propio campo ha de poseer la propiedad de isospín. En cambio, en el electromagnetismo, las variaciones locales en la función de onda de una partícula no alteran la carga eléctrica de ésta. Lo cual queda reflejado en el hecho de que el campo electromagnético no cambia las cargas eléctricas. La carga eléctrica puede ser definida como la fuente del campo electromagnético, pero el campo electromagnético en sí no es una fuente de carga eléctrica. En la teoría de Yang-Mills, sin embargo y aunque suene endogámico, el campo es una fuente de sí mismo.

La ecuación de Yang-Mills es la ecuación del movimiento de ese campo. Equivale a las ecuaciones de Maxwell —o a las del movimiento de Newton— y se escribe de una forma parecida. En la notación empleada por Yang y Mills, la ecuación dice así:⁹⁹

⁹⁹ C. N. Yang y R. L. Mills, «Conservation of isotopic spin and isotopic gauge invariance», *Physics Review*, 1954, 96, págs. 191-195.

$$\partial f_{\mu\nu} / \partial x_\nu + 2\varepsilon(b_\nu \times f_{\mu\nu}) + J_\mu = 0$$

Donde $f_{\mu\nu}$ representa la intensidad del campo de Yang-Mills y $\partial/\partial x_\nu$ indica que la ecuación depende del modo en que la intensidad del campo varía en el espacio y el tiempo; ε desempeña el papel de «carga» y J_μ representa la corriente asociada; b_ν es el potencial del campo. El término $(b_\nu \times f_{\mu\nu})$ plasma la gran diferencia con el electromagnetismo, ya que expresa la dependencia del campo de Yang-Mills respecto a sí mismo. En las ecuaciones de Maxwell, el término equivalente es nulo, ya que los campos fundamentales no se ven afectados mutuamente.

El problema de la masa

En la teoría de Yang y Mills aún quedaba un obstáculo por superar: el relativo a las «partículas de campo». En la teoría cuántica de campos —el marco teórico en el que Yang y Mills desarrollaron sus trabajos—, los campos están representados por partículas. Esas «partículas de campo» son más que una oportuna convención matemática; bajo ciertas circunstancias, emergen de ese campo como entes detectables, tan reales como los protones o los electrones. En la teoría electromagnética, las partículas de campo son los fotones, que pueden emerger del campo electromagnético en forma de luz.

Las partículas de campo se comportan como pelotas de un juego de «captura cuántica» entre «partículas de materia» que interaccionan,

tales como los protones y los electrones. En el caso electromagnético, las partículas cargadas interactúan jugando a *capturarse* con fotones. Los fotones no tienen masa, con lo cual las interacciones pueden tener lugar a distancias muy grandes; en principio, infinitas. (Podemos imaginar una *pelota* fotón lanzada infinitamente lejos). En cambio, el alcance de la fuerza fuerte entre protones y neutrones parece estar limitado a las dimensiones del núcleo. (Esto implica que la *pelota* fuerte ha de tener cierta masa para que el intercambio —la interacción— se desarrolle siempre durante un tiempo limitado o, dicho de otro modo, a corta distancia).

El nuevo campo que Yang y Mills habían hallado ajustaría el isospín lo que fuera preciso en cada punto del espacio y el tiempo, convirtiendo los protones en neutrones o los neutrones en protones o dejándolos a todos inalterados. Para ello, se requerían tres partículas portadoras, con otros tantos estados de isospín. El campo podía también cambiar la carga eléctrica, transformando, por ejemplo, un protón con carga positiva en un neutrón sin carga. Por consiguiente, dos de las partículas portadoras tenían que poseer carga —positiva o negativa—, mientras que la tercera sería neutra y participaría en las interacciones entre protones y protones o entre neutrones y neutrones. Yang y Mills conocían, pues, la carga y el isospín de las nuevas partículas de campo, pero no tenían ni idea de sus masas y admitían que ello constituía un punto débil en su teoría. Cuando, en febrero de 1954, Yang presentó la teoría en un

seminario en Princeton, fue increpado nada menos que por Pauli. Apenas había escrito en la pizarra una expresión relativa al nuevo campo, cuando el gran físico le preguntó: «¿Cuál es la masa de ese campo?». Yang explicó que se trataba de un problema complejo y que Mills y él no habían llegado a conclusión definitiva alguna, a lo que Pauli replicó, tajante: «Eso no es suficiente excusa».

Aunque en febrero de 1954 habían completado la mayor parte de su trabajo, Yang y Mills se abstuvieron de publicar artículo alguno. Como escribiría Yang, «La idea era *hermosa* y debía ser publicada. Pero ¿cuál era la masa de la partícula *gauge*? No disponíamos de conclusiones firmes, sólo de experiencias frustrantes que demostraban que este caso era mucho más enrevesado que el electromagnetismo. Pensábamos, basándonos en razones físicas, que las partículas *gauge* cargadas no podían carecer de masa». Es el propio Yang quien pone en cursiva la palabra *hermosa* y no hay duda alguna de que tenía razón. Mills y él enviaron su artículo a la prestigiosa revista *Physical Review* a finales de junio de 1954, y fue publicado tres meses después, el 1 de octubre. Su penúltimo párrafo concluía lamentando que aún «no hubieran podido concluir nada acerca de la masa del cuanto *b'* es decir, el portador de su nuevo campo».

La unificación electrodébil

La progresiva comprensión de las partículas elementales y las fuerzas se derivó, como casi siempre, de una combinación de ideas y

descubrimientos: de la teoría y la experimentación. Como en un dúo musical, ambas se complementan recíprocamente: unas veces prevalece una de ellas y otras, la otra. En ocasiones un instrumento desarrolla una nueva melodía, mientras el otro continúa con un tema anterior. Más tarde, uno de los fragmentos puede convertirse en el tema dominante. De manera similar, los físicos exploran distintos caminos mediante ideas teóricas e investigaciones experimentales. Algunos resultan ser callejones sin salida y son abandonados, mientras que otros pueden llegar a constituir el fundamento de una etapa posterior. El enfoque de Yang-Mills tal vez fracasó en su intento de explicar el misterioso comportamiento de la fuerza fuerte, pero hoy es el eje en torno al que gira nuestro conocimiento de las partículas y las fuerzas. En cualquier caso, sólo avances teóricos y descubrimientos experimentales posteriores pusieron de relieve el modo en que dicho enfoque tiene que ver con la naturaleza de las fuerzas entre las partículas.

En octubre de 1979, un cuarto de siglo después de la publicación del artículo de Yang y Mills, tres físicos teóricos fueron galardonados con el Premio Nobel de Física. Sheldon Glashow, Abdus Salam y Steven Weinberg habían construido, de manera independiente, un nuevo marco teórico basado en el principio de invariancia local. Las ideas de Yang y Mills y las de Weyl, antes que ellos, habían fructificado, si bien de una forma inesperada.

La nueva teoría trataba sobre dos fuerzas a la vez: no eran la fuerte y la electromagnética, como cabría esperar de la línea seguida por

Yang y Mills, sino las fuerzas electromagnética y débil. La «teoría electrodébil» resolvía también el problema de la masa, incorporando partículas de campo pesadas. Más aún, la teoría predecía las masas de esas partículas (con la ayuda de ciertas magnitudes que era posible medir).

La fuerza débil se halla detrás de algunas clases de radiactividad, en las que los núcleos atómicos «se desintegran» cuando alguno de los neutrones que contienen se transforma en un protón, o viceversa. Estos procesos constituyen verdadera alquimia, ya que alteran el número de protones del núcleo y esto, a su vez, cambia la naturaleza química del átomo al que el núcleo pertenece. El carbono puede convertirse en nitrógeno, el plomo en bismuto, etc. De manera parecida, los protones se transforman en neutrones en la cadena de reacciones nucleares que tienen lugar en el corazón del Sol y otras estrellas. Así pues, aunque la fuerza débil es unas cien mil veces más pequeña que la fuerza fuerte en el seno del núcleo atómico, tiene una influencia directa y profunda en la naturaleza de nuestro universo y, a través de nuestro Sol, en la vida misma.

En el contexto de los fenómenos de la vida diaria, sorprende que las fuerzas débil y electromagnética estén tan estrechamente vinculadas. El elevado alcance de la electricidad y el magnetismo da pie a fenómenos a gran escala, como los relámpagos y las auroras boreales, mientras que la fuerza débil actúa clandestinamente, a escala subatómica. La energía vital que recibimos del Sol nos llega en forma de fotones de luz —las partículas del campo

electromagnético—, si bien dicha energía es liberada en reacciones desencadenadas por las interacciones débiles que tienen lugar en núcleos alojados en las profundidades del astro rey. Es esa relación entre fenómenos aparentemente inconexos la que hallaron Glashow, Salam y Weinberg, aunque ninguno de ellos se lo había propuesto.

En el Reino Unido, Abdus Salam trataba de explicar las interacciones débiles entre partículas en términos de invariancia local. La fuerza débil puede alterar la carga eléctrica de aquéllas —al convertir un neutrón en un protón, por ejemplo—. Salam propuso que la fuerza débil podía proceder de un campo como el descrito por Yang y Mills, con tres «partículas de campo» de carga eléctrica positiva, negativa y nula, respectivamente. Era fácil asociar las partículas de campo con carga positiva y negativa a interacciones débiles conocidas que alteraban la carga, pero la partícula neutra resultaba más problemática. Una opción tentadora era identificarla con una partícula de campo neutra ya conocida —el fotón del electromagnetismo—, con lo que la idea de la «unificación electrodébil» comenzó a tomar cuerpo en la mente de Salam.

En Estados Unidos, Sheldon Glashow adoptaba un camino similar, aunque con otro objetivo. El problema que intentaba resolver eran los infinitos sin significado físico que siempre aparecían en las teorías sobre la fuerza débil. Pensaba que, al incluir en una misma teoría tanto la fuerza débil como la electromagnética, las partes de los cálculos que producían esos infinitos se cancelarían. Eligió como punto de partida el enfoque de Yang-Mills y, al igual que Salam,

asumió que la partícula neutra era el fotón. No obstante, tanto Salam como Glashow se dieron cuenta, cada uno por su lado, de que la teoría funcionaba mejor si se trataban las simetrías electromagnética y débil de manera independiente. El resultado era una teoría con dos partículas de campo neutras: el fotón del electromagnetismo y una partícula neutra distinta para el campo débil.

La teoría presentó al principio algunos obstáculos, uno de los cuales era el problema de la masa, que ya había puesto en dificultades a Yang y a Mills. Al igual que en el caso de la fuerza fuerte, el alcance de la débil resulta ser muy pequeño, lo que implica que las *pelotas* débiles en el juego de la *captura cuántica* debían ser muy pesadas. En la teoría electrodébil, mientras los fotones carecían de masa, las partículas positiva, negativa y neutra del campo débil tenían que poseer una gran masa todas ellas. Pero al dotar de masa a las partículas del campo se destruía la invariancia local y, con ello, la razón de ser de este enfoque. Más aún: para consternación de Glashow, el problema de los infinitos seguía sin resolverse y, lo que era peor, ninguna evidencia experimental avalaba la existencia de la partícula neutra y pesada que la teoría requería.

La solución al problema de la masa tuvo un origen inesperado: un área completamente distinta de la física, que se ocupa del modo en que se comportan colectivamente los átomos de los sólidos. La clave fue el concepto de que un sistema físico puede existir en un estado carente de simetría, aunque las ecuaciones que lo describan sean

simétricas. Los átomos de hierro, por ejemplo, se comportan como imanes diminutos. En un fragmento ordinario de hierro, esos imanes atómicos apuntan en direcciones aleatorias, con lo que existe simetría, pues no hay ninguna dirección preferente. Sin embargo, el hierro puede ser magnetizado, en cuyo caso los imanes atómicos se alinean en la dirección del campo magnético. La simetría parece evaporarse, aunque las ecuaciones que describen el movimiento de los átomos mantengan su simetría original. Varios teóricos, entre los que se encontraba Peter Higgs, de la Universidad de Edimburgo, observaron que podían aplicar esas ideas y permitir que las partículas adquirieran masa, introduciendo un campo adicional en sus ecuaciones, un campo que hoy se conoce con el nombre de Higgs.

El campo de Higgs es inusual en el sentido de que, aunque el potencial asociado a él es simétrico, las soluciones de las ecuaciones de movimiento en su seno son asimétricas. En efecto, el potencial de Higgs es como el fondo curvado de una botella de vino: la forma en su conjunto es simétrica, pero un guisante colocado en equilibrio en lo alto de ella acabará rodando en alguna dirección, rompiendo la simetría. Traducido a las ecuaciones que describen las interacciones entre partículas, éstas serían como el guisante: inicialmente no tienen masa, pero al interactuar con el campo de Higgs, rompen la simetría y la adquieren.

En Estados Unidos, a Steven Weinberg le pareció prometedor usar las ideas de la ruptura de simetrías en una teoría de tipo Yang-Mills

para describir la interacción fuerte. Al principio no tuvo éxito al tratar de identificar las partículas de campo con y sin masa de su teoría con las partículas conocidas de la interacción fuerte. Luego, «en cierto momento, a finales de 1967», tal como recordaba en su discurso de aceptación del Premio Nobel, «mientras conducía hacia mi oficina en el Instituto Tecnológico de Massachusetts, me di cuenta de que había estado aplicando las ideas correctas al problema equivocado». La partícula sin masa que necesitaba era el fotón y las partículas masivas, las correspondientes al campo débil. «Las interacciones electromagnética y débil podían, pues, ser descritas de manera unificada en términos de una simetría *gauge* exacta, aunque rota espontáneamente».

Cuatro años más tarde, en 1971, tendría lugar el añadido teórico final que convertiría «al sapo Weinberg-Salam-[Glashow] en un príncipe encantado», en la evocadora frase de Sidney Coleman. Gerard 't Hooft, que trabajaba con Martin Veltman en Utrecht, demostró que los infinitos de la teoría se cancelaban (mediante un proceso conocido como «renormalización»). Glashow contemplaba, pues, cómo desaparecían los últimos obstáculos. «Al buscar la renormalizabilidad», escribiría después, «había trabajado diligentemente, pero en una senda errónea. La simetría *gauge* es una simetría exacta, pero se halla escondida. No se puede introducir la masa sin más» (como había intentado). Los trabajos de Hooft y Veltman convirtieron el enfoque electrodébil en teoría acreditada. Ambos físicos recibieron el Premio Nobel en 1999 en reconocimiento

al *toque mágico* que había convertido al *sapo electrodébil* en un príncipe entre las teorías.

En el transcurso de una década, entre 1973 y 1983, muchas de las piezas clave ocuparon su lugar. En 1973, los experimentos revelaron los primeros indicios de «corrientes débiles neutras». Esas reacciones, nunca antes observadas, revelaban la existencia de la partícula neutra pesada de la fuerza débil. En 1983, tanto la partícula del campo débil cargada como la neutra fueron generadas y detectadas en colisiones de alta energía y sus masas correspondieron a las predichas por la teoría electrodébil. Se trataba de una espectacular confirmación de las ideas básicas de Yang y Mills.

La fuerza del color

¿En qué situación quedaba la fuerza fuerte, la que trataban de describir Yang y Mills, tras todos esos progresos en las fuerzas débil y electromagnética? La década de 1960 había traído grandes cambios —tanto en el terreno social como en el científico—, y uno de ellos fue el modo de contemplar la naturaleza real de las partículas elementales. El protón, el neutrón y toda la pléyade de partículas de vida corta resultaban estar formados por otras partículas aún más fundamentales, los denominados quarks. El protón y el neutrón, por ejemplo, constaba cada uno de tres quarks unidos por la fuerza fuerte. Estaba claro que dicha fuerza tenía que ver con alguna propiedad de los quarks.

Los teóricos comenzaron a pensar que, si tres quarks idénticos debían formar una partícula similar al protón, esos quarks tenían que poseer cierta propiedad característica. Para cumplir las leyes de la teoría cuántica, la propiedad debía diferenciar unos quarks que, en caso contrario, serían idénticos. Por analogía con los tres colores primarios de la luz, esa propiedad con tres valores posibles fue denominada «color» y los valores que podía adoptar, rojo, verde y azul. En cualquier caso, estaba claro que era el color y no el isospín lo que constituía la «carga fuerte» —la fuente de las interacciones fuertes entre los quarks.

La teoría que los físicos han construido en torno a los quarks coloreados es del tipo de la que exploraron Yang y Mills, pero como el color tiene tres valores posibles y no los dos que Yang y Mills consideraron para el isospín, los campos resultantes son más complicados. En vez de tres partículas de campo, son necesarias ocho. Esas nuevas partículas son conocidas como gluones y, al igual que los quarks, deben poseer color, de modo que el nuevo campo que garantiza la invariancia local es del tipo del de Yang-Mills: es una fuente de sí mismo. La teoría que describe los campos fuertes originados por los «cambios de color» se denomina cromodinámica cuántica, o CDC, en analogía con la electrodinámica cuántica, o EDC, la teoría cuántica de la fuerza electromagnética. Si la CDC ha resultado ser una teoría exitosa, ¿cómo resuelve el problema de las partículas de campo masivas previstas en la fuerza fuerte de corto alcance?

La respuesta tiene que ver con la complejidad de las interacciones que pueden tener lugar entre los propios gluones, algo que simplemente no se da en la EDC con sus fotones sin carga. Las interacciones de los gluones en la CDC hacen que la intensidad efectiva de la fuerza alrededor de una «carga fuerte» —un quark rojo, por ejemplo— sea menor a corta distancia. El descubrimiento constituyó toda una sorpresa, pues los físicos sabían desde hacía dos siglos que la fuerza debida a una carga eléctrica crece a medida que nos aproximamos a ella. Sin embargo, el nuevo efecto parecía explicar ciertas observaciones paradójicas registradas en la década de 1970, en experimentos que sondeaban protones y neutrones con electrones de alta energía. Los experimentos mostraban que, cuando los electrones exploraban a distancias más cortas, comenzaban a interactuar con los quarks del interior de los nucleones como si estuvieran casi libres o en absoluto vinculados con el ente superior. Esto se ajustaba a la idea de una fuerza fuerte que reducía su intensidad al disminuir la distancia.

¿Qué sucede, entonces, a distancias más grandes? La fuerza fuerte parece incrementar su magnitud. La conclusión sería que un único quark no puede ser extraído de un protón o un neutrón del mismo modo que un electrón puede ser arrancado de un átomo. Realmente, no existe evidencia alguna de que un quark aislado haya sido observado alguna vez. Así pues, la fuerza fuerte aparenta tener corto alcance, limitado al tamaño de las partículas en las que los quarks están atrapados. De todo ello resulta que los gluones no

tienen por qué ser masivos para justificar el corto alcance de la fuerza fuerte. En la CDC, los gluones carecen de masa y no causan problema alguno a las simetrías locales de la teoría.

Una idea de la época

Como en muchos otros avances científicos, el camino desde los trabajos de Yang y Mills hasta la emergencia de la teoría electrodébil y la CDC fue largo y tortuoso. Al aceptar el Premio Nobel en 1979, Glashow hablaba de cómo «el tosco centón» de los años cincuenta se había «convertido en un tapiz» en los setenta. «Los tapices», continuaba, «son elaborados por muchos artesanos que trabajan conjuntamente. Las contribuciones de los distintos operarios no se distinguen unas de otras en el producto acabado y las hebras sueltas o erróneas quedan ocultas. Ésa es nuestra imagen de la física de partículas». En la misma ocasión, cuando a mitad de discurso Salam culminaba su síntesis de la teoría electrodébil, hizo observar que ya había mencionado el nombre de una cincuentena de teóricos.

Ningún científico trabaja en total soledad y menos aún en la actualidad. Incluso a menudo se tiene la sensación de que el progreso científico y los descubrimientos son fruto de la época. Yang y Mills pudieron estar adelantados a su tiempo en el sentido de que llevó casi dos décadas que su creencia en un principio básico diera su fruto, pero también eran hijos de su tiempo. En 1953 y en diferentes partes del mundo, otros científicos empezaban también a

elaborar teorías del mismo tipo. Pauli, cuyos artículos sobre la teoría de campos inspiraron a Yang, comenzó a indagar la posibilidad de extender las transformaciones locales de fase del electromagnetismo al isospín, pero nunca publicó sus trabajos; hoy sólo se tiene noticia de ellos por sus cartas a Abraham Pais. Al parecer, Pauli comprendió que la teoría debía implicar partículas sin masa para los campos, en aparente contradicción con el corto alcance de las interacciones fuertes.

En Cambridge, Ronald Shaw, el alumno de Salam, investigó la invariancia local para el isospín de una forma similar a la de Yang y Mills y dedujo el mismo campo transportado por tres partículas. Shaw escribía después, en 1982:

«Mi campo gauge surgió de mi fascinación por las ideas relativas a la invariancia lanzadas en una edición preliminar (más bien tosca) de Schwinger, con la que tropecé en 1953».

Completó su trabajo en enero de 1954, pero se limitó a incluirlo como un capítulo más en la segunda parte de su tesis doctoral. Esa Parte II, según Shaw, «constaba de varios fragmentos inconexos, incluyendo el capítulo 3 sobre los campos *gauge* SU (2). Recuerdo que me pareció insuficiente [...], y por ello seguí buscando durante el resto del año, hasta que en 1955 creé la Parte I de mi tesis». ¡Le pareció insuficiente! Shaw presentaba su tesis en septiembre de 1955; el artículo de Yang y Mills había aparecido en la *Physical*

Review de octubre del año anterior. Por supuesto, tanto Shaw como Yang y Mills creían que sus teorías describían partículas sin masa que no existían en la realidad, así que hemos de disculpar sus reticencias.

La tercera deducción de la misma teoría tuvo lugar en Japón, donde Ryoyo Utiyama buscaba una estructura matemática que pudiera enlazar la gravedad con el electromagnetismo. En marzo de 1954 completó ciertos trabajos sobre «la idea de una teoría *gauge* general». Tal como el teórico irlandés Lochlainn O’Raifeartaigh explicaba en su libro *The Dawning of the Gauge Theory*, «dado que Utiyama abarcaba también la gravedad, debemos reconocer que su aproximación era la más amplia y generalista. Pero, desde el punto de vista de la prioridad, su contribución apareció después que las de Yang y Mills...». Utiyama había sido invitado a visitar el Instituto de Estudios Avanzados de Princeton y llegó allí en septiembre de 1954. Tuvo enseguida noticia de que Yang había anunciado una teoría similar a la suya y recibió una copia de la edición preliminar. En un libro publicado en 1983, Utiyama recordaba: «Me di cuenta de inmediato de que [Yang] había encontrado la misma teoría que había desarrollado yo. Me sentí muy impresionado al examinar su artículo con detalle y compararlo con mis propios trabajos». Utiyama no regresaría a su teoría *gauge* general hasta marzo de 1955, fecha en la que analizó más detenidamente la propuesta de Yang y Mills. Tras constatar que su enfoque era más general, escribió un artículo para la *Physical Review*, que fue publicado a comienzos de 1956. La

originalidad del trabajo de Utiyama ha sido, por lo general, infravalorada, probablemente —en opinión de O’Raifeartaigh— «porque Utiyama consideró grupos generales y citó a Yang y a Mills, con lo que su artículo de 1956 suele verse como una simple generalización de la teoría de estos últimos». Sin duda, Utiyama tenía razón al reconocer «lo mucho que lamentaba no haber enviado el artículo a una publicación japonesa en marzo de 1954, cuando había completado el trabajo».

Yang y Mills sólo escribieron dos artículos juntos: el famoso artículo de 1954, en el que aparecía por primera vez su ecuación, y otro mucho menos conocido sobre el fotón, escrito en 1966. Por otra parte, mientras que Yang es hoy ampliamente reconocido entre los físicos como uno de los teóricos más brillantes de la segunda mitad del siglo XX, Mills nunca más volvió a aparecer en escena. Tan sólo tres años después de su trabajo con Mills, Yang compartió el Premio Nobel de Física de 1957 con Tsung Dao (T. D.) Lee, un compañero, estadounidense también. Habían hallado que el único modo de explicar las misteriosas propiedades de ciertas partículas subatómicas inusuales era asumir una diferencia entre izquierda y derecha cuando las partículas interactúan mediante la fuerza débil. Proponían la forma de comprobar experimentalmente esa idea en apariencia tan extravagante; y para gran asombro de los físicos, incluyendo el escéptico Pauli, un experimento realizado por Chien Sung Wu y sus colegas demostró que, en efecto, la fuerza débil establece diferencias entre izquierda y derecha. La colaboración

entre Yang y Lee dio lugar a muchos trabajos importantes a lo largo de bastantes años pero, lamentablemente, cesó a finales de los setenta.

A diferencia de Yang, Mills continuó sus investigaciones en física en relativo anonimato. En 1956 entró a formar parte de la Universidad del Estado de Ohio, donde permaneció hasta su jubilación, en 1995. En cualquier caso, Yang siempre le profesó un gran respeto: «Bob tenía una mente brillante. Era muy rápido a la hora de aprehender nuevas ideas», escribía con motivo de la muerte de su colega, acaecida en 1999. «Siempre perdurarán en mi memoria nuestra estrecha colaboración y nuestras muchas discusiones».

Por la misma época, Yang comentaba también que, si bien Mills y él «estaban entusiasmados con la belleza de su trabajo», ninguno de los dos «hubiera previsto su gran impacto en la física». En la actualidad, a comienzos del siglo XXI, su obra sostiene la teoría electrodébil y la cromodinámica cuántica, componentes clave ambas del Modelo Estándar de partículas y fuerzas fundamentales. Parece que la belleza, a través de la simetría, y los fenómenos del mundo físico se hallan inextricablemente vinculados, tal como Werner Heisenberg, uno de los padres de la teoría cuántica, había empezado a constatar. En un ensayo sobre «La belleza y la física teórica», Yang cita unas palabras de Heisenberg, quien en 1973 decía: «Tendremos que abandonar la filosofía de Demócrito y el concepto de partículas elementales. En su lugar, deberemos adoptar la noción de simetrías fundamentales».

Cuando, en el verano de 1953, Yang y Mills trataron de explicar la fuerza fuerte a partir del isospín, se toparon con un principio, basado en la simetría, que da lugar a ecuaciones que enlazan fuerzas y partículas elementales. Con este descubrimiento, avanzaron un gran paso hacia ese ideal que Newton formulara tres siglos atrás. A los teóricos del siglo XXI les corresponde descubrir si del citado principio es posible derivar la unificación total — incluyendo la gravedad—, haciendo que el deseo newtoniano se vea por fin cumplido.

Lecturas recomendadas

Y. Nambu, «Quarks», *World Scientific*, 1985. Una atractiva guía sobre los desarrollos clave en la comprensión de las fuerzas y las partículas.

G. 't Hooft, *In search of the ultimate building blocks*, Cambridge University Press, 1997. Un informe sobre el estado actual de la física de partículas, a cargo de un reciente ganador del Premio Nobel.

G. Fraser, ed., *The Particle Century*, Institute of Physics, 1998. La historia de la física de partículas y el desarrollo del modelo estándar de partículas y fuerzas a lo largo del siglo XX.

C. N. Yang, *Selected Papers with Commentary*, Freeman, 1983. Contiene una recopilación de artículos de Yang e incluye sus trabajos con Mills.

L. O’Raifeartaigh, *The Dawning of Gauge Theory*, Princeton University Press, 1997. Contiene todos los artículos importantes mencionados en el capítulo, junto con un comentario interesante, aunque muy técnico. Muy recomendable para físicos.

Parte 8

Un espejo en el cielo

La ecuación de Drake

Oliver Morton

Desde Galileo, la mirada del astrónomo ha sido siempre el símbolo de una poderosa forma de percepción. El astrónomo ve más y ve más allá que ningún otro ser humano, y el objeto de su mirada es el que más lejos se halla de nuestro alcance. Es indudable que el poder atribuido a la mirada del astrónomo deriva en parte de connotaciones astrológicas. No obstante, el poder del astrónomo como símbolo de la forma de observación científica más pura tiene una profunda relación con el hecho obvio de que no existe conexión alguna entre quien hace la observación y el universo observado. El astrónomo lo ve todo y todo lo que hace es ver, y la circunstancia de que lo haga desde lo alto de una montaña hace que todo sea más romántico.

En las décadas posteriores a la segunda guerra mundial, los astrónomos abandonaron las cumbres. La tecnología ofrecía cada vez más posibilidades de observar mediante instrumentos remotos y en distintas longitudes de onda. Una de esas nuevas formas de ver consistía en el uso de ondas de radio. El método condujo a la creación de una nueva y fascinante forma de astronomía, algo que polarizó a la ciencia y atrajo al público en igual medida. La radioastronomía ha proporcionado un nuevo modo de abordar la

cuestión de si existe vida inteligente fuera de la Tierra, una cuestión que se halla, a la vez, en las fronteras de las ciencias naturales y en su mismo núcleo. En las fronteras no sólo por la dificultad de concretar el objeto, sino también porque éste escapa al alcance de la ciencia natural; por su propia naturaleza, se trata de algo artificial. Y esto, a su vez, lo coloca en pleno corazón de la ciencia. El uso de radiotelescopios para buscar civilizaciones alienígenas supone abordar la cuestión del lugar del hombre en el universo con las herramientas de una ciencia natural. Se trata de responder a la pregunta «¿Qué es el hombre?» no mediante la introspección subjetiva, sino a través de la mirada objetiva del astrónomo: escudriñando a lo lejos, identificando cierto fenómeno lejano y diciendo: «El hombre es una cosa como *ésta*».

La radioastronomía se hallaba aún en su infancia, a finales de la década de los cincuenta, cuando un joven operador llamado Frank Drake se dio cuenta de que había algo muy singular en ella. Él y el resto de los radioastrónomos eran las primeras personas que podían, en principio, detectar la presencia de gente como ellos haciendo su vida alrededor de otras estrellas. Gracias a los radares militares y a las emisoras comerciales de radio y televisión, la Tierra estaba emitiendo ya más energía en la banda de radio del espectro electromagnético que el propio Sol alrededor del que estaba orbitando. Una señal transmitida en forma de haz estrecho desde un radiotelescopio como en el que trabajaba Drake —la parábola de 26 metros del Observatorio Nacional de Radioastronomía de Green

Bank, en las montañas de Virginia occidental— podría ser captada por un telescopio similar en un sistema solar cercano. El observatorio en el que trabajaba podía dedicarse no sólo a detectar estrellas y galaxias, sino también civilizaciones.

Drake no fue la única persona a la que se le ocurrió esta idea; el mismo año, 1959, dos físicos de Cornell, Giuseppe Cocconi y Philip Morrison, publicaron en *Nature* un artículo sobre el tema.¹⁰⁰ Pero Drake no ha sido famoso por la prioridad en la publicación, sino por ser el primero que puso la idea en práctica. El 8 de abril de 1960, antes de que despuntara el alba, trepó hasta el foco de la parábola —una altura de cinco pisos— e instaló en él un amplificador ajustado para captar señales en la estrecha gama de frecuencias que consideró de interés. (El amplificador era un préstamo del MIT y se lo había traído expresamente, colocado en el asiento del acompañante de un Morgan deportivo, su creador, Sam Harris, reputado ingeniero y radioaficionado). Durante más de una hora, Drake manipuló los potenciómetros de ajuste; después, de vuelta en la sala de control, apuntó el telescopio hacia Tau Ceti, situada a 12 años-luz. No detectó nada. Cuando Tau Ceti se ocultó, movió la parábola hacia Épsilon Eridani, a 10,5 años-luz, y entonces recibió una fuerte señal, aunque fugaz. Durante cuatro días, los observadores no supieron interpretar su significado; al quinto, estuvo claro que se había tratado del paso de un avión, en lo que se

¹⁰⁰ Cocconi y Morrison, «Searching for interstellar communications», *Nature*, 1959, 184, pág. 844 y sigs.; reeditado en *The Quest for Extraterrestrial Life: A Book of Readings*, En: Donald Goldsmith, ed., University Science Books, 1980, junto a muchos otros artículos clásicos en este campo.

convirtió en la primera de las muchas falsas alarmas que han sobresaltado a esta ciencia desde entonces.

Drake no sólo fue pionero en el campo de la comunicación con mentes extraterrestres (práctica denominada al principio CETI y más tarde SETI, *Search for Extra-Terrestrial Intelligence* o búsqueda de inteligencia extraterrestre). Un año después de su primer intento elaboró un dispositivo retórico sorprendentemente sólido con el que estructurar toda futura discusión sobre el tema. Aunque la investigación de Drake —conocida como «Proyecto Ozma», en honor de la hija del Mago de Oz— no arrojó resultado alguno, unida al artículo publicado en *Nature* por Cocconi y Morrison, generó el suficiente interés como para que pareciera adecuado organizar un congreso al respecto. La reunión tuvo lugar en Green Bank a finales de 1961 y Drake se encargó de dar forma al programa científico. Decidió estructurarlo como una investigación sobre el número probable de civilizaciones emisoras de radio en la galaxia. Contar las fuentes de diferentes clases de señales era un método de trabajo típico en radioastronomía. Y una estimación del número de fuentes, N , no sólo serviría para establecer los límites de la probabilidad de que los radioastrónomos captaran señales de inteligencias alienígenas; determinaría también la mejor estrategia de búsqueda. Si N era grande, merecería la pena investigar de forma individualizada las estrellas cercanas. Si, por el contrario, era pequeño, sería necesario explorar todo el firmamento.

Si la reunión iba a girar en torno a un número, ese número tenía que ser calculado. El punto de partida de Drake era la tasa, R^* , con la que son creadas en nuestra galaxia estrellas razonablemente semejantes al Sol. Establecida ésta, se trataba de estimar el porcentaje de esas estrellas que poseen planetas, un área de investigación en la que había abierto camino Otto Struve, el jefe y principal valedor de Drake. A continuación, había que determinar el número de planetas en torno a una estrella dada que pudieran ser habitables. Después, la fracción de esos planetas habitables en los que se habría originado la vida. Seguidamente, el porcentaje de estos parientes de la Tierra en los que habría evolucionado la inteligencia y, por último, la fracción de esas inteligencias que producirían civilizaciones tecnológicas. En la asombrosamente pragmática argumentación que dio forma a la idea, la civilización era definida como la infraestructura que hace posible la radioastronomía.

Cada una de esas consideraciones puede ser expresada mediante un simple número. Drake decidió que el porcentaje de estrellas con planetas sería f_p ; el número medio de planetas habitables alrededor de tales estrellas, n_e ; las probabilidades de la existencia de vida, inteligencia y civilización serían f_l , f_i y f_c , respectivamente. Por último, era preciso considerar la vida media de una civilización tecnológica, L . Con relativamente poco esfuerzo, Drake creó una fórmula para calcular N . Expresada en palabras, N es igual a la tasa con la que los planetas habitables aparecen, multiplicada por la

probabilidad de existencia de una civilización tecnológica en un planeta habitable dado y por el periodo típico en que esas civilizaciones se comunican. Expresada al modo en que Drake la escribió en la pizarra del Congreso de Green Bank dice así:

$$N = R^* \times f_p \times n_e \times f_l \times f_i \times f_c \times L$$

Esta simple expresión llegaría a ser conocida como la ecuación de Drake, aunque la fórmula de Drake sería probablemente más precisa. No expresa ley alguna de la naturaleza, simplemente nos dice cómo calcular un número a base de multiplicar siete factores numéricos, tales como R^* , f_p , n_e , etc. La mayoría de las ecuaciones suponen la conclusión de un proceso creativo; destilan nociones profundas, las generalizan y extienden su alcance. La ecuación de Drake, por el contrario, supone un punto de partida. No es una herramienta analítica, sino pedagógica; no es una ecuación para ser utilizada, sino para hablar sobre ella. «No me supuso ningún esfuerzo intelectual ni una profunda reflexión», recordaría Drake más tarde. «Pero [...] expresaba una gran idea de una forma que cualquier científico, incluso un principiante, podía asimilar».¹⁰¹ La síntesis realizada por Drake poseía una elegancia retórica que la ha hecho perdurar. Estructuraba el tema bajo un formato de aspecto científico y, al hacerlo, parecía que sus gigantescas incógnitas se

¹⁰¹ Frank Drake y Dava Sobel, *Is Anyone Out There: The Scientific Search for Extraterrestrial Intelligence*, Delacorte Press, 1992. Se trata de un informe sobre el Congreso de Green Bank que recoge, sobre todo, el punto de vista de Drake.

hacían más abordables. En definitiva, creaba la ilusión de un puente entre los radicalmente distintos tipos de problemas que surgen al pensar en civilizaciones alienígenas, un puente entre cuestiones como de dónde proceden las ondas de radio y de dónde proceden los propios radioastrónomos.

Por medio de sus factores, la ecuación de Drake recorre desde conceptos astronómicos hasta otros de biología y sociología. En ese recorrido, se mueve desde cuestiones de tipo ordinario hasta las del todo excepcionales. A menudo, los astrónomos se dedican a descubrir cosas tales como la tasa de formación de estrellas en una galaxia similar a la nuestra; los sociólogos, en cambio, es raro que consideren la probabilidad de que una especie inteligente dada desarrolle una civilización. Al ignorar esas diferencias, la ecuación de Drake cae en el clásico error de confundir la capacidad de expresar una cuestión en el lenguaje de la ciencia con la de resolver tal cuestión mediante las prácticas de esa misma ciencia. Lo primero tan sólo requiere pensar; lo segundo exige un conjunto de conocimientos y saber aplicarlos.

Los once científicos reunidos en Green Bank se entusiasmaron de inmediato con la ecuación y dedicaron la mayor parte de los tres días del congreso a asignar valores a sus coeficientes. Estaban de acuerdo en que había alrededor de diez mil millones de estrellas de tipo solar en nuestra galaxia y, dado que ésta tenía una edad de unos diez mil millones de años, parecía razonable pensar en una tasa de formación de estrellas de una por año. Struve estimó que la

mitad de ellas podían poseer planetas. A juicio de Phil Morrison, la fracción debía ser menor, en torno al veinte por ciento. Las conjeturas acerca del número de planetas por sistema que podían ser habitables variaban entre uno (sobre la base de que en el nuestro sólo la Tierra lo es) y cinco (sobre la base de que Marte, Júpiter y algunos de los satélites de los planetas exteriores podrían serlo también). El más joven de los presentes, Carl Sagan, opinaba que la aparición de vida en el planeta adecuado constituiría una certidumbre; la vida surgió en la Tierra a partir de acciones y reacciones físico-químicas predecibles de materiales muy comunes a nivel cósmico y emergería en cualquier otra parte del mismo modo. El reputado biólogo Melvin Calvin compartía esa opinión. La inteligencia se consideraba también más o menos segura; basándose en sus estudios sobre los delfines, John Lilly argumentaba que, en la Tierra, la inteligencia había surgido no una, sino dos veces, aunque —como Morrison señaló— era difícil imaginar a los delfines dedicándose a la radioastronomía.

La segunda sesión comenzó con un terremoto —o, al menos, una cierta conmoción— al llegar a Green Bank la noticia de que Calvin acababa de obtener el Premio Nobel. En un brindis con champán tras el almuerzo, Struve nombró a Calvin delfín honorario. El nombramiento fue extendido a todos los presentes, lo que dio lugar a la fundación de la Orden del Delfín, de la que Elvar —el delfín del investigador jefe John Lilly— se convirtió en mascota, representando así a los cetáceos en el proyecto CETI. Al regresar al trabajo, Calvin

se mostró convencido de que, tarde o temprano, toda criatura inteligente haría uso del espectro electromagnético. No se examinó la cuestión de si el desarrollo de una civilización que desplegara dicha habilidad de una forma tecnológica era imprescindible o no. Se debatió tanto la idea de que algunas de esas criaturas tal vez no desearan comunicarse, como el que esos alienígenas *de clausura* no podrían evitar la fuga al espacio abierto de sus señales de uso doméstico. Finalmente, a f_c se le asignó, de forma provisional, el valor de una décima parte.

El siguiente parámetro era L , la vida probable de una especie tecnológica. Debió de parecerles obvio a todos los de la Orden que el valor de L tenía una inquietante actualidad en el marco de la guerra fría, incluso aunque uno de sus miembros, Phil Morrison, no hubiera sido el encargado de armar la bomba de Nagasaki en Tinian, una remota isla del Pacífico. Todos albergaban temores acerca de un futuro nuclear, por no mencionar otros peligros, tales como la superpoblación y la polución del medio ambiente. Una vida probable de menos de cien años parecía, por desgracia, plausible, aunque no inevitable. Al final, el margen adoptado fue muy amplio: entre mil y cien millones de años. Reuniendo todas las estimaciones

$$(R^* = 1 \text{ a } 10; f_p = 0,5; n_e = 1 \text{ a } 5; f_l = 1; f_i = 1; f_v = 0,1; L = 10^3 \text{ a } 10^8),$$

los *delfinianos* ubicaron N en algún punto entre mil y mil millones, una cifra que satisfizo a todos. La estimación concordaba con la

premisa astronómica —atribuida a Copérnico— de que la Tierra no ocupa lugar especial alguno en el universo. Si la humanidad fuese una civilización más entre un millón de civilizaciones, sólo en nuestra galaxia, ese principio de *mediocridad* se cumpliría puntualmente. La idea de que el hogar del ser humano sea un lugar sin relevancia podrá repugnar a muchos, pero para los astrónomos y los biólogos, acostumbrados a transitar a través de las revoluciones copernicana y darwiniana, otra opción hubiera resultado impropia.

Para algunos —Sagan, en particular— la idea más importante que se derivaba de la ecuación de Drake era que SETI podía contemplarse como un modo de medir la probabilidad de supervivencia de la especie humana. Si SETI tenía éxito gracias a un valor alto de N , esto implicaría también un valor grande para L , lo cual significaría que las civilizaciones tecnológicas no estarían condenadas a la autodestrucción. Barney Oliver, un importante investigador de Hewlett-Packard que participó en la reunión de Green Bank, consideraba prioritario este punto en su informe sobre el «Proyecto Cyclops», publicado en 1971. Cyclops consistía en un costosísimo conjunto de cientos de radiotelescopios dedicados al Proyecto SETI. En la propaganda del proyecto, Oliver hacía unas estimaciones sobre el número de planetas, la tasa de formación de estrellas, el origen de la vida, etc., que se compensaban unas con otras. En el informe Cyclops, la ecuación de Drake quedaba

reducida a $N = L$. Y L «resulta ser ¡el factor más incierto de todos!». ¹⁰²

El argumento no bastó para que Cyclops fuera financiado. (La única manera de obtener la inmensa suma de dinero requerida, como señaló el agudo comentarista sobre política espacial John Pike, hubiera sido distribuir las antenas de tal modo que cada distrito del Congreso poseyera una). Pero L aún era capaz de causar cierto impacto político. En 1978, el senador William Proxmire decidió de pronto retirar la financiación a una propuesta de la NASA para el desarrollo de un programa de radioastronomía SETI mucho menos ambicioso. Sagan, ya famoso entonces, fue a ver al senador —un hombre que Sagan sabía que estaba preocupado por las armas nucleares— y le explicó la ecuación de Drake punto por punto, haciendo un énfasis especial en L . En palabras de Ann Druyan, la esposa de Sagan, la actitud de Proxmire pasó de «en cuanto este sabiondo termine de hablar le echo educadamente del despacho» a «me ha dejado boquiabierto, sorprendido y asombrado», tras lo cual «reconoció su error, afortunadamente». ¹⁰³

Aunque sugerentes y provocadores, esos argumentos basados en L tenían sus fallos. Por una parte, los programas SETI no van a proporcionar el valor de L , simplemente, porque en realidad no tratan de medir N . Todo lo que buscan es un único contacto: nadie está pensando en hacer un censo. Dejando esto aparte, el

¹⁰² *Project Cyclops: A Design Study of a System for Detecting Extraterrestrial Intelligent Life*. El informe sobre un taller de verano en el Ames Research Center, CR114445 (1971).

¹⁰³ *Carl Sagan: A Life*, de Keay Davidson, Wiley, 1999, pág. 348.

argumento de que L tiene cierta relevancia para la humanidad asume implícitamente que las vidas de las civilizaciones se hallan uniformemente distribuidas. Pero ¿y si no fuera así? De hecho, esto es lo que los asistentes al Congreso de Green Bank temían y lo que les llevó a proponer una gama tan amplia de valores para L . Sebastian von Hoerner, uno de los radioastrónomos de Green Bank, señaló que bastaría con que unas pocas civilizaciones duraran miles de millones de años para que L , la vida media, tuviese un valor alto aunque la mayoría de las civilizaciones se autoinmolaran en holocaustos nucleares. Si el uno por ciento de las civilizaciones tecnológicas durara mil millones de años y el otro noventa y nueve por ciento, sólo un siglo, L valdría nada menos que diez millones de años, aunque la probabilidad de extinción para cualquier civilización en el umbral nuclear fuera abrumadoramente grande. Podemos apostar a que Sagan no se detuvo en estas minucias en su entrevista con Proxmire.

La relevancia de SETI en relación con el fin del mundo no se limitaba a cálculos cuestionables sobre su proximidad inminente. El primer libro popular acerca del Congreso de Green Bank y sus ideas, *We are not alone*, del periodista del *New York Times* Walter Sullivan,¹⁰⁴ está dedicado a «quienes, desde cualquier lugar, tratan de que “ L ” sea grande». Y eso era parte de lo que los *delfinianos* confiaban en lograr de varias maneras. La más obvia de ellas era establecer contacto. Si L era superior a algunas décadas, la mayoría

¹⁰⁴ Walter Sullivan, *We Are Not Alone*, McGraw-Hill, 1964.

de las civilizaciones tecnológicas serían más antiguas que la de la Tierra. En este caso, si se establecía contacto, sería, casi con toda seguridad, con una civilización más antigua y sabia, tanto en tecnología como en ética. Más sabia en tecnología, simplemente, porque habría dispuesto de más tiempo para desarrollar cosas. Más sabia en el aspecto ético porque habría sobrevivido a los peligros de una era nuclear erradicando de algún modo las causas de los conflictos. De este modo, mientras la cosmología de un universo en expansión convertía los telescopios en máquinas para ver el pasado, SETI hacía de ellos instrumentos para observar el futuro. Tal como señalaba A. G. W. Cameron, uno de los primeros promotores de la idea, «Si logramos dar el paso siguiente y nos comunicamos con una de esas sociedades, cabe esperar obtener un inmenso enriquecimiento en todas las áreas de nuestras ciencias y nuestras artes, y tal vez recibiríamos también valiosas lecciones sobre cómo gobernar el mundo».¹⁰⁵ Aunque los alienígenas no nos aportaran ningún beneficio práctico, su mera presencia debería hacer que los pueblos de la Tierra se sintieran unidos. Los alienígenas no constituirían una amenaza contra la que unirse —las primeras discusiones en SETI fueron unánimes al respecto, basándose en que ninguna civilización belicosa sobreviviría mucho tiempo tras desarrollar tecnologías nucleares—, pero su majestuosa y profunda alteridad nos haría olvidar nuestras insignificantes disputas.

¹⁰⁵ Introducción de *Interstellar Communication*, en: A. G. W. Cameron, ed., Benjamin, 1963; citado en *The Biological Universe*, de Steven Dick, pág. 508. El libro de Dick constituye un excelente informe histórico sobre SETI y otras disciplinas relacionadas.

El escritor Ed Regis subrayaba que los proponentes de SETI — Sagan, en particular— empleaban argumentos como el anterior de una forma poco sólida y aparentemente contradictoria. Por un lado, afirmaban que encontrar vida por ahí nos haría olvidar nuestras diferencias y, por ello, ser menos propensos a exterminarnos nosotros mismos. Y por otro, que si finalmente no se hallaba vida fuera de nuestro planeta, esto nos llevaría a apreciar más profundamente nuestro valor y a renunciar a las guerras. Como decía Regis, ninguno de los dos argumentos resulta verosímil. ¿Se imaginan a LBJ apartando el dedo del botón y explicando a sus ayudantes: «Si hubiera otras especies inteligentes ahí afuera, pulsaría el botón sin dudarlo, pero, como estamos solos en el universo, no sé si debo...»? El hecho de que Sagan llegara a la conclusión de que la humanidad debía renunciar a las armas de destrucción masiva —tanto si el proyecto SETI tenía éxito como si no— sugiere que dicha conclusión fue en realidad un punto de partida (y, en efecto, lo fue).

La relación entre SETI y la salvación iba más allá de la mera retórica, sin embargo. Para Sagan, SETI no era un simple instrumento para medir la probabilidad de supervivencia. Era una forma de promover el tipo de valores que podían incrementar esa probabilidad: la racionalidad científica y la cooperación internacional. Para que una civilización pudiera sobrevivir a la crisis nuclear, debería estar unida y ver más allá. Debería ser consciente de su lugar en el cosmos y mirar lejos en el espacio y el tiempo. El

proyecto SETI, concebido como una empresa en la que soviéticos y estadounidenses podían trabajar juntos, no sólo era un medio para detectar dichas civilizaciones, sino para intentar construir una de ellas.

En este sentido, el SETI de la década de 1960 guarda una notable semejanza con el mundo de *Star Trek*. A diferencia de la mayor parte de la ciencia ficción escrita, *Star Trek* albergaba una fuerte carga utópica. Combinando la retórica kennedyana de las «nuevas fronteras» con la «frontera sin límites» del famoso informe de Vannevar Bush sobre ciencia y gobierno, publicado en 1946, la «frontera final» de *Star Trek* proyectaba un futuro sustentado en la ciencia. Tal como Constance Penley indica en su estudio sobre la influencia de la serie televisiva, *NASA/TREK*, el programa contribuyó al proceso por el cual «Viajar al espacio se convirtió en la principal metáfora mediante la que intentábamos comprender el mundo de la ciencia y la tecnología, e imaginábamos un lugar en él para cada uno de nosotros». ¹⁰⁶ Frente al temor a la guerra nuclear, la superpoblación, el agotamiento de los recursos, la contaminación, etc., los científicos de SETI trataban de dar sentido a la ciencia y la tecnología, y «viajar al espacio» —con la mirada del astrónomo y no físicamente— era su modo de hacerlo.

Como la Orden del Delfín, la tripulación de la nave *Enterprise* «buscaba nuevas civilizaciones» y encamaba una de ellas. En el puente de mando había un ruso y un africano, además de varios

¹⁰⁶ Constance Penley, *NASA/TREK: Popular Science and Sex in America*, Verso, 1997.

norteamericanos y un nativo de Vulcano. Según el creador de la serie, Gene Roddenberry, «el planteamiento constituía el “mensaje” de la serie: debemos aprender a vivir juntos si no queremos perecer juntos». ¹⁰⁷

El papel de los soviéticos en SETI era bastante más sustancial que el del alférez Chekov en la *Enterprise* (el cual consistía básicamente en atraer jovencitas con debilidad por tipos como el David Jones de *The Monkees*). En los años sesenta, los científicos soviéticos también mostraban interés en escuchar voces del más allá. Mientras, en Estados Unidos, SETI era poco más que un foro de discusión —la única investigación en marcha era la de Drake—, en la Unión Soviética se dedicaban a escudriñar sistemáticamente el cielo con sus antenas (aunque esas antenas fueran poca cosa). De forma aún más explícita que en Estados Unidos, las civilizaciones colonizadoras del espacio eran consideradas en la Unión Soviética como una etapa obligada en una progresión histórica, en una utopía en la que la Unión Soviética lideraría a los pueblos de la Tierra hacia el futuro con sus *Sputniks*. Según las influyentes teorías de Nicolai Kardashev, la Tierra se estaba convirtiendo en una civilización de «tipo I», en la que una especie ha llegado a controlar los recursos energéticos de todo un planeta. En un futuro lejano se

¹⁰⁷ Stephen E. Whitfield y Gene Roddenberry, *The Making of Star Trek*, Ballantine/Del Rey, 1968, pág. 112. En el mismo libro, Roddenberry describe cómo introdujo, al principio del guión original de la serie, un cálculo del número de planetas habitables que podían ser explorados (presumiblemente, tras haber oído hablar de la ecuación de Drake). Por desgracia, no disponía de material de investigación alguno, con lo que improvisó tanto la fórmula, $F^2(MgE) - c^1 R^1 \times M = L/So^*$, como el enorme número de planetas que se deducían de ella. Ninguno de los responsables de la cadena de televisión cuestionó este punto.

hallaba la posibilidad de que se transformara en una civilización de «tipo II», que controlaría toda la producción de energía de una estrella, o en una de «tipo III», que tomaría el control sobre toda una galaxia. Aunque las civilizaciones de tipo I serían las más comunes, las más fáciles de detectar serían las de los tipos II y III, miles de veces más escasas, pero millones de veces más brillantes. Freeman Dyson, un físico y matemático de Princeton, había especulado sobre civilizaciones que usaran la mayor parte de la energía de una estrella y sugerido que sus residuos térmicos podrían ser detectables mediante satélites sensibles al infrarrojo. Las tecnologías de tipo II que abarcaran una estrella fueron bautizadas como «esferas de Dyson», aunque el propio Dyson reconoció después que la idea se la había sugerido la novela *El hacedor de estrellas* (*Star Maker*), de Olaf Stapledon, en la que, en un futuro remoto, «cada sistema solar estaría rodeado de una red de receptores de luz que captarían la energía solar emitida para destinarla a un uso inteligente, con lo que la galaxia en su conjunto se vería oscurecida».¹⁰⁸ Stapledon plasmaba así tanto los anhelos espirituales de SETI como sus especulaciones tecnológicas. Tras la búsqueda de esa inteligencia que retrataba *El hacedor de estrellas* «no había una simple ansia de observación científica, sino también la necesidad de desarrollar algún tipo de tráfico mental y espiritual con otros mundos, orientado al enriquecimiento mutuo».

¹⁰⁸ *Star Maker*, de Olaf Stapledon (1937). [Trad, esp.: *El hacedor de estrellas*, Barcelona, Minotauro, 1997].

Convencido de que SETI debía ser verdaderamente global, y de que había que asegurar la emergencia de una civilización de tipo I, Sagan dedicó parte de sus inagotables energías a organizar una secuela del Congreso de Green Bank en el este. Él y Kardashev convocaron una reunión en el observatorio de Byurakan, Armenia, en 1971. El contingente americano incluía a veteranos de Green Bank y a algún nuevo talento, entre ellos Francis Crick, Thomas Gold, William McNeil y Marvin Minsky (quien llevó consigo algunos *frisbees*¹⁰⁹ algo nunca visto en Armenia y que causó sensación). Se les unieron unos treinta científicos soviéticos, incluyendo el colega de Kardashev Iosif Shklovskii, quien con ocasión del quinto aniversario del primer *Sputnik* había publicado un libro titulado *Vselennaia, Zhizn, Razum (Universo, Vida, Mente)*. Sagan se las arregló para que fuera traducido al inglés y añadió una buena cantidad de material propio. El volumen resultante, *Intelligent Life in the Universe*, que rendía homenaje a la ecuación de Drake, fue el primer tratado científico de envergadura sobre el tema y se convirtió, en palabras del historiador Steven J. Dick, en «la biblia del movimiento SETI».¹¹⁰

En Byurakan, la ecuación de Drake volvió a ser el punto de partida. Crick no estaba tan de acuerdo como los astrónomos de Green Bank con la afirmación de Sagan de que la vida era inevitable. Los últimos parámetros de la ecuación, como siempre, fueron los que desencadenaron las mayores controversias. El neurólogo David

¹⁰⁹ Discos voladores [N. del T.]

¹¹⁰ Iosif S. Shklovskii y Carl Sagan, *Intelligent Life in the Universe*, Holden-Day, 1966.

Hubel admitía tras el almuerzo que, incluso antes de haberse atiborrado de coñac armenio, no tenía ni idea de por qué ciertas criaturas desarrollaban inteligencia, mientras que otras parecían apañárselas perfectamente sin ella. El antropólogo y etnógrafo Richard Lee opinaba que, para una civilización tecnológica, el lenguaje resultaba crucial, pero no era suficiente, tal como demostraba la vida del pueblo bosquimano, culturalmente rica pero tecnológicamente atrasada. (Lee, por cierto, acabaría brindando en la cena en lenguaje Kung). A pesar de todas las discrepancias, se adoptó para N el valor de un millón, el mismo que Shklovskii y Sagan proponían en su libro.

Aunque la ecuación de Drake fuera el punto de partida del congreso, durante la semana que éste se prolongó los participantes divagaron tanto en sus discusiones que el término «amplio espectro» aplicado a ellas se quedaría muy corto. Por la reunión pasó una extraordinaria selección de las fantasías y los temores de unos científicos en libertad total para especular: enormes máquinas horadando la Tierra a 60 kilómetros bajo la superficie, vida en las estrellas de neutrones, inteligencias artificiales, universos atrapados en el seno de partículas elementales, nuevas leyes físicas, efecto de las manchas solares sobre la creatividad, taquiones, «autodestrucción genética de la razón» (es decir, reproducción ilimitada de los imbéciles), la futura edad de oro, nanotecnologías, agujeros negros, agujeros blancos, cohetes de antimateria, etc. El

esforzado estenotipista Floy Swanson dejó testimonio de todo ello para la posteridad.¹¹¹

A lo largo de la primera década de estudios SETI, los radioastrónomos consiguieron mantener bajo control los términos del debate que ellos mismos habían iniciado. No es que no hubiera otras ideas que fueran aún más lejos; Byurakan había sido un ejemplo. Pero, como Phil Morrison argumentara allí, añadir al debate alternativas imposibles de comprobar no iba a ser de gran ayuda. SETI comenzó como un proyecto de radioastronomía debido a que a Morrison, Cocconi y Drake les pareció que la radioastronomía proporcionaba el único medio de que una civilización como la nuestra pudiera ser detectada por otra civilización como la nuestra. Los alienígenas que no fueran fuentes potentes de radio simplemente serían ignorados —por ser nubes inteligentes de gas o sibaritas en una edad de oro posttecnológica o criaturas de materia nuclear que pululan en la superficie de una estrella de neutrones—. Los aspectos tecnológicos de SETI fueron, pues, los asociados a la radioastronomía y, en particular, la elección de las frecuencias adecuadas (frecuencias que pudieran ser consideradas *naturales* por los radioastrónomos de cualquier época y especie) y las estrategias de señalización. La diferencia entre las civilizaciones de tipo I y tipo II, por ejemplo, radicaría básicamente en la potencia de sus radiotransmisores y en la estrategia que las

¹¹¹ Carl Sagan, ed., *Communication with Extraterrestrial Intelligence*, MIT Press, 1973. Quienes no deseen abordar el tema en su totalidad disponen de un excelente informe en *Carl Sagan: A Life in the Cosmos*, de William Poundstone (Henry Holt, 1999).

hiciera potencialmente detectables. (En la época del Congreso de Byurakan, Drake estaba cada vez más interesado por las civilizaciones de tipo II: las supercivilizaciones. En 1974, Sagan y él iniciaron una búsqueda de ellas. En vez del radiotelescopio de 26 metros de Green Bank en el que Drake realizó su primera búsqueda, emplearon el recién estrenado radiotelescopio de Arecibo, una enorme parábola ubicada en las montañas de Puerto Rico. En lugar de investigar estrellas individuales, exploraron galaxias enteras, muestreando simultáneamente las emisiones de miles de millones de estrellas lejanas y estimando que el reducido número de posibles supercivilizaciones se vería compensado por la potencia de sus radiofaros).

El hecho de que los radioastrónomos fueran las únicas personas que pudieran hacer algo más que hablar sobre civilizaciones más allá del sistema solar no significó que los demás interesados se limitaran a dejar el asunto en sus manos. Muchos biólogos opinaban, por una parte, que el origen de la vida no estaba tan claro como Sagan afirmaba y, por otra, que incluso emergiendo la vida, la probabilidad de que la evolución diese lugar a seres como nosotros era increíblemente pequeña. La versión más influyente de este argumento era «el no predominio de los humanoides», debida al eminente teórico evolucionista George Gaylord Simpson.¹¹² La evolución, según Simpson, era contingente y no determinista. No tenía interés alguno en producir al hombre y si lo había hecho era

¹¹² George Gaylord Simpson, «The nonprevalence of humanoids», *Science* 1964, 143, pág. 769 y sigs.; reeditado por Goldsmith, *op. cit.*

por casualidad. No daría lugar otra vez al ser humano, ya que nunca había producido una misma cosa dos veces. Y si jamás generaría de nuevo al hombre en la Tierra, ¿por qué iba a hacerlo en otra parte? La respuesta de Sagan era que, aunque la historia evolutiva específica del ser humano fuera improbable, habría muchas historias evolutivas posibles que condujeran a criaturas tan inteligentes como el hombre. Y aunque cada una de esas historias resultase en sí misma altamente improbable, si su número fuera suficientemente grande, la probabilidad de que una de ellas se diera sería bastante elevada.

La respuesta a la pregunta de cuál sería esa probabilidad —o lo que es lo mismo, de cuántas historias evolutivas podrían converger hacia la inteligencia— se convirtió en una cuestión de gustos. Algunos evolucionistas, como Stephen Jay Gould, opinaban que, aunque la evolución no se repita ciertamente en los detalles, la inteligencia podría ser una de esas cosas que la evolución inventa una y otra vez, como el ojo o el ala.¹¹³ Si así fuera, sería razonable pensar que podría surgir también en otros planetas, con lo que SETI tendría sentido. Otros expertos en evolución, como Jared Diamond, pensaban que, si bien la inteligencia de tipo humano —capaz de construir radiotransmisores— posee claramente cierta utilidad, esto no significa que tenga que evolucionar repetidamente. Picotear los árboles, señalaba Diamond, es una estrategia ecológica muy exitosa, pero de muy difícil evolución. Sólo una especie de criaturas en toda

¹¹³ Stephen Jay Gould, «SETI and the wisdom of Casey Stengel», *The Flamingo's Smile*, W. W. Norton, 1984. [Trad. esp.: *La sonrisa del flamenco*, Barcelona, Crítica, 1995]

la historia de la Tierra —el pájaro carpintero— ha hecho uso de ella. Si no hubiera surgido el pájaro carpintero, no hay razón para creer que su nicho habría sido ocupado; allí donde nunca hubo pájaros no tendría por qué haberlos. En la opinión de Diamond, la clase de inteligencia que buscaba SETI se parecía más a la estrategia de picotear árboles que a la de ver o volar.¹¹⁴

Después de Simpson, hubo que aceptar que, aunque fuera lícito mantener una opinión concreta al respecto, no había un modo real de estimar la frecuencia de la inteligencia. Esta incertidumbre hacía que la ecuación de Drake resultara un tanto problemática. En rigor, multiplicar factores como f_i y f_c implica multiplicar también las incertidumbres con las que estimamos dichos factores. Las incertidumbres en las estimaciones realizadas en Green Bank para R^* (entre uno y diez) y L (entre mil y diez millones) se traducían en una gran incertidumbre en el resultado final, que podía oscilar entre mil y mil millones. Y, tal como Francis Crick subrayó en Byurakan, las incertidumbres a la hora de estimar cosas como la probabilidad de emergencia de la vida eran enormes. Podía ser algo seguro o una posibilidad en mil billones. No había forma de saberlo. Y esa incertidumbre se trasladaba ineludiblemente a la estimación final de N . El matemático Alfred Adler criticó duramente este punto en su cáustico informe para el *Atlantic* sobre las actas del congreso.¹¹⁵

¹¹⁴ Jared Diamond, «Alone in a crowded universe», *Natural History*, 1990; reeditado en *Extraterrestrials: Where are They?*, en: Ben Zuckerman y Michael H. Hart, eds., 2.ª ed., Cambridge University Press, 1995.

¹¹⁵ Alfred Adler, «Behold the stars», *Atlantic Monthly* 1974, 234, pág. 109 y sigs.; reeditado por Goldsmith, *op. cit.*

Adler cita a Crick («No es posible [...] hacer estimación razonable alguna del factor f_i »), L. M. Mukhin («No entiendo cómo podríamos estimar f_i ») y Sagan («Nos enfrentamos a [...] graves dificultades para extrapolar [...] en el caso de L , del que no hay ejemplo alguno»), y concluye que «el propósito de la conferencia, realizar una estimación del número N , fue un completo fiasco».

«Es increíble», continúa Adler, «que científicos distinguidos como los que había entre los asistentes (y, de hecho, había unos cuantos) participasen de buen grado, y hasta con entusiasmo, en una parodia de todo lo que quienes aman y valoran la ciencia y el intelecto consideran con respeto». Adler culpaba de esta actitud a la perniciosa influencia del estilo intelectual del joven Sagan:

«El moderno tecnólogo es un imbécil con talento; bien entrenado y oportunista, aunque poco imaginativo y sin gracia. [...] Cabalga en las sutilezas y profundidades que los prudentes apenas se atreven a recorrer de puntillas e invade terrenos de los que no conoce nada. [...] El hombre civilizado y racional ya está cansado, harto de soportar las embestidas de estos jóvenes maestros que atraen grandes cantidades de dinero e influencias, les dejan sin aliento en los congresos, lanzan un torbellino de vaguedades y agotan a sus debilitados mayores con una prepotencia natural, ostensible y confiada».

Aunque Adler era especialmente severo, el resentimiento hacia la arrogancia y la financiación obtenida por los científicos del espacio era un denominador común en muchas de las críticas contra SETI y otros proyectos afines en el recién nacido campo de la exobiología.

Adler tenía razón en que la presencia de grandes incertidumbres hace que, en sentido estricto, la ecuación de Drake pierda todo significado. Pero limitar la crítica a ese terreno supone ignorar su verdadero propósito. La ecuación de Drake era una forma de decir que el universo produce civilizaciones del mismo modo que produce estrellas y planetas y que las herramientas que detectan estos últimos, en realidad, también podrían detectar civilizaciones. Todos los implicados eran conscientes de las limitaciones de la ecuación y algunos se hallaban muy molestos con las especulaciones sociológicas a que había dado lugar: «Al diablo con la filosofía. He venido aquí a aprender sobre observaciones e instrumentos», decía Dyson en el Congreso de Byurakan. También eran conscientes de que la ecuación representaba no una respuesta, sino un tipo de aproximación. Para derribarla no bastaba con demostrar que nunca podría proporcionar respuestas; había que demostrar que el enfoque que encarnaba jamás las encontraría. El razonamiento de que el universo nunca habría producido civilizaciones con las que contactar representaba un argumento contra SETI mucho más poderoso que cualquier crítica sobre su metodología.

Ese argumento aparentemente potente surgió a mediados de la década de 1970, cuando a la visión del universo por parte de los

astrónomos se añadieron las ideas de los vuelos espaciales. Una de las premisas básicas de SETI era que las civilizaciones permanecían en los sistemas estelares en los que habían evolucionado, comportándose como fuentes puntuales esparcidas por el cielo y titilando durante el tiempo L que les correspondiera. Pero ¿y si las civilizaciones se movían? Si los alienígenas podían viajar de estrella en estrella, aunque lo hicieran muy despacio, el tiempo que les llevaría visitar todos los sistemas solares de la galaxia sería solamente unos cien millones de años. En este caso y suponiendo que la galaxia tuviera una edad de unos diez mil millones de años, cabría esperar que ya nos hubieran encontrado. Muchos remontan este enfoque a una pregunta lanzada por Enrico Fermi, el gran físico experimental constructor del primer reactor nuclear, en un almuerzo en Los Alamos, allá por 1950. Según Edward Teller, Fermi «se descolgó de pronto con una pregunta totalmente inesperada: “¿Dónde diablos están?”. Hubo una carcajada general porque, curiosamente, aunque la pregunta de Fermi era por completo inocente, todos alrededor de la mesa parecieron entender a la vez que se refería a la vida extraterrestre». (En realidad, había habido una conversación previa sobre los extraterrestres, al hilo de una tira cómica publicada en *The New Yorker*, según la cual la ausencia de cubos de basura en la ciudad se debía a que los platillos volantes se los estaban llevando; a Fermi le pareció una buena teoría, pues

explicaba a la vez la presencia de platillos volantes y la falta de cubos de basura).¹¹⁶

Los fundadores de SETI interpretaron esta «paradoja de Fermi» como una evidencia de que los viajes espaciales interestelares eran imposibles en la práctica. En un estudio sobre «los límites del viaje a través del espacio», publicado el año anterior al de la reunión de Green Bank, Sebastian von Hoerner escribía: «Personalmente, llego a esta conclusión: los viajes por el espacio, incluso en un futuro remoto, se limitarán a nuestro sistema planetario y una conclusión parecida sería aplicable a cualquier otra civilización, por muy avanzada que ésta fuese. El único medio de comunicación entre distintas civilizaciones, por lo tanto, serían las ondas electromagnéticas».¹¹⁷ De los *delfinianos* originales, sólo el profesor de Stanford Ron Bracewell consideraba el contacto directo como la forma ideal de que las civilizaciones antiguas se comunicaran con otras más jóvenes e imaginaba sondas enviadas por civilizaciones avanzadas a sistemas estelares cercanos para buscar vida inteligente y, en caso de encontrarla, conversar con ella. Frente al argumento de que lanzar estas sondas resultaría prohibitivamente caro, Bracewell insistía en que las sondas serían pequeñas, «del tamaño de un balón de fútbol [...], con una inteligencia equivalente a la de un ser humano encapsulada en un objeto del tamaño de una

¹¹⁶ Eric Jones, «Where is everybody» [carta], *Physics Today*, 1985. Recoge varios relatos, de Teller y otros, sobre el famoso almuerzo, además de la tira cómica instigadora.

¹¹⁷ Von Hoerner, «The general limits of space travel», *Science*, 1962, 137: pág. 18 y sigs.; reeditado por Goldsmith, *op. cit.*

cabeza humana»,¹¹⁸ y que una sola podría visitar varias estrellas si lograba sobrevivir lo suficiente. Aunque fuese tan razonable como muchas otras de las especulaciones en circulación, la idea de Bracewell se alejaba demasiado de la premisa básica —«las civilizaciones son como estrellas»— como para formar parte del punto de vista general en SETI.

La idea de Bracewell presentaba, además, otro problema: situaba a los alienígenas demasiado cerca. Hasta los más imaginativos se abstenían de pensar en civilizaciones extraterrestres que atraviesan los límites de nuestro sistema solar. Cuando, en el Congreso de Byurakan, Dyson expuso su opinión de que el hábitat más grande y apetecible de la galaxia no eran las superficies de los planetas, sino las de los cometas, y de que la vida podría esparcirse poco a poco de una nube cometaria a otra —sin llegar a estar nunca cerca de una estrella caliente que la perturbara—, Thomas Gold le preguntó de inmediato si la vida procedente de alguna parte estaría ya presente en la nube de cometas que giran alrededor de nuestro Sol. Hasta un pensador tan ágil como Dyson tuvo que reconocer que no había considerado esa posibilidad.

Bracewell continuó divagando lejos de SETI y especulando a la manera de Fermi sobre si la primera civilización que se lanzara al espacio acabaría dominando el universo. Pero la idea quedó «escondida en una o dos páginas [de un libro que escribió en 1974], debido a que al editor le pareció mejor transmitir la idea de una

¹¹⁸ Entrevista realizada por David W. Swift y publicada en *SETI Pioneers* (University of Arizona Press, 1990); una excelente fuente oral.

inteligencia que se extendía que la de un pensamiento único y colonizador». ¹¹⁹ Un año después, sin embargo, el físico Michael Hart proponía una especie de variante del argumento de Fermi. ¹²⁰ Hart partía de lo que denominaba el *hecho A*: «En la actualidad no hay en la Tierra seres inteligentes que procedan del espacio exterior».

Hart imaginaba cuatro clases de explicaciones para este hecho. La primera era física —el viaje interestelar es demasiado duro—. Pero unos alienígenas longevos o que viajaran hibernados o en naves-ciudad autosuficientes, o unos robots, no tendrían por qué desanimarse ante un viaje de siglos de duración, ya que habrían de hacerlo a una velocidad realista por debajo de la de la luz. La segunda clase de explicación era *sociológica*: a los alienígenas no les interesaría la exploración, sus civilizaciones acabarían estancándose o autodestruyéndose o no desearían perturbar el desarrollo de sociedades primitivas como la nuestra. Pero Hart argumentaba que para explicar la ausencia de extraterrestres en nuestro planeta, dichos razonamientos debían ser aplicables a todas las civilizaciones alienígenas; si tan sólo una de ellas fuera expansionista —en el sentido que imaginamos para nosotros mismos— ya habría llegado hasta aquí. Hart descartaba las explicaciones temporales —el que no hubieran tenido tiempo de llegar hasta la Tierra— sobre la base de que el tiempo necesario para colonizar una galaxia sería despreciable comparado con la edad de la galaxia o los tiempos

¹¹⁹ Bracewell en Swift, *op. cit.*

¹²⁰ Michael Hart, «An explanation for the absence of extraterrestrials on Earth», *Quarterly Journal of the RAS*, 1975, 16, pág. 128 y sigs.; reeditado por Goldsmith, *op. cit.*

evolutivos. La cuarta clase de explicación —el que *ya* nos hubiesen visitado— hacía uso de argumentos similares. Pero si nos hubieran visitado hace poco, ¿por qué se habrían ido tan pronto? Y si hubieran llegado hace mucho, ¿por qué no seguirían aquí? Hart concluía que «la idea de que miles de civilizaciones avanzadas se hallen esparcidas por la galaxia es muy poco verosímil a la luz del hecho A. [...] Tal vez nuestros descendientes acaben encontrando algunas [...], pero su número debería ser muy pequeño y hasta podría ser nulo».

Algunos años más tarde, el físico Frank Tipler empleó argumentos similares de una forma menos matizada, pero particularmente llamativa.¹²¹ Tipler argumentaba que era plausible que las civilizaciones avanzadas construyeran «máquinas de Von Neumann» —robots autorreproductores, preparados también en este caso para el viaje interestelar— y las enviaran por toda la galaxia. Si así fuera, la evidencia de estas máquinas sería amplia e inconfundible, ya que en unas decenas de millones de años se habrían extendido por todos los sistemas solares. Como no vemos máquinas de Von Neumann por ninguna parte, no existen dichas civilizaciones. Tipler se tomó el argumento lo suficientemente en serio como para hacer campaña activa contra el *despilfarro* de dinero en SETI, dado que, según sus razonamientos, se trataba de un fraude. Por el contrario, Hart integró sus críticas en el lenguaje de la comunidad SETI, convirtiendo su *hecho A* en un indicio de que *N* era muy pequeño y

¹²¹ Frank Tipler, «Extraterrestrial intelligent beings do not exist», *Quarterly Journal of the RAS*, 1981, 21, pág. 267 y sigs.

de que la ecuación de Drake debía ser revisada. Según él, la causa estaba en la probabilidad de existencia de planetas habitables, que era muy baja. En un simposio dedicado a SETI durante el Congreso de la Unión Astronómica Internacional celebrado en Montreal en 1979,¹²² Hart presentó una gama de valores para los factores de la ecuación de Drake y demostró que, al adoptar supuestos simplemente moderados o ligeramente pesimistas, N se hacía igual a uno o incluso más pequeño.¹²³

En realidad, nunca antes se había cuestionado el valor de N ; se aceptaba que estaba entre diez mil y diez millones y nadie se había decantado por un valor concreto. En tanto fuera un número similar en tamaño a los que representaban cantidades de galaxias, nebulosas o fuentes de radio en un catálogo astronómico típico, servía para que SETI pareciese una rama astronómica más. La magnitud del impacto causado en SETI por Tipler y Hart queda reflejada en el hecho de que en el simposio de 1979 hubo seis artículos que defendían otras tantas gamas de valores muy diferentes para N y a los cuales se dio precedencia. Los argumentos en la línea del «¿Dónde diablos están?» sacudieron al propio núcleo de SETI como nunca lo habían hecho los razonamientos evolutivos, ganando muchos adeptos. El coautor de Sagan, Shklovskii, fue uno de ellos. Otro fue Von Hoerner, quien sintetizó el tema en la incisiva observación de que a los humanos les gustaría, sin duda, explorar y

¹²² Las actas se hallan en *Strategies for the Search for Life in the Universe*, en: Michael Paggiannis, ed., Reidel, 1980.

¹²³ Por ejemplo, $R^* = 20$, $f_p = 1/40$, $n_e = 1/1.000$, $f_l = 1/10$, $f_i = 1/10$, $f_c = 1/2$ y $L = 10.000$ dan $N = 0,025$.

colonizar y, por lo tanto, «si fuésemos típicos, no deberíamos existir». SETI había partido de la idea de que tenía que haber otros usuarios de ondas de radio más o menos como nosotros esparcidos por el cielo, de que podíamos apuntar hacia las estrellas y decir: «El hombre es algo similar a *eso*». Si nuestro arquetipo humano era un explorador y no un radioastrónomo, la paradoja de Fermi nos había agitado la fiesta.

En la década de 1970, los viajes espaciales parecían al alcance de la mano y la humanidad, una raza de exploradores espaciales en potencia. Como escribía Hart, «Tras el éxito del *Apolo 11*, resultaba extraño escuchar a alguien decir que los viajes espaciales eran imposibles». La NASA había lanzado ya cuatro vehículos interestelares en la época en la que Tipler hizo sus contribuciones; las sondas *Pioneer 10*, *Pioneer 11*, *Voyager 1* y *Voyager 2* estaban a punto de abandonar el sistema solar. Otra explicación de por qué la paradoja de Fermi llevó a revisar a la baja las estimaciones sobre N en ese momento fue que el crecimiento exponencial de la población se acababa de convertir en motivo de preocupación. A la bomba atómica se unía la *bomba poblacional*; el Club de Roma había llegado a la conclusión de que el mundo estaba agotando rápidamente los recursos no renovables. Para la clase de gente que era consciente de estos temas y se tomaba el proyecto SETI en serio, la expansión hacia el espacio era un paso necesario y, por lo tanto, imprescindible también para cualquier otra civilización avanzada; tarde o temprano, la escasez de recursos conduciría a la expansión

a través de la galaxia. Finalmente, un N pequeño ponía a SETI en línea con el estado de ánimo del vecino y emergente campo conocido como exobiología. La década inicial de SETI, los años sesenta, había coincidido con los primeros planes para buscar vida en otros planetas; Drake estuvo presente en una de las primeras reuniones de esa nueva comunidad exobiológica y Sagan era su más ardiente defensor. A finales de los años setenta, sin embargo, la exobiología se hallaba en horas bajas. La misión *Viking* a Marte, cuyo objetivo principal era la búsqueda de vida, no había encontrado nada de interés exobiológico. La vida en el sistema solar parecía confinada a la Tierra, una impresión reforzada por las imágenes de nuestro planeta enviadas por los astronautas del proyecto Apolo (verdaderos iconos de la fragilidad de la vida en la inerte inmensidad del cosmos). En un universo mayoritariamente estéril, la soledad cósmica parecía verosímil, nos gustase o no.

Los partidarios de un N alto disponían de varios contraargumentos, aunque ninguno fuera convincente del todo. Hart y sus colegas podían estar subestimando las dificultades del viaje interestelar. Con las matemáticas usadas para explicar las plagas de ratas, Sagan subrayaba que la expansión de las civilizaciones a través de la galaxia podía ser mucho más lenta que la velocidad de sus naves espaciales, debido a que el tiempo necesario para ocupar y consolidar los mundos recién visitados sería mucho mayor que el que les llevaría alcanzar el siguiente sistema solar. Las civilizaciones avanzadas tal vez encontrarán sus interacciones con otras mucho

más interesantes —o más destructivas— que cualquier otra actividad, lo que podría también frenar su expansión. Finalmente, nadie construiría máquinas de Von Neumann porque podrían quedar fuera de control y devastar toda la galaxia.¹²⁴

No obstante, el contraargumento de más peso fue ignorado. Se trataba de que el *hecho A* de Hart no era tal; la ausencia de extraterrestres no necesitaba explicación, sencillamente porque no estaban ausentes. Hart asimilaba la creencia en la presencia extraterrestre con que la observación de OVNI fuera una evidencia de tal presencia, y rechazaba la idea con el argumento poco convincente de que «dado que pocos astrónomos creen en la hipótesis OVNI, juzgo innecesario analizar mis propias razones para descartarla». Pero la creencia de que existen alienígenas de incógnito en nuestro sistema solar, o incluso en la Tierra, no parece más descabellada que muchas otras de las hipótesis que formaban parte del debate SETI. Algunos se la tomaron lo suficientemente en serio como para sugerir la búsqueda de residuos industriales en el cinturón de asteroides. Pero la idea de una presencia más cercana que la del ámbito astronómico era algo que pocos estaban dispuestos a aceptar.

En parte, esto se debía sin duda a que, en la práctica, era muy difícil distinguir dicha idea de las asociadas a la teoría OVNI, las cuales se consideraban poco científicas, a diferencia de las que

¹²⁴ Tal vez el informe más amplio y sistemático sobre las posibilidades sea «The "Great Silence": the controversy concerning extraterrestrial intelligent life», de Brin, publicado en *Quarterly Journal of the RAS*, 1983, 24, pág. 283 y sigs.

formaban parte del proyecto SETI. Otra razón para no aceptar extraterrestres cerca podría ser el poder mítico de la mirada del astrónomo. Si hubiera naves espaciales alienígenas por aquí, ya habrían sido detectadas por nuestros telescopios. En realidad, no tendría por qué ser así. Hay una enorme cantidad de sistemas solares que nunca hemos escudriñado. Existen millones de asteroides no clasificados de tamaño suficiente como para ser naves espaciales interestelares. Pero asumimos que no lo son porque, de lo contrario, ya nos habríamos dado cuenta. Acostumbrados a la mirada del astrónomo, asumimos que todo el universo visible está a la vista, aunque la mayor parte de él nunca haya sido inspeccionado. Haciendo una analogía con la vida cotidiana, asumimos que el primer plano astronómico se ve más fácilmente que el fondo, aunque éste no sea el caso. Al igual que las civilizaciones lejanas de tipo II quedarían eclipsadas por otras más próximas de tipo I, todo lo que vemos en el cielo, por muy distante que esté, es mucho más brillante que un trozo de roca cercano (o una nave espacial).

Hay una objeción más profunda a la idea de que el *hecho A* sea falso. En presencia de inteligencias indetectables, la propia ciencia hace aguas: no se puede establecer claramente qué se entiende por natural. El sujeto de SETI fue siempre una curiosa mezcla de lo natural y lo artificial. Las civilizaciones alienígenas eran tratadas en muchos sentidos como entes naturales —su distribución probable era definible mediante un cálculo científico como el de la ecuación

de Drake— pero, a la vez, eran vistas como algo artificial. La habilidad de los observadores de SETI radicaba en distinguir las señales alienígenas tanto de las verdaderamente naturales como de las artificiales terrestres. Se trataba de una tarea conceptualmente difícil, ya que, como Marvin Minsky señalaba en *Byurakan*, la ley de Shannon establece que las comunicaciones codificadas de forma eficiente son indistinguibles del ruido aleatorio si se desconoce el protocolo de codificación. Lo artificial podría, pues, parecerse natural; sólo dejaría de parecérnoslo si los alienígenas se hubieran propuesto que así fuera, si hubiesen diseñado sus emisiones para que se comporten como un radiofaro. Pero la naturaleza también puede asemejarse a un radiofaro. Cuando se descubrieron los pulsares, llevó tiempo aclarar que la naturaleza podía ser responsable de esas emisiones discontinuas; el equipo de Cambridge que los descubrió se refería jocosamente a esas fuentes de señal como LGM, la abreviatura de *little green men* (hombrecitos verdes). En 1965, Kardashev afirmó que las emisiones de radio del cuásar CTA-102 eran artificiales, una afirmación retirada poco después, cuando las observaciones ópticas las vincularon a fluctuaciones en el brillo del cuásar. Y, desde Ozma, toda búsqueda SETI ha captado señales de origen humano, desde radiofaros de navegación aérea hasta satélites espía.

Las diferencias entre lo natural, lo artificial y el singular terreno intermedio de los alienígenas inteligentes eran problemáticas, pero manejables en el altamente especializado campo de la

radioastronomía SETI, que se ocupaba sólo de objetos demasiado lejanos como para producir cualquier efecto causal sobre nuestro mundo. Si estaban más cerca, sin embargo, la cosa cambiaba. Los científicos confían en que pueden identificar lo natural, pero la presencia de alienígenas entre nosotros cuestionaría esta confianza. Un mundo en el que pulularan extraterrestres de incógnito ya no sería natural, sino engañoso. En un mundo así, la premisa científica más básica —la idea de que es posible «clasificar los iguales»— dejaría de ser válida. Poco después del primer Congreso de Green Bank, y poco antes de su prematuro fallecimiento, Otto Struve escribía: «Hoy deben quedar pocas dudas de que el libre albedrío del ser humano no tiene por qué estar restringido a la Tierra. Hemos de cambiar nuestra forma de pensar a la luz de este hecho». Pero dicho cambio, en la práctica, no es posible en las ciencias naturales, a no ser que los alienígenas aceptaran entrar en nuestro mundo social a través de la comunicación.

Unos extraterrestres libres resultaban inaceptables en la Tierra e inescrutables fuera de ella. Los argumentos posteriores a Hart sobre el valor de N se vieron condicionados por la nueva necesidad de explicar la estrategia de los alienígenas, de hacer hipótesis sobre sus tácticas, sus intenciones, sus temores y aversiones y sus objetivos. Estos supuestos sobre comportamiento resultaron ser mucho más difíciles de reconciliar con la lógica de las ciencias naturales que las premisas relativas a la elección de longitudes de onda y estrategias de señalización que desde siempre venía

asumiendo SETI (aunque, en realidad, la diferencia estaba en el alcance y no en el concepto). De ciencia natural sin datos, SETI se convertía en algo aún más difícil de llevar adelante con convicción: una ciencia social sin comunicación. Para muchos fue demasiado; si no era ciencia natural, perdía todo interés, como explicaba Dyson al responder a una petición de apoyo hecha por Sagan. «Me pregunto por qué te tomas tan en serio a Tipler. Creo que a sus argumentos se le presta más atención de la que merecen. No puedo aceptar ni sus números, ni los tuyos. Ningún modelo específico del futuro tiene por qué ser tan absurdamente estrecho y poco imaginativo». ¹²⁵

A pesar de los problemas asociados a N , las investigaciones de SETI continuaron en las décadas de 1980 y 1990. Como decía Stephen Jay Gould, uno de los muchos eminentes divulgadores científicos a los que Sagan solicitó suscribir la «petición SETI» en 1982, «Soy lo bastante egoísta como para desear ver algunos resultados exobiológicos [...] durante mi vida y SETI es lo único que tenemos por ahora». Más recientemente, sin embargo, nuevos enfoques y descubrimientos han hecho que la exobiología ajena a SETI vuelva a ser un campo mucho más prometedor. Como parte de ese proceso, el campo ha sido redefinido por su principal promotor —la NASA— como *astrobiología*: la principal diferencia es que la astrobiología trata de incluir aspectos del estudio de la vida en la Tierra en el mismo marco que el estudio de la vida en cualquier lugar del

¹²⁵ Carta de Dyson a Sagan, reproducida en *Captured by Aliens: The Search for Life and Truth in a Very Large Universe*, de Joel Achenbach (Simon & Schuster, 1999) (no se informa sobre la fecha de la carta). El libro de Achenbach es tremendamente divertido.

cosmos. Se trata de un paso lógico que también busca acallar la vieja crítica de que la exobiología es una ciencia sin materia de estudio. En el sistema solar, la nueva materia de estudio incluye la posible vida fosilizada en Marte, algo que ha despertado un interés creciente desde mediados de los ochenta, antes incluso de la controvertida evidencia de bacterias en el meteorito marciano ALH 84001. Hay gran interés por la posibilidad de vida en los océanos cubiertos de hielo de Europa, una de las lunas de Júpiter. Ha generado un gran entusiasmo el descubrimiento de sistemas planetarios alrededor de otras estrellas y un sistema de telescopios para detectar planetas similares a la Tierra y buscar indicios químicos de vida en sus atmósferas se ha convertido en pieza clave del programa astronómico de largo alcance de la NASA. (Curiosamente, esos sistemas se derivan de una de las propuestas de Ron Bracewell; la técnica que cabría utilizar en la práctica para detectar vida en planetas lejanos se remonta a una sugerencia hecha por James Lovelock en los años sesenta, cuando compartía despacho con Sagan en el Laboratorio de Propulsión a Chorro de la NASA).

En lo que se refiere a la NASA, la astrobiología es una fuerte candidata a convertirse en la nueva y revolucionaria ciencia del siglo XXI. Pero en la astrobiología, a diferencia de la exobiología, no hay un lugar para SETI. Esto en parte se debe a una batalla política que la agencia espacial perdió en 1993, cuando el Senado de Estados Unidos retiró la financiación a todos los ambiciosos programas SETI

que la NASA se había propuesto finalmente abordar. Pero si SETI no era políticamente factible, tal vez no resultaba ya intelectualmente necesario. Si la vida en planetas semejantes a la Tierra puede ser detectada directamente, a través de la paleontología marciana, la oceanografía de Europa o la espectroscopia infrarroja extrasolar, ya no es imprescindible que «nos hable». La astrobiología se desenvuelve perfectamente en el espacio definido por los cuatro primeros términos de la ecuación de Drake, los relativos a la frecuencia de los planetas habitables y a la probabilidad de vida. Los últimos términos pueden quedar temporalmente en suspenso, junto con todas las preocupaciones características de SETI acerca de lo no natural. Si esta nueva investigación astrobiología arrojara también un valor pequeño para N —hipótesis manejada por Peter Ward y Donald Brown lee en *Rare Earth* (Copernicus, 2000)—, tanto mejor, al menos desde el punto de vista de la coherencia en la materia. Un bajo valor de N certificado por la astrobiología haría de SETI un callejón sin salida históricamente interesante.

Sin embargo, el proyecto SETI continúa; de hecho, sigue creciendo. Aunque no dispone de una financiación centralizada, la aportación filantrópica y discrecional de distintas universidades ha permitido una amplia gama de investigaciones, entre ellas buena parte de las incluidas en el programa original de la NASA, sostenidas de forma privada a través del Instituto SETI en Mountain View, California. Con motivo del cuadragésimo aniversario del Proyecto Ozma, Drake —en la actualidad jefe de la junta directiva del Instituto SETI—

subrayaba entusiasmado que, gracias a los perfeccionamientos en las tecnologías de radio y procesamiento de señal, que permiten analizar simultáneamente un enorme número de frecuencias, los dispositivos de búsqueda son hoy cien billones de veces más potentes que el equipo original que él había empleado en Green Bank. Las capacidades de SETI han crecido incluso más deprisa que la potencia de los ordenadores, duplicándose cada diez meses más o menos. La mirada del astrónomo se va haciendo aún más poderosa a medida que va escudriñando más lejos.

Pero quizá en absoluto se trate de una mirada. La radioastronomía, debido a las connotaciones de la palabra *radio*, siempre se ha encontrado a caballo de dos metáforas contrapuestas: la visión y la audición. Cuando la radioastronomía produce imágenes —chorros procedentes de un cuásar, discos alrededor de un agujero negro— es una forma de la mirada del astrónomo. Pero SETI no genera imágenes. Por ello, siempre ha tenido una especial afinidad con la escucha y no con la visión, en particular en su versión popular, y dado que SETI es actualmente una investigación popular, financiada mediante suscripción y sostenida por miles de voluntarios que realizan el procesamiento de datos en sus ordenadores domésticos, parece oportuno tener esa versión popular muy en cuenta. Una de las primeras novelas basadas en SETI fue *The Listeners*, de James Gunn. En la representación más famosa del proyecto SETI en los medios de comunicación —la película de 1997 basada en la novela *Contact*, de Carl Sagan—, al espectador le llama

la atención de inmediato el que uno de los investigadores sea ciego (como, de hecho, lo es uno de los miembros del Instituto SETI, Kent Cullers). El fenómeno más llamativo de los que aparecen en el filme no es visual, sino auditivo: una palpitación visceral que satura los sentidos. La protagonista aparece tan deslumbrada por su intelecto que es incapaz de escuchar su corazón; su justificación final, sin embargo, es solamente ruido. A diferencia de la novela en la que está basada, la película contempla favorablemente ese famoso fenómeno ciego: la fe. El hecho ha disgustado a algunos de los partidarios más científicos de Sagan, pero no hay por qué verlo como un comentario negativo hacia SETI como estos partidarios han querido entender.

La mirada del astrónomo es poderosa porque la visión es la metáfora sensorial que define el conocimiento objetivo: «Ver es creer». Por otra parte, la audición es una metáfora primordial de la comprensión. La experiencia de ver se halla necesariamente enmarcada en un modelo espacial del mundo, en el que siempre hay una distancia entre el observador y el sujeto observado. La audición es directa, inmediata; la sentimos en nuestros oídos. La visión objetiviza el mundo; la audición abre la puerta al lenguaje y al sentimiento. Escuchar a alguien y mirarlo son dos cosas muy distintas, pues nadie puede ser invisible y, sin embargo, cualquiera, para ser oído, debe hablar. Escuchar es pasivo; mirar, activo. En todos esos sentidos, SETI resulta más auditivo y menos visual. No necesita números ni catálogos; no precisa de ecuaciones o de

cálculos. En última instancia, no es un estudio del universo, sino una comunicación con algo que no podremos conocer hasta que hable.

Cuando rezamos, cerramos los ojos.

Parte 9

Las ecuaciones de la vida

Las matemáticas de la evolución

John Maynard Smith

Se supone a menudo que los biólogos pueden prescindir de las matemáticas. Al fin y al cabo, en *El origen de las especies* de Darwin no hay una sola línea de álgebra. Pero esta idea es falsa, como vemos enseguida si analizamos la posterior historia de la biología evolutiva. La teoría de la evolución mediante la selección natural sólo funciona si los descendientes se parecen a sus padres, pero el propio Darwin desconocía el proceso de herencia por el cual esto sucede. Cuando comenzó a desarrollarse la ciencia genética, después de que Gregor Mendel descubriera las leyes genéticas en el año 1900, los estudiosos de la evolución pronto se dividieron en dos bandos. Por un lado, los «biométricos», un grupo de estadísticos liderados por el combativo Karl Pearson, argumentaban que lo que de verdad importaba en la evolución era la selección natural de caracteres en continua variación, tales como el tamaño o las proporciones corporales, y que los genes eran un producto irrelevante de la imaginación de Mendel. Por otro lado, muchos criadores de animales y plantas aceptaban la teoría mendeliana de los factores hereditarios discretos, o genes, e iban más allá, argumentando que las nuevas especies surgían por mutaciones genéticas de gran alcance y que la selección natural era irrelevante.

Vistos en retrospectiva, ambos argumentos parecen absurdos. Pero fueron necesarios estudios matemáticos —en especial, los llevados a cabo por los fundadores de la genética de poblaciones, los británicos R. A. Fisher y J. B. S. Haldane y el norteamericano Sewall Wright— para demostrar que ambos puntos de vista podían ser reconciliados. Las observaciones de los biométricos relativas a los caracteres en continuo cambio, las correlaciones entre parientes y los efectos de la selección, pueden ser explicadas como los efectos combinados de muchos genes, cada uno de ellos de pequeño alcance.

En el presente ensayo, comenzaremos describiendo una ecuación muy simple que predice la tasa de evolución cuando la selección no actúa; asimismo, explicaremos cómo se ha utilizado para datar sucesos evolutivos pasados. Pero el tema central se refiere a los muchos casos en los que es difícil predecir los efectos de la selección natural por no existir una «estrategia óptima» para el animal o la planta: en lugar de ello, la mejor estrategia depende de lo que hagan los otros miembros de la población. Veremos cómo una nueva rama de las matemáticas, la «teoría del juego evolutivo», ha sido empleada para resolver esos problemas.

Comenzaremos examinando un argumento cuantitativo relativo al proceso de mutación y evolución genéticas. Los fundadores de la ciencia habían mostrado que es la selección y no la mutación la que determina la dirección evolutiva. En la década de 1950, se descubrió que un gen consta de una cadena de cuatro tipos de moléculas, las bases A, C, G y T. Afortunadamente, no es preciso

que el lector conozca las fórmulas químicas de esas bases (ni siquiera las conozco yo). Lo que importa es que la secuencia específica de las bases en un gen determina la secuencia de aminoácidos en la proteína fabricada y las proteínas, a su vez, determinan la clase de organismo que se desarrolla.

Pronto se dispuso de datos sobre cambios evolutivos en las secuencias de aminoácidos de las proteínas (y más tarde, en las secuencias de las bases A, C, G y T que codifican las secuencias de aminoácidos). Analizando esos datos, el genetista japonés Motoo Kimura tuvo una idea original (aunque escandalosa para algunos). Aunque aceptaba que la evolución adaptativa es consecuencia de la selección natural que actúa sobre mutaciones aleatorias, pensó que muchos —de hecho, la mayoría— de los cambios evolutivos en los aminoácidos eran no adaptativos o «neutros». Es decir, en una población, la sustitución evolutiva de un aminoácido por otro —por ejemplo, la leucina por la treonina— ocurre no porque la selección favorezca la treonina, sino por puro azar: en una población, algunas proteínas tienen leucina en un sitio determinado, mientras que otras poseen treonina en el mismo sitio; dado que los genes presentes en la población de una generación son una muestra aleatoria de los que hay en la generación anterior, la proporción entre treonina y leucina cambiará gradualmente hasta que, finalmente, uno de los dos aminoácidos desaparece, a la vez que el otro se fija como único para toda la población.

Kimura subrayaba que su premisa de neutralidad tenía una consecuencia interesante sobre la tasa de evolución. Para entender su argumento, necesitamos definir primero el significado de «tasa de mutación». Supongamos que heredamos un gen concreto de uno de nuestros progenitores (p. ej., de nuestra madre). La tasa de mutación es simplemente la probabilidad de que, cuando transmitamos ese gen a un hijo nuestro, éste experimente un cambio o mutación genética. Habitualmente, la mayoría de las alteraciones consisten en el cambio de una única base (A, C, G o T) por otra; por lo general —aunque no siempre— este cambio afecta a un aminoácido. Kimura razonaba del siguiente modo. Supongamos que la «tasa de mutación neutra» del gen (la probabilidad, en una generación, de que surja una nueva mutación que no afecte a la supervivencia) es m . La mayor parte de esas mutaciones neutras se perderán por azar en pocas generaciones. De forma muy ocasional, y tras muchas generaciones, la nueva mutación quedará *fijada* en la población, es decir, en toda la población, el correspondiente gen será descendiente directo del gen mutante original, con lo que habrá tenido lugar un cambio evolutivo.

¿Cuál es la probabilidad de que esto suceda? Es obvio que depende del tamaño de la población: será más alta en una población pequeña. Imaginemos una población de mil ratas. Existen dos conjuntos de cromosomas en cada una, con lo que habrá dos mil copias de cada gen en la población. Supongamos ahora que la probabilidad de una nueva mutación neutra es uno sobre cien (una

cifra extraordinariamente elevada). Habría dos mil sobre cien, es decir, veinte mutaciones neutras nuevas en cada generación. Toda mutación neutra tiene exactamente la misma probabilidad de quedar fijada —al fin y al cabo, esto está implícito en el significado de *neutra*—. Como hay dos mil genes, todo gen mutante tiene una probabilidad de uno entre dos mil de quedar fijado. Así pues, el número de mutaciones neutras que surgen y quedan fijadas en cada generación es igual a $20 \times 1/2.000 = 1/100$. Obsérvese que este valor es igual a la tasa de mutación y resulta independiente del tamaño de la población, la cual se cancela en los cálculos: si doblamos la población, se duplica el número de imitantes, pero se reduce a la mitad la probabilidad de que cualquiera de ellos quede fijado. Si el tamaño de la población es N y la tasa de mutación neutra por generación vale m , el número de mutaciones neutras que se fijan en cada generación es $2Nm \times 1/2N = m$. Expresado en forma de ecuación:

$$\begin{aligned} \text{número de mutaciones neutras fijadas en cada generación} &= \\ &= \text{tasa de mutación neutra por generación} \end{aligned}$$

¡Ya nos gustaría que todas las matemáticas biológicas fueran así de fáciles!

De este modo, la teoría de Kimura dice que la tasa de evolución molecular neutra depende solamente de la tasa de mutación y es independiente del tamaño de la población. Esto último es

importante, ya que, por lo general, no tenemos ni idea del tamaño de las poblaciones pasadas, pero cabe asumir plausiblemente que la tasa de mutación es más o menos constante. La idea de Kimura se convirtió en el punto de partida de un gran cuerpo de teoría matemática, que ha sido aplicada de dos maneras. En primer lugar, proporciona una «hipótesis nula» frente a la cual medir la selección: las desviaciones respecto a las predicciones de la teoría neutralista indican que ha habido selección. En segundo lugar, existen cambios que probablemente fueron casi neutros (los mejores candidatos son los denominados cambios en bases «sinónimas», es decir, cambios que no alteran el aminoácido y, por lo tanto, no modifican el funcionamiento de una proteína). Los citados cambios nos sirven para datar sucesos pasados. Si podemos estimar la tasa de mutación, comparando la secuencia de ADN de un mismo gen en dos especies animales actuales y contando el número de cambios en bases sinónimas entre ambas, es posible datar el último ascendiente común de las dos (p. ej., de hombres y chimpancés o de pájaros y mamíferos).

La teoría neutralista de la evolución molecular se parece a una teoría física —por ejemplo, a la mecánica de Newton— en que es un cuerpo de predicciones matemáticas derivado de un pequeño conjunto de premisas simples y difiere en que dichas premisas son, como mucho, buenas aproximaciones. Por ejemplo, las tasas de mutación no son constantes y las mutaciones sinónimas no son neutras del todo. Por este motivo, en biología empleamos las teorías

matemáticas de una manera distinta. Si una teoría física —la ley de la gravitación universal, por ejemplo— hace predicciones que difieren, aunque sea levemente, de la observación, los físicos se preocupan y buscan la razón de la discrepancia. En el caso antes elegido, las leyes de Newton predicen que los planetas deben recorrer órbitas elípticas, con el Sol en uno de los focos. En dos ocasiones, medidas cuidadosas revelaron que el movimiento de un planeta se apartaba ligeramente de la predicción. En la primera, las irregularidades en el movimiento de Neptuno sugirieron que el planeta era perturbado por la atracción de un objeto celeste entonces desconocido; siguiendo esta pista, se descubrió Plutón. En la segunda, las desviaciones en la trayectoria de Mercurio sirvieron para confirmar las predicciones de la teoría general de la relatividad de Einstein.

En biología usamos las ecuaciones de manera diferente. No nos podemos permitir el lujo de estudiar las interacciones entre dos únicos cuerpos, cada uno de los cuales puede ser tratado como si su masa se concentrara en un solo punto. Por el contrario, estudiamos las interacciones entre un gran número de organismos, cada uno de los cuales es de enorme complejidad. ¿En qué forma nos servimos de las ecuaciones a la hora de abordar estas complicaciones? En primer lugar, aislamos cierto fenómeno para su estudio. El latido del corazón, el ritmo diario de sueño y vigilia o el ciclo de diez años en la población de liebres, lince, aves de caza y otros animales en el Ártico canadiense son todos ellos oscilaciones

periódicas, pero no es probable que respondan al mismo mecanismo subyacente. Así pues, las estudiamos de una en una. Luego, mediante una mezcla de experimentos e intuición, suponemos un mecanismo o —en términos más pomposos— establecemos una hipótesis sobre él. Para averiguar si la suposición es correcta, suele ser útil escribir ecuaciones que describan el mecanismo propuesto y, resolviendo (o simulando) dichas ecuaciones, comprobar si generan la *clase* de comportamiento que observamos. En otras palabras, esperamos que nuestras ecuaciones predigan cualitativamente el comportamiento correcto. Esperar que los números se ajusten con precisión es pretender demasiado. Por ejemplo, cabría esperar que un modelo matemático del ciclo ecológico canadiense prevea oscilaciones regulares en los números con un periodo más o menos correcto, pero no es realista confiar en obtener el periodo o la amplitud exactos (en la naturaleza son muy variables en cualquier caso).

Una razón por la que sólo cabe esperar predicciones cualitativas es que, en todo modelo concreto, obviamos demasiadas cosas. Por ejemplo, al modelar el ciclo canadiense, probablemente sólo incluiremos algunas de las especies más abundantes (y nuestra elección puede ser fácilmente errónea), e ignoraremos fluctuaciones en el clima de un año a otro (aunque algunas teorías consideran cruciales dichas fluctuaciones) y diferencias espaciales en el entorno. Los estudiantes recién llegados a la biología evolutiva a menudo me preguntan por qué se excluye de un modelo algo que

seguramente afecte al resultado final. La respuesta es, en primer lugar, que si dejamos fuera algo verdaderamente importante, el modelo no hará las predicciones correctas, ni siquiera cualitativamente, y, en segundo lugar, que si tratásemos de incluirlo todo en un modelo, éste acabaría resultando inútil. Tendremos que utilizar simulaciones por ordenador para estudiar su comportamiento y no sabremos a ciencia cierta cuáles de las características que hemos incorporado son importantes. Como los antropólogos Robert Boyd y Peter Richerson señalaban en cierta ocasión, «De nada sirve sustituir un mundo que no comprendes por un modelo de él que tampoco comprendes».

Así pues, en biología sólo son útiles los modelos sencillos. El precio que pagamos por esa simplicidad es la ausencia de precisión cuantitativa en nuestras predicciones. Pero, si tenemos suerte, tal vez nos topemos con que nuestro modelo explica aspectos del mundo en los que no habíamos pensado cuando lo estábamos creando. Trataré de ilustrar este punto describiendo un modelo de comportamiento animal que elaboré junto con George Price hace ahora treinta años. Su versión original era tan simple que casi resultaba trivial. Sin embargo, hasta en esa forma simple explicaba ciertos aspectos del comportamiento que, aunque conocidos por los biólogos, no teníamos en mente al inventar el modelo. Y, lo que es más importante, el método que adoptamos —la construcción de un «juego evolutivo»— ha resultado tener muchas más aplicaciones de las que nunca habiéramos soñado. Se ha utilizado para analizar

conceptos tan variados como la señalización animal, el crecimiento de las plantas, la evolución de los virus y la proporción entre machos y hembras en una población. Examinaremos este último problema, así como el inicial relativo al comportamiento que dio origen al modelo.

Comenzaremos con el problema original. Cuando era estudiante de zoología, allá en los años cuarenta, mis compañeros y yo estábamos fascinados por los recientes descubrimientos de los etólogos Konrad Lorenz y Niko Tinbergen —en parte, sospecho, porque nuestros profesores no parecían haber oído hablar de ellos y era agradable sentirse superior—. Lorenz, en particular, señalaba que, cuando dos animales compiten por un recurso, no suelen utilizar todas sus armas —dientes, cuernos, garras— en una lucha sin cuartel, sino que se limitan a enfrentamientos rituales y a menudo saldan la pugna sin que haya una pelea encarnizada. En aquella época, ese comportamiento ritual se explicaba diciendo que era bueno para la especie; como un distinguido zoólogo decía, si las luchas encarnizadas fueran habituales, muchos individuos resultarían heridos, lo cual «iría en contra de la supervivencia de la especie».

Aunque era un simple estudiante, la explicación me parecía errónea. La teoría darwiniana de la evolución mediante la selección natural —que es para la biología lo que las leyes de Newton fueron para la física durante trescientos años— es básicamente una teoría de selección individual. En un entorno dado, ciertos tipos de individuos tienen más posibilidades de sobrevivir y reproducirse que

otros. Cuando se reproducen, transmiten sus características a los descendientes. El resultado es una población compuesta por individuos dotados de rasgos que contribuyen a su supervivencia. Si éstos son perjudiciales para la especie, mala suerte. ¿Cómo, pues, explicar el comportamiento ritual en la lucha?

Aunque ya era consciente del problema en 1950, no reflexioné seriamente sobre él hasta veinte años después, cuando, casi por accidente, decidí informarme sobre la «teoría de juegos» para ver si era aplicable al tema. Esta teoría había sido desarrollada en primera instancia a comienzos de la década de 1940 por John von Neumann y Oskar Morgenstern, con objeto de estudiar los *juegos* humanos, es decir, las interacciones humanas en las que la mejor acción a realizar depende de lo que haga nuestro *oponente*. Por ejemplo, en una partida de póquer, ¿cuándo conviene echarse un farol? Para un biólogo, el principal obstáculo en esta teoría es que parte de la premisa de que los contendientes se comportan de manera racional y asumen en sus cálculos que sus oponentes son racionales también. Obviamente, no cabe hacer esta hipótesis tratándose de animales. Sin embargo, los teóricos del juego han dado a luz una sencilla idea que me pareció muy útil: la «matriz de pagos».

Imaginemos el *juego* siguiente: dos animales compiten por cierto recurso (un territorio, un trozo de comida o una hembra) de valor V . Cada individuo puede elegir entre una de estas dos *estrategias*: halcón o paloma. (Una estrategia significa aquí, simplemente, un tipo de comportamiento, pero en posteriores contextos implicará

una característica heredable. Los términos halcón y paloma resultaban muy gráficos debido a que el juego fue inventado en Chicago durante la guerra de Vietnam —no son demasiado aplicables a las aves en cuestión—.) En una contienda, un halcón lucha con todas sus armas hasta que vence y obtiene el recurso (el beneficio V) o hasta que resulta gravemente herido (lo que supone un coste C ; el significado de beneficio y coste se discute con posterioridad). La paloma hace ostentación de fuerza; si su oponente ataca, huye antes de sufrir daño, pero si se topa con otra paloma, comparte el recurso, de forma que ambas obtienen un beneficio $V/2$. A partir de estas estrategias, podemos establecer las *recompensas* para un individuo dado, dependiendo de su estrategia y la de su oponente:

1. Un halcón, jugando contra una paloma, obtiene el recurso sin sufrir daño: su beneficio es V . La paloma no obtiene nada, pero tampoco sufre daño: su beneficio es 0 .
2. Si un halcón se topa con otro halcón, tiene tantas posibilidades de ganar (beneficio V) como de perder (coste C): en promedio, su beneficio es $1/2(V - C)$.
3. Si se enfrentan dos palomas, comparten el recurso, obteniendo ambas un beneficio $V/2$.

Estos resultados pueden ser expresados mediante una «matriz de pagos», en la que en las filas se indica la recompensa para un

individuo si su oponente adopta la estrategia indicada en las columnas:

	<i>Estrategia del oponente</i>	
	Halcón	Paloma
Recompensa para el halcón	$1/2(V - C)$	V
Recompensa para la paloma	0	$V/2$

Imaginemos una población que desarrolla este juego. ¿Cómo evolucionará? Cuando un halcón se encuentra con una paloma, gana, pero de ello no se deduce que los halcones reemplacen a las palomas en la población, ya que cuando un halcón se tope con otro, le puede ir muy mal. No deseamos saber el resultado de un único encuentro, sino cómo evolucionará la población a lo largo del tiempo. Para ello asumiremos lo siguiente:

1. Todo individuo entabla una lucha contra un oponente al azar; obtenemos los mismos resultados en individuos que entablen una serie de enfrentamientos contra oponentes aleatorios.
2. Un individuo produce después un número de descendientes que depende del beneficio obtenido en la lucha.¹²⁶ En otras palabras, interpretaremos V y C como el cambio, resultante de la lucha, en el número esperado de descendientes.

¹²⁶ Como no es posible que los individuos tengan un número negativo de descendientes, supondremos que dicho número es igual a una constante más el beneficio.

3. Cuando los individuos se reproducen, los halcones generan halcones y las palomas, palomas. Por supuesto, asumimos reproducción asexual e ignoramos los detalles de la genética mendeliana. En la época en la que creamos el modelo, la justificación era que, por lo general, se sabía poco sobre la genética de los rasgos concretos de comportamiento, pero existía cierta semejanza entre padres e hijos para la práctica totalidad de los rasgos analizados. Con posterioridad, las matemáticas han demostrado que los resultados derivados de nuestros supuestos originales concuerdan bastante bien con lo que sucede en una población sexual, aunque pueda haber diferencias.

Tenemos, pues, el modelo. ¿Qué predicciones hace? Supongamos, en primer lugar, que el beneficio V es mayor que el coste C ; por ejemplo, $V = 10$ y $C = 4$. La matriz de pagos sería:

	<i>Estrategia del oponente</i>	
	Halcón	Paloma
Recompensa para el halcón	3	10
Recompensa para la paloma	0	5

¿Qué sucederá? Para esta matriz, la respuesta es fácil. Necesitamos saber la cantidad relativa de descendientes que producen, respectivamente, los halcones y las palomas. Los primeros salen

ganando, sea cual sea la estrategia de su oponente ($3 > 0$ y $10 > 5$). En otras palabras, cualquiera que sea la proporción inicial entre unos y otras en la población, los individuos halcón tendrán, en promedio, más descendientes, con lo que aquélla terminará compuesta exclusivamente por halcones.

Es más interesante el caso en el que el coste C es superior al beneficio V ; ya no merece la pena correr el riesgo de resultar herido para obtener el recurso. Hagamos, por ejemplo, $V = 4$ y $C = 10$. La nueva matriz de pagos será:

	<i>Estrategia del oponente</i>	
	Halcón	Paloma
Recompensa para el halcón	-3	4
Recompensa para la paloma	0	2

Ahora, si la población está compuesta mayoritariamente por halcones, lo ideal es ser paloma, pero si lo está mayoritariamente por palomas, es preferible ser halcón. Dicho de otra manera, una población de halcones podría ser invadida por un mutante paloma y una población de palomas, por un mutante halcón. ¿Qué ocurrirá en este caso? Al parecer, al final acabaríamos con una población de individuos que unas veces se comportarían como halcones y otras, como palomas —lo que se denomina una «estrategia mixta»—, al igual que un buen jugador de póquer algunas veces va de farol y

otras, no. Pero ¿es realmente así? Y si lo es, ¿en qué proporción se adoptarán los papeles de halcón y de paloma?

La clave está en buscar una «estrategia evolutivamente estable», o EEE. Una EEE es una estrategia «que no es posible invadir», en el sentido siguiente. Supongamos que casi todos los miembros de una población adoptan cierta estrategia que denominaremos S. Esto significa que, típicamente, un individuo S se topará con otro S y obtendrá la «recompensa de S contra S». Imaginemos ahora un raro mutante, Y. Éste, también típicamente, se enfrentará a un individuo S y obtendrá la «recompensa de Y contra S». Supongamos que la «recompensa de Y contra S» es inferior a la «recompensa de S contra S» para todas las mutaciones posibles. En este caso, ningún mutante podrá invadir la población y diremos que S es una EEE. En el lenguaje de la calle, una EEE es una estrategia que funciona mejor contra sí misma que cualquier otra estrategia que se enfrentase a ella. De manera más formal, una EEE puede ser definida como una estrategia S tal que, si casi toda la población la adopta, ninguna estrategia mutante Y puede invadirla; esto será así si la estrategia S funciona frente a ella misma mejor que cualquier estrategia mutante Y frente a la propia S. Volviendo a la primera matriz, con $V = 10$ y $C = 4$, la del halcón es una EEE, ya que, frente a sí misma, funciona mejor que la estrategia de la paloma. Para la segunda matriz buscamos una estrategia que consista en «comportarse como un halcón con una probabilidad p y como una paloma, con una probabilidad $1 - p$ ». Para hallar p nos basamos en

el hecho de que, en una EEE, las recompensas obtenidas al ser paloma o halcón han de ser, en promedio, iguales; en caso contrario, la población no estaría en equilibrio. Esto conduce a un valor de $p = 0,4$. Es decir, la EEE consiste en que un individuo se comporte como halcón el cuarenta por ciento de las veces y como paloma, el sesenta por ciento restante. La conclusión es que, si las únicas tácticas posibles en una contienda son «luchar hasta la muerte» y «hacer ostentación de fuerza, pero salir corriendo si el oponente ataca», la única EEE disponible es de tipo mixto: optar por una u otra estrategia según convenga. Existe una pequeña complicación. Hay dos estados estables posibles para una población que desarrolle un juego de este tipo. Uno corresponde al caso en el que todos los individuos adoptan la estrategia mixta; el otro, a aquel en el que un cuarenta por ciento de la población se comporta siempre como un halcón y el otro sesenta por ciento, como una paloma.

¿Es cualitativamente correcta esta conclusión para los animales del mundo real? Existen ciertas situaciones en las que los organismos adoptan una EEE mixta de esta clase. Por ejemplo, en el pez sol de agalla azul existen dos tipos de machos. Uno de ellos crece sin procrear durante más de cinco años y luego establece un territorio de cría, fertilizando los huevos de las hembras que entran en él. El otro, conocido como *furtivo*, se oculta en el territorio de un macho en fase de procrear; cuando las hembras depositan sus huevos, el furtivo sale de su escondite, vierte esperma sobre ellos y escapa. Sin

embargo, no creo que la solución sea aplicable a la clase de contienda que el juego del halcón y la paloma trataba de modelar. ¿Dónde está el fallo? Price y yo propusimos otras posibles estrategias que podría adoptar un animal. Examinaremos dos de ellas.

La primera es la del *vengador*: «Compórtate como una paloma, pero si tu oponente lucha, contraataca». En nuestro ejemplo numérico, la matriz de pagos sería la siguiente:

	<i>Estrategia del oponente</i>		
	Halcón	Paloma	Vengador
Recompensa para el halcón	-3	4	-3
Recompensa para la paloma	0	2	2
Recompensa para el vengador	-3	2	2

Vengador es casi una EEE. Una población de vengadores no puede ser invadida por halcones ($2 > -3$) o por cualquier tipo de estrategia mixta, pero la evolución de una población compuesta sólo por vengadores y palomas es ambigua, ya que se comportarán de manera idéntica. Para soslayar este punto, Price y yo introdujimos una estrategia más complicada, la del *sondeador-vengador*, el cual, frente a las palomas, intentaba ocasionalmente un «ataque de prueba», volviendo a la simple ostentación de fuerza si su oponente contraatacaba. La nueva estrategia demostró ser una EEE y podría constituir una buena descripción de lo que sucede en ciertas luchas

animales. Nuestra estrategia *sondeador-vengador* se ha hecho luego popular en las discusiones sobre juegos evolutivos bajo el nombre de «Donde las dan, las toman».

Existe, no obstante, otra solución que es, a la vez, más elegante y una descripción mejor de lo que habitualmente sucede. Supongamos que dos humanos jugasen a este juego. ¿Se pondrían de acuerdo para compartir el recurso? ¿Y si el recurso es indivisible? Porque el lector se habrá dado cuenta ya, probablemente, de que la premisa de que el recurso puede ser compartido es, a menudo, inviable. En un caso así, tal vez los dos humanos acordarían lanzar una moneda al aire e incluso buscarían un testigo que certificase el veredicto. No cabe imaginar dos animales lanzando una moneda. Pero ¿cuál es la función real de ésta? La moneda se limita a introducir una asimetría en una situación hasta ese instante simétrica. Este hecho nos sugirió que los animales podrían basarse en alguna asimetría para saldar sus enfrentamientos. La asimetría más obvia es la existente entre el *propietario* del recurso y un intruso. Por supuesto, no pretendíamos que los animales poseyeran el concepto de propiedad; bastaba con que el animal tuviese que luchar más duramente por un recurso — como puede ser un territorio— que hubiese ocupado durante algún tiempo que por otro que, simplemente, acababa de encontrar.

Consideremos la estrategia que, por razones obvias, denominamos del *burgués*: luchar encarnizadamente por un recurso si lo poseemos ya, y no hacerlo en caso contrario. Es decir, «ser halcón si

se es propietario y paloma si se es intruso». Si asumimos razonablemente que un burgués será con la misma probabilidad propietario o intruso, la matriz de pagos resulta ser:

	<i>Estrategia del oponente</i>		
	Halcón	Paloma	Burgués
Recompensa para el halcón	-3	4	0,5
Recompensa para la paloma	0	2	1
Recompensa para el burgués	-1,5	3	2

La estrategia del burgués es una EEE: en una población compuesta mayoritariamente por burgueses, la recompensa media para un burgués es 2, mientras que para un halcón y una paloma vale 0,5 y 1, respectivamente. Los burgueses, pues, no pueden ser invadidos ni por halcones ni por palomas. Obsérvese que la conclusión no requiere asumir nada acerca de si el propietario de un recurso tiene más o menos posibilidades de vencer en una lucha abierta.

Muchos animales siguen esta sencilla estrategia. La evidencia más importante es que, cuando dos animales se consideran propietarios de un mismo recurso, sobreviene la lucha; en ocasiones es posible crear esta situación de manera experimental. Estimulado por el modelo que acabamos de describir, el zoólogo Nick Davies investigó el comportamiento territorial de la mariposa de los muros. En un bosque, los machos mantienen áreas bañadas por la luz solar como territorios de cría. Si un intruso penetra en un área dominada ya

por otro macho, se produce un breve enfrentamiento en el que ambos vuelan hacia arriba en espiral, hasta que el intruso se bate en retirada. Davies apartó al propietario de un territorio y permitió que otro macho ocupara la zona de luz. Luego liberó al macho original. Ambos se comportaron como propietarios. El resultado fue un prolongado vuelo en espiral, mucho más largo que un encuentro típico. Cuando existen otras asimetrías, tales como diferencias en tamaño o edad, la cosa se complica. Cabe preguntarse qué sucedería con la estrategia contraria: «Actuar como paloma si se es propietario o como halcón si se es intruso». Si los animales se limitaran a jugar una sola vez en la vida, parecería una alternativa válida a la estrategia del burgués, aunque nunca he sabido muy bien cómo llamarla. Pero si el juego se reitera, está claro que la estrategia es un problema. En cuanto un animal se ha convertido en propietario de un recurso, debe entregárselo al primer intruso que aparezca. Conozco un estudio sobre un animal, cierta araña semisocial, que parece seguir esta estrategia.¹²⁷ Estas arañas fabrican una telaraña comunitaria, con agujeros separados, cada uno de ellos ocupado por un individuo. Cuando una araña es expulsada de su agujero y éste queda destruido, la araña recorre la tela y entra en otro agujero. La ocupante de este último se marcha y busca nuevo alojamiento; el proceso se repite hasta que una araña encuentra un agujero vacío. Me extrañaría que un comportamiento tan paradójico fuera frecuente.

¹²⁷ J. W. Burgess, *Scientific American*, 1976, 234, págs. 100-106.

Antes de dejar estas contiendas simples, me gustaría examinar un último juego, el denominado «guerra de desgaste» —mucho más complejo matemáticamente que los juegos que hemos visto hasta ahora—. Supongamos que dos individuos compiten por un recurso indivisible de valor V y que físicamente no pueden luchar. Lo único que pueden hacer es continuar amenazándose hasta que uno de los dos abandone. ¿Cuánto tiempo debería seguir insistiendo un individuo? Obviamente, si amenazar no supone coste alguno y V es mayor que cero, el juego se bloquearía, ya que ambos contendientes seguirían amenazándose eternamente, lo cual es absurdo. Debemos asumir que amenazar supone un coste y que ese coste se incrementa con el tiempo de igual modo para los dos contendientes; supongamos que el enfrentamiento dura un tiempo t , con un coste kt para ambos. Supongamos que éstos eligen los tiempos t_1 y t_2 , respectivamente. Si t_1 es mayor que t_2 , la contienda dura un tiempo t_2 , con un coste kt_2 para ambos participantes, y el primero de ellos es el que obtiene el recurso. Las recompensas son: $V - kt_2$ para el primer contendiente (sólo ha de pagar kt_2 , aunque estaba dispuesto a pagar kt_1) y $-kt_2$ para el segundo.

¿Cómo debería comportarse un individuo? Dicho de forma más precisa, ¿existe alguna regla de comportamiento que sea evolutivamente estable, en el sentido de que, si todos la adoptan, ninguna regla *mutante* funcionaría mejor? Resulta que la EEE consiste en que todo individuo tenga la misma (y constante) probabilidad de abandonar por segundo, independientemente de lo

que haya durado la contienda hasta ese momento. Por ejemplo, «abandonar en el próximo segundo con una probabilidad de uno sobre cien». El valor de la probabilidad depende del valor de V : cuanto mayor sea éste, menor será la probabilidad de abandonar y más grande la duración media de una contienda. El resultado de adoptar una regla así aparece en la *figura 9.1*, conocida como ley de desintegración exponencial.

Este tipo de curva describe el nivel de radiactividad, a lo largo del tiempo, de un fragmento de material radiactivo, como un residuo nuclear. Corresponde a la distribución esperada si todo átomo radiactivo de la muestra tiene una probabilidad por segundo constante de *desintegrarse* y escindirse en átomos más pequeños. De manera parecida, en la *guerra de desgaste* los individuos tienen una probabilidad fija por segundo de abandonar.

¿Por qué la semejanza? ¿Por qué los contendientes tendrían una probabilidad fija de abandonar? Encontré la solución mediante el razonamiento siguiente. Consideremos un individuo en el curso de una contienda así. Ha estado haciendo alarde de fuerza durante un tiempo t_1 ; ¿cuánto debería continuar? La respuesta es que debería seguir exactamente el mismo tiempo que estaba dispuesto a aguantar al comenzar la contienda. Al fin y al cabo, aún tiene ante él la misma recompensa (y el mismo castigo) que había al principio: el recurso de valor V a adquirir y el coste kt por cada intervalo de tiempo t adicional que la competición se prolongue. Es cierto que ya ha gastado kt_1 , pero eso es agua pasada —ya no puede hacer nada

al respecto—. Si las pérdidas y beneficios potenciales futuros son exactamente los que eran al principio, su futuro comportamiento debería ser exactamente igual al inicial. En otras palabras, su probabilidad de abandonar en cualquier intervalo de tiempo debería permanecer constante, del mismo modo que un átomo radiactivo tiene una probabilidad fija de desintegrarse en la unidad de tiempo, no importa cuánto tiempo lleve en ese estado radiactivo.

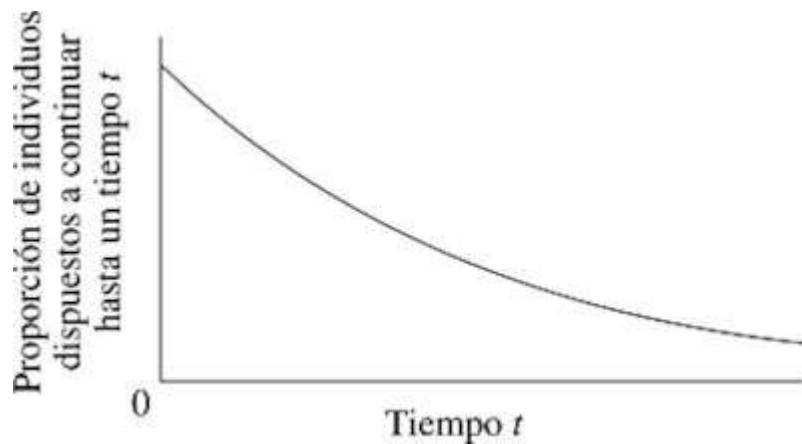


Figura 9.1. El juego de la «guerra de desgaste». Distribución de frecuencias en una población de individuos que hacen alarde de fuerza durante distintos periodos. La forma de la curva viene determinada por una ecuación sencilla.

¿Juegan los animales este juego? Creo que no: siempre les merecerá la pena encontrar una asimetría que resuelva la contienda. Pero existe un juego similar, aunque ligeramente más difícil, en el que cierto número de animales compiten simultáneamente por un recurso. Mi ejemplo favorito es un estudio de Geoff Parker — realizado cuando era estudiante en Bristol— sobre la mosca del

estiércol macho, que aguarda sobre una boñiga de vaca la llegada de hembras vírgenes. ¿Cuánto tiempo debía permanecer allí? Geoff encontró que las moscas adoptaban una solución cuantitativa bastante precisa para el tiempo de espera evolutivamente estable; por supuesto, para descubrirlo, Geoff tuvo que esperar también en el mismo sitio.

Una última sugerencia antes de dejar las contiendas animales. Si el lector es experto en programación, puede intentar analizar la dinámica de una población mediante el juego «halcón-paloma-vengador-matón». Las tres primeras estrategias son ya conocidas. *Matón* es lo contrario de vengador: «Ser halcón frente a una paloma y paloma frente a un halcón». La dinámica de este juego es fantástica.

En un juego simple como el del halcón y la paloma, en el que las estrategias posibles sean las dos citadas más un conjunto de estrategias mixtas compuestas por porcentajes diversos de ambas, existe siempre una EEE. Dependiendo de los valores de coste y beneficio, resultará ser halcón, paloma o las dos; en este último caso, la población evolucionará hasta estar compuesta sólo por halcones o sólo por palomas, según sean las frecuencias iniciales. Pero no todos los juegos tienen una EEE. A primera vista, parece extraño: la población terminará en alguna parte, ¿no? Por supuesto, pero no necesariamente en un punto estable; puede que siga evolucionando cíclicamente para siempre (o, en la práctica, hasta

que cambien las circunstancias). Pero para que un juego no disponga de EEE deben existir más de dos estrategias puras.

Un juego muy sencillo que puede no presentar una EEE es el conocido juego infantil «piedra, papel o tijeras». En él, piedra vence a tijeras (la piedra mella el filo de las tijeras), tijeras vencen a papel (las tijeras cortan el papel) y papel vence a piedra (el papel envuelve la piedra). La matriz de pagos es la siguiente:

	<i>Estrategia del oponente</i>		
	Piedra	Papel	Tijeras
Recompensa para piedras	$1 + e$	0	2
Recompensa para papel	2	$1 + e$	0
Recompensa para tijeras	0	2	$1 + e$

Hemos supuesto que vencer tiene una recompensa de 2 unidades. Si dos oponentes adoptan la misma estrategia, obtienen ambos una recompensa de $1 + e$ (asumiendo que pueden compartir la ganancia), donde e es una recompensa adicional (digamos, por el hecho de haber evitado la disputa). Pero e no tiene por qué ser positiva; cabría asumir también que es negativa y representa un pequeño coste para los individuos que adoptan la misma estrategia. Como veremos, el resultado depende por completo de si e es positiva o negativa.

Está claro que ninguna de las estrategias puras —piedra, papel y tijeras— puede ser una EEE: una población que adopte piedra

puede ser invadida por papeles y así sucesivamente. La única candidata a EEE sería la estrategia mixta «comportarse como piedra, papel o tijeras con igual probabilidad». Si se analizan las condiciones de estabilidad, se observa que, cuando e es positiva, la estrategia mixta no es una EEE: puede ser invadida por cualquiera de las estrategias puras. Pero ninguna de éstas es estable. ¿Qué sucede entonces? El comportamiento de una población puede ser representado como una trayectoria en un «espacio de estados», tal como muestra la *figura 9.2 B*. El sistema entra en una oscilación permanente.

Sin embargo, si la recompensa e es negativa, la estrategia mixta se convierte en EEE y la dinámica del sistema es la de la *figura 9.2 C*. ¿Qué ocurre cuando $e = 0$? El equilibrio se hace «neutralmente estable» y la dinámica es como en la *figura 9.2 E*. Este tipo de dinámicas, en forma de un conjunto de ciclos cerrados, se denominan «conservativas». No hallamos sistemas así en el mundo real, ya que el más mínimo cambio en las circunstancias llevaría a la situación de las *figuras 9.2 B* o *9.2 C*. En cambio, es posible encontrar sistemas dinámicos en oscilación permanente con una amplitud constante —el mundo vivo está lleno de ellos—. Pero su comportamiento dinámico es el de la *figura 9.2 D*: el sistema se instala en un ciclo de amplitud fija, cualquiera que sea su punto de partida.

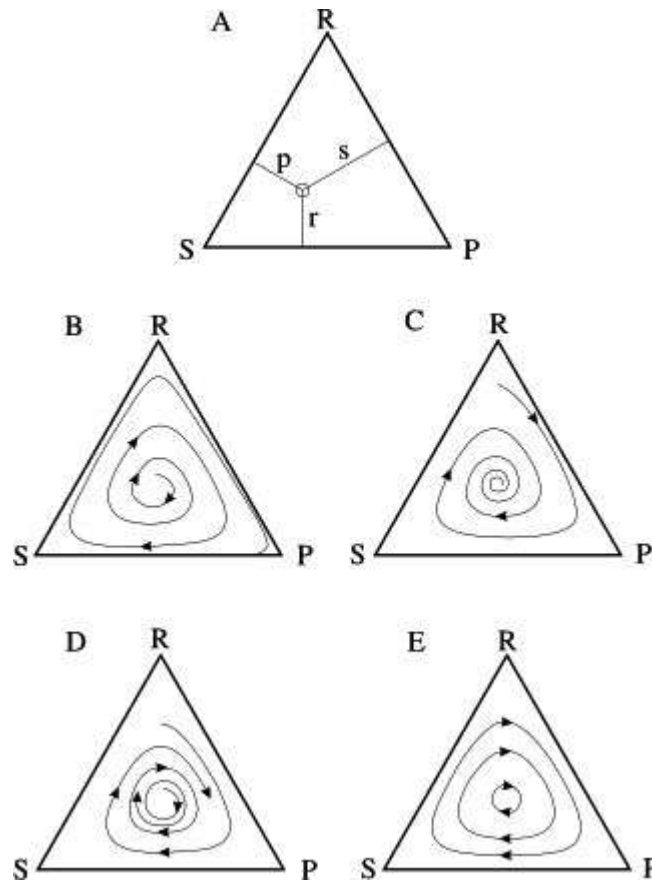


Figura 9.2. La dinámica del juego de «piedra, papel o tijeras» [rock (R), paper (P), scissors (S), en inglés]. El estado de la población en un momento dado puede ser definido por los valores de r , p y s (las frecuencias de R, P y S en la población, respectivamente). Como $r + p + s = 1$, cabe representar dicho estado como un punto en el interior de un triángulo equilátero —diagrama A— y describir la forma en que evoluciona en el tiempo. El diagrama B muestra lo que sucede cuando la recompensa adicional e es positiva y el C, cuando es negativa. El diagrama D muestra un ciclo límite (un comportamiento que no surge en la forma más simple del juego piedra, papel o tijeras, pero que se aproximaría más a la realidad). Finalmente, el diagrama E muestra lo que sucede cuando e vale cero.

En 1982 utilicé el juego piedra, papel o tijeras para ilustrar la posibilidad teórica de un juego sin EEE; nunca esperé toparme con animales que desarrollaran un juego tan simple. Por eso me sorprendió, catorce años después, tropezar con un artículo publicado en *Nature* y titulado «Los lagartos juegan a piedra, papel o tijeras».¹²⁸ El artículo describía una especie de lagarto con tres tipos de macho. El macho de cuello naranja domina un territorio que aloja a varias hembras. Una población de machos de este tipo puede ser invadida por machos «furtivos» de cuello verde, que aguardan a que un macho de cuello naranja se dé la vuelta para aparearse con una de sus hembras. Pero una vez la población está formada mayoritariamente por machos de cuello verde, puede ser invadida por machos de cuello azul, cada uno de los cuales domina un territorio de suficiente tamaño para alojar a una hembra. Ni qué decir tiene que, cuando los machos de cuello azul son dominantes, pueden ser invadidos a su vez por los machos de cuello naranja originales, completándose el ciclo.

Para cualquier teórico, es motivo especial de satisfacción el que un animal acabe por hacer algo que él había predicho teóricamente, pero que parecía muy raro que se diera en la realidad. Sin embargo, creo que los juegos de ciclo permanente y sin EEE pueden ser más habituales de lo que a simple vista parece. Un modelo de situación que podría estar muy extendido es el denominado «juego del ciervo

¹²⁸ B. Sinervo y C. M. Lively, *Nature*, 1996, 380, págs. 240-243.

rojo». Imaginemos una especie —por ejemplo, el ciervo rojo— en la que el macho crece sin procrear hasta que ha alcanzado un determinado tamaño. Llegado el momento, en la época de celo, los machos compiten y los más fuertes acumulan un harén de hembras con las que se aparean. Durante las luchas, un macho gasta tanta energía que a partir de entonces crece muy poco o, incluso, nada. ¿A qué edad o tamaño debería un macho empezar a competir? Si lo hace demasiado pronto, no conseguirá el harén, y si espera demasiado puede no sobrevivir hasta la época de celo —al fin y al cabo, existe el continuo riesgo de morir de hambre, enfermedad o depredación—. Un modelo del juego, en el que se asume que el éxito reproductor de un macho es una función creciente de la proporción de la población en celo que es más pequeña que ella misma, sugiere que puede no existir una EEE. Los machos tendrán cada vez más edad y tamaño al abordar su primera competición, hasta que sean demasiado viejos para procrear. En ese momento pueden ocurrir dos cosas. La población puede ser invadida por «furtivos» que roban hembras sin intentar poseer un harén. Existe un número sorprendentemente alto de especies en las que hay una mezcla de machos que poseen harén y machos furtivos (no está claro si se trata de una EEE mixta estable o una etapa transitoria en un ciclo). También puede ocurrir que la especie se extinga en competencia con otra similar de menor tamaño y más eficiente desde el punto de vista ecológico.

Existe un hecho curioso en relación con los mamíferos que sugiere que el juego del ciervo rojo puede estar diciéndonos algo. La evidencia fósil muestra que la mayoría de los linajes de mamíferos incrementan progresivamente su tamaño; por ejemplo, los primeros caballos no eran más grandes que un perro de tamaño medio. Sin embargo, los mamíferos en su conjunto no son hoy más grandes de lo que lo eran hace cincuenta millones de años (y esto es así incluso ignorando la extinción reciente de muchas especies grandes debido, probablemente, a la caza humana). Una posible explicación es que muchas especies crecieron en tamaño a causa de la competición entre machos, tal como sugiere el citado juego, pero se extinguieron al competir con especies más pequeñas. Pero esto es sólo una especulación.

Examinaremos ahora un tema en el que las predicciones cuantitativas son, en ciertos casos, posibles. Se trata de la evolución de la proporción entre sexos. ¿Por qué, en la mayoría de las especies, hay el mismo número de machos que de hembras y, sin embargo, en algunos casos no es así? (Por ejemplo, ciertas avispas tienen diez veces más hijas que hijos).

La respuesta básica a esta cuestión fue formulada en 1930 por R. A. Fisher. Aunque Fisher no hizo uso de ninguna analogía explícita con un juego humano (al menos, no en este contexto), su razonamiento era, en esencia, el de una EEE y puede ser parafraseado como sigue. Suponiendo que una hembra pudiera elegir el sexo de sus descendientes, ¿cuál debería escoger? Bajo el

punto de vista darwiniano, debería optar por una proporción de sexos que maximice el número de nietos a los que se transmiten sus genes. ¿Cómo hacerlo? Su elección viene determinada por un razonamiento muy simple. Si todo hijo tiene un padre y una madre, los miembros del sexo menos abundante tendrán, *como media*, más hijos. Así pues, debería tener descendientes del sexo menos frecuente.

Es obvio que el único estado estable, o EEE, es aquel con igual número de machos que de hembras. De ahí viene la proporción 1:1. En realidad, Fisher fue más lejos y exploró la posibilidad de que fuera más costoso tener hijas que hijos, o viceversa. Su conclusión fue que los padres emplearían los mismos recursos tanto al producir machos como al producir hembras.

Se trata de una predicción que a menudo se cumple con gran precisión. Pero hay que decir que se debe a que el mecanismo de la determinación del sexo —que generalmente produce igual número de gametos con los cromosomas X e Y— es tal que sólo puede generar una proporción de 1:1. El argumento tiene una validez relativa. La proporción entre sexos es uno de los dos únicos rasgos de la mosca del vinagre que, elegidos entre cientos, han sido sometidos a selección artificial prolongada sin observar efecto alguno. Sin embargo, me cuesta creer que la selección natural no habría alterado este mecanismo de haber merecido la pena. Existen, de hecho, ciertas especies en las que las hembras pueden alterar la proporción de sexos en su descendencia y lo hacen como respuesta

a determinadas circunstancias. En los mamíferos, una hembra puede alterar dicha proporción con un coste relativamente bajo, seleccionando cuál de los blastocitos fertilizados debería implantarse y crecer en su útero. No obstante, el caso más claro de control del sexo de los descendientes por parte de la madre tiene lugar en los himenópteros (hormigas, abejas y avispas). A tal efecto, como en muchos otros insectos, la hembra almacena el esperma tras el apareamiento. Si fertiliza un huevo, se desarrolla una hembra, y si no, se desarrolla un macho con un único conjunto de cromosomas. Los experimentos demuestran que las hembras pueden elegir el sexo de un descendiente individual y que, de hecho, lo hacen. Esta notable característica hizo que el ecólogo Eric Chamov dedicara un artículo científico «A Dios Todopoderoso, por haber creado los himenópteros, permitiéndonos comprobar la teoría de la proporción entre sexos».

¿Qué uso hacen los himenópteros hembra de esa capacidad? Uno de los casos, descubierto por el biólogo matemático Bill Hamilton, se refiere a las avispas parásitas que depositan huevos en las orugas de la polilla. Las larvas se desarrollan dentro de la oruga, matando a su anfitrión, y a menudo se aparean entre ellas en cuanto emergen; los machos mueren poco después y las hembras se dispersan en busca de otras orugas. Si, típicamente, sólo una avispa hembra deposita huevos en una única oruga, ¿qué sexo sería el más conveniente para la prole? Como en el razonamiento de Fisher, las hembras deberían actuar de modo que el número de nietos fuera el

máximo posible. Como un solo macho puede producir suficiente esperma para muchas hembras, deberían engendrar un macho y todo lo demás, hembras. Las proporciones de sexos con fuerte mayoría de hembras se dan, de hecho, en este tipo de parásitos.

Pero las cosas no son así de simples. Supongamos que una segunda hembra depositara huevos en la misma oruga. Podría serle ventajoso engendrar varios machos. Consideremos el siguiente modelo, muy simplificado. Cada avispa hembra deposita la totalidad de sus huevos en una única oruga y toda oruga es parasitada por dos avispas. Parece razonable asumir que el número de huevos depositado por una hembra es constante y vale n (el número se cancela, pero me parece más claro incluirlo en las ecuaciones). El apareamiento en una oruga es aleatorio y cada hembra se aparea sólo una vez. ¿Cuál es la proporción estable de sexos? Se trata de un problema bastante complejo desde el punto de vista matemático; requiere utilizar cálculo diferencial y no resulta obvia la manera de aplicarlo. Siguiendo la filosofía de la EEE, buscamos una proporción de sexos según el cual, si todas las hembras de una población la adoptan, a ninguna hembra mutante que produzca otra proporción le irá mejor —en el sentido de que transmita sus genes a más descendientes—. La respuesta es que las hembras deberían engendrar un macho por cada tres hembras. En otras palabras, la EEE para las hembras en este juego es producir una proporción de sexos con más machos que cuando sólo una de ellas parasitaba una oruga. Si el lector es capaz de deducir esto a partir de principios

fundamentales, puede considerarse todo un biólogo: necesitamos gente como usted.

La aplicación de la teoría de juegos a la evolución de la proporción entre sexos da lugar a predicciones cuantitativas. Pero la comprobación sigue siendo difícil, ya que, como se ha señalado en el anterior modelo, es necesario siempre partir de ciertos supuestos. En la práctica no tiene por qué darse el que toda oruga sea parasitada exactamente por dos avispas, que la mortalidad del parásito resulte independiente de la cantidad de huevos depositados, que el apareamiento sea aleatorio, y así sucesivamente. Por ello, incluso en la teoría de la proporción de sexos, los ensayos son generalmente cualitativos.

La teoría del juego evolutivo puede ser aplicada siempre que la mejor conducta a adoptar por parte de un individuo —su grado de «adaptación»— dependa de lo que los demás hagan. Su gama de aplicaciones es, por lo tanto, muy amplia; por ejemplo, ha sido utilizada no sólo con animales, sino también con plantas e incluso con los elementos genéticos *egoístas* que se replican por su cuenta con el resto del genoma. Un tema que ha suscitado mucho interés en los últimos tiempos es el de la comunicación animal. La cuestión es, en principio, muy simple: ¿Por qué los animales no mienten? Supongamos que, en el juego del halcón y la paloma, un animal pudiera indicar: «voy a luchar». Si la señal fuese sincera, lo más sensato para el oponente sería dejarlo. Hacer la señal sería, pues, una forma económica de obtener el recurso sin pelea. Pronto, todos

señalarían la intención o no de entrar en combate. Y, poco después, ninguno se fiaría de esa señal, con lo que la comunicación se habría acabado. Esta dificultad ha sido abordada, con considerable éxito, tratando la comunicación como un juego asimétrico entre dos personas.

Conviene aclarar que lo que se ve sometido a prueba en estos modelos no es la propia teoría de la evolución mediante la selección natural. Dicha teoría ha de ser comprobada por otros medios. El mejor modo de constatar si la evolución ha tenido lugar es el examen del registro fósil; como J. B. S. Haldane señalaba en cierta ocasión, un simple fósil de conejo entre las rocas del Cámbrico demostraría que la evolución no ha sucedido. La teoría de que el mecanismo de la evolución es la selección natural podría quedar en entredicho si se demostrara que los descendientes no se parecen a sus progenitores (una posibilidad teórica, pero poco verosímil) o que las características adquiridas son a menudo heredadas (lo que podría constituir un mecanismo alternativo). Los modelos de la teoría de juegos asumen la validez de la selección natural y se limitan a comprobar una explicación concreta de la evolución de una característica determinada, ya sea el comportamiento en la lucha, la proporción entre sexos o la coloración de advertencia.

La ventaja de un modelo matemático frente a un modelo verbal es doble. En primer lugar, para definir el modelo se ha de ser absolutamente claro con lo que se está asumiendo. Además, si el autor del modelo ha asumido algo inconscientemente, siempre es

posible para otros, examinando el modelo, comprobar que sólo es válido si se parte de esa premisa inconsciente. Por ejemplo, cuando George Price y yo escribimos por primera vez la matriz de pagos del juego halcón-paloma, no afirmamos explícitamente que asumíamos que el recurso podía ser compartido, pero la matriz implica esto último. Esta faceta de la construcción de modelos es, a mi juicio, muy importante. Muchas veces, al reflexionar sobre un problema biológico, me doy cuenta de que sólo empiezo a entenderlo cuando he creado un modelo matemático. Existen cuestiones sobre las que es muy difícil razonar sin la ayuda de las matemáticas.

La otra misión del modelo, por supuesto, es hacer predicciones que puedan ser comprobadas. Como reiteradamente hemos subrayado aquí, en biología es difícil llegar a hacer predicciones cuantitativamente precisas, debido a la enorme complejidad de las situaciones que tratan de reflejar los modelos. Pero espero haber convencido al lector de que cabe hacer ciertas predicciones cualitativas no obvias y de que, a veces, esas predicciones se cumplen.¹²⁹

¹²⁹ Los temas analizados en este ensayo son tratados con más detalle y de una manera no técnica por K. Sigmund en *Games of Life*, Penguin Books, 1995.

Parte 10

El mejor tiempo posible para vivir

*El mapa logístico**Robert May*

§. I

Soy tan feliz... Estar otra vez al principio, sin saber casi nada... Una puerta como ésta se ha abierto sólo cinco o seis veces desde que comenzamos a caminar erguidos. Es el mejor tiempo posible para vivir, cuando casi todo lo que creías saber resulta estar equivocado.

Tom Stoppard,

Arcadia, Acto 1, Escena 4

Este pasaje de *Arcadia* es, sin duda, un tanto desmesurado, pero refleja cómo nos sentíamos a principios de los setenta en medio de la revolución causada por la teoría del caos.

El drama de Stoppard entreteje brillantemente tres temas: la jardinería, la poesía de Byron y el caos.¹³⁰ Preocupado siempre por el rigor, me pidió que le echara un vistazo al texto para ver si contenía algún error científico (la verdad es que se bastaba solo). Aproveché descaradamente la oportunidad y asistí a los ensayos de la obra a comienzos de 1993 como un aficionado más. Escribí

¹³⁰ *Arcadia*, incluida en *Plays 5*, de Tom Stoppard (Faber & Faber, 1999). En *Tom Stoppard: A Faber Critical Guide* (Faber & Faber, 2000) aparece un útil comentario sobre la obra.

también un artículo para el programa de mano que, sospecho, se ha convertido en mi publicación más leída, aunque no me proporcione citas fuera del reducido círculo de los especialistas.

En *Arcadia*, Stoppard afirma —correctamente, desde mi punto de vista— que todos los conceptos radicalmente nuevos de la teoría del caos podrían haber surgido doscientos años atrás, mucho antes de que se inventaran los ordenadores electrónicos.

Contrariamente a la extendida opinión de que la teoría del caos fue un descubrimiento propiciado por los ordenadores, todo lo que se necesita es papel y lápiz y un montón de paciencia; el ordenador se limita a incrementar —drásticamente, eso sí— la velocidad con la que podemos hacer los cálculos. Fue precisamente con esos materiales de baja tecnología como empecé a trabajar en el caos; por aquel entonces sólo disponía de una de aquellas primeras máquinas de sobremesa que, frente a los estándares actuales, hoy nos parecen antediluvianas. En ciencia, el caos se refiere a la idea de que el comportamiento de algo pueda ser impredecible a todos los efectos, aunque venga descrito por una sencilla ecuación «determinista». (Por determinista entendemos que la ecuación y todos los parámetros que intervienen en ella son del todo conocidos, sin que existan elementos estadísticos o inciertos). Una ecuación así parece predecir con certidumbre el futuro de algo a partir de su estado en un instante inicial.

La existencia de este comportamiento caótico en ecuaciones deterministas sencillas supuso una gran conmoción para los

científicos, imbuidos como estaban de la poderosa visión de Newton y de todos los que le siguieron en el Siglo de las Luces. El mundo newtoniano es ordenado y predecible y está gobernado por leyes y reglas que pueden ser expresadas matemáticamente mediante ecuaciones. Si las circunstancias son lo bastante sencillas —un planeta que se mueve alrededor de un sol, por ejemplo—, el sistema puede comportarse de un modo simple y predecible. Se pensaba que las situaciones verdaderamente impredecibles —la bola de una ruleta, cuyo veredicto, el número ganador, viene determinado por una compleja concatenación en la que participan la mano del *croupier*, el giro de la rueda, etc.— surgían simplemente porque las reglas eran complejas y muy numerosas.

En los últimos treinta años, más o menos, desde el advenimiento de la moderna teoría del caos, esa visión newtoniana ha quedado hecha añicos. Ahora sabemos que las ecuaciones simples pueden generar comportamientos tan complicados como lo más complejo que podamos imaginar. Examinaremos aquí la más simple de esas ecuaciones generadoras de caos, una ecuación tan sencilla que la puede entender un niño (de hecho, mi hija Naomi se topó con ella en una clase de ordenadores mientras estudiaba en una escuela primaria norteamericana).

¿De qué trata, pues, esa ecuación? Piense en un número entre cero y uno; multiplíquelo por su diferencia hasta uno y, seguidamente, multiplique el resultado por una constante, que podemos denominar como queramos, por ejemplo, a . El resultado es otro

número. En términos matemáticos, si llamamos $x_{inicial}$ al número de partida, el número $x_{siguiente}$ que hemos generado en el proceso se puede expresar mediante la ecuación:

$$x_{siguiente} = a x_{inicial} (1 - x_{inicial})$$

Esta ecuación es la protagonista del presente ensayo. Es muy fácil de aplicar; por ejemplo, si $x_{inicial}$ es 0,25 y a vale 10, la ecuación nos dice que $x_{siguiente}$ es $10 \times 0,25 \times 0,75 = 1,875$. Generalmente, al denominar una expresión como ésta, los matemáticos no emplean el término ecuación, sino el de *mapa*, ya que describe el *mapeado* de un número ($x_{inicial}$) sobre otro ($x_{siguiente}$). El caso que nos ocupa suele ser conocido como *mapa logístico*.

El mapa logístico fascina a los matemáticos debido a su asombrosa complejidad. Cuando se toparon con él por primera vez, parecía el paradigma de lo simple: eligiendo un número inicial, generamos otro número; si cambiamos el de partida una fracción diminuta, cabría esperar que el que obtengamos sea siempre ligera y predeciblemente distinto. Al menos, eso era lo que yo esperaba cuando me enfrenté por primera vez al mapa a comienzos de los setenta. Pero pronto me di cuenta de que no siempre es así; para ciertos valores de a , el mapa aporta resultados que semejan ser completamente aleatorios e impredecibles. Tal como finalmente alcancé a entender, y como veremos en este artículo, esto sucede cuando el mapa describe el caos.

Lo que hace que el mapa logístico sea tan atractivo para los científicos es la observación de que puede ser aplicado con éxito a la ecología, la rama de la biología que trata de las relaciones que los organismos tienen unos con otros y con el entorno. En particular, el mapa hace revelaciones notables sobre los cambios que tienen lugar en las poblaciones animales a lo largo del tiempo. El desove de los bancos de salmones, la cantidad de hormigas que se afanan alrededor de su hormiguero y hasta las fluctuaciones en la población de gallos lira en el páramo —el problema que Valentine, el personaje de la *Arcadia* de Stoppard, está estudiando para su tesis doctoral— obedecen el dictado del mapa. Como veremos, la constatación de que el caos puede subyacer en ciertos casos tras el comportamiento a gran escala de las poblaciones animales revolucionó el campo de la ecología.

El caos ha recorrido un largo camino desde sus inicios. Los científicos saben que, sin él, no podríamos comprender una enorme gama de fenómenos. La actividad eléctrica del corazón o el cerebro, el goteo de un grifo, las acumulaciones de vehículos en una autopista o incluso el intrincado comportamiento del átomo de hidrógeno, están todos ellos relacionados con el caos. El propio término *caos*, en su sentido científico, forma parte hoy del vocabulario de mucha gente. ¿Quién no ha oído hablar del «efecto mariposa», la clásica ilustración del modo en que las ideas del caos son aplicables a la predicción meteorológica?

Volveremos a ello más tarde. Antes vamos a examinar cuál era el punto de vista de los biólogos teóricos sobre las poblaciones animales hasta finales de la década de 1960. Después veremos cómo las revelaciones derivadas del mapa logístico revolucionaron el campo y nos enseñaron un nuevo modo de contemplar la naturaleza. Stoppard tiene razón: para quienes nos vimos envueltos en la aventura, la década de 1970 fue «el mejor tiempo posible para vivir».

§. II

Me convertí en biólogo teórico sólo tras una serie de transmutaciones profesionales. Estudiante de ingeniería química en Sydney, a finales de los cincuenta, acabé siendo físico y obtuve el doctorado con una tesis sobre superconductividad. Después pasé un par de años en la Universidad de Harvard, en la división de ingeniería y física aplicada. A comienzos de los sesenta, volví a la Universidad de Sydney para enseñar física teórica, y me convertí en profesor de esa asignatura. A finales de los sesenta, mi papel como miembro fundador de la entidad Responsabilidad Social de la Ciencia en Australia me condujo, casi por accidente, a interesarme por la relación entre complejidad (en el sentido de número de especies y riqueza de la red de interacciones entre ellas) y estabilidad (en el de capacidad para soportar o recuperarse de una perturbación) en ecosistemas y, poco después, a convertirme en profesor de biología en la Universidad de Princeton. Tuve la suerte

de tropezar con el recién nacido campo de la ecología teórica en su *fase romántica*, similar a la de la física teórica en 1920-1930, cuando una serie de cuestiones simples estaban siendo planteadas en el adecuado marco matemático y como consecuencia emergían sorprendentes respuestas.

La ecología es una ciencia joven. Seguramente, el primer texto ecológico es *Natural History of Selborne*, del clérigo Gilbert White, publicado en 1789. El libro va más allá del descriptivismo arrobado que caracteriza otras obras anteriores de historia natural y comienza a plantear cuestiones analíticas como, por ejemplo, cuáles son los mecanismos que gobiernan la abundancia de vencejos o de avispas en las ciudades. El siglo siguiente presencié enormes avances gracias al advenimiento de la teoría de la evolución por la selección natural, de Darwin y Wallace —en mi opinión, el más importante avance en la historia intelectual de la humanidad—. Al describir la «lucha por la supervivencia» que subyace tras la evolución, Darwin empleaba la metáfora de las cuñas en un barril para ilustrar lo que podríamos denominar «competición por los nichos entre especies». Pero nunca cuantificó esas ideas, y los estudios ecológicos se vieron postergados por los evolutivos en aquella época.

La Sociedad Ecológica Británica, la entidad más antigua de esta clase, fue fundada en 1913, mucho después que la mayoría de las sociedades científicas del Reino Unido. La Sociedad Ecológica de América se creó poco después, en 1915. Hasta mediados del siglo

XX, las publicaciones de ambas sociedades eran predominantemente descriptivas y clasificadoras, y centraban su atención sobre todo en las comunidades vegetales. No obstante, hacia los años cincuenta, los ecólogos animales debatían algunas cuestiones teóricas como, por ejemplo, por qué las poblaciones de ciertos mamíferos boreales suelen sufrir cambios periódicos, creciendo y decreciendo cíclicamente con el tiempo. Los estudios matemáticos proporcionaban indicios de unas respuestas muy simplificadas a esas cuestiones: por ejemplo, las poblaciones de las comunidades con un único predador y una única presa tienen una tendencia inherente a comportarse de forma cíclica. En el tercer cuarto del siglo, algunos biólogos teóricos, entre ellos el muy influyente Robert MacArthur, dieron un gran impulso al proceso mediante la combinación de observaciones empíricas y aproximaciones analíticas (a menudo, netamente matemáticas) que establecían líneas de ataque sobre las cuestiones ecológicas.

Así estaban las cosas a mi llegada. Ante todo, me llamó la atención que las ecuaciones que usaban los ecólogos fueran, en sentido trascendente, distintas de las más habituales en física. Las principales diferencias no estaban en la naturaleza de las ecuaciones en sí, sino en que las de la física se proponen dar cuenta exacta de aquello que describen. Por ejemplo, las ecuaciones de la relatividad general de Einstein están diseñadas para describir la deflexión de un haz de luz por parte del Sol con la precisión que queramos; cuanto más exacta sea la información que

introduzcamos en las ecuaciones (para la masa del Sol, la energía del haz de luz, etc.), más exacta será la predicción de la ecuación sobre la deflexión del haz. En biología poblacional, las cosas son, por lo general, muy distintas. En este ámbito de estudio, las ecuaciones se suelen referir a modelos de sistemas vivos demasiado complicados como para poderlos representar en su totalidad mediante ecuaciones como las que les gustan a los físicos.

Los modelos de comunidades biológicas tienden más bien a ser de tipo muy general y estratégico: son caricaturas de la realidad. Y al igual que una buena caricatura capta lo esencial del sujeto representado, pero aceptamos su vaguedad en los detalles intrascendentes, lo más que cabe esperar de las ecuaciones de biología poblacional es que plasmen los puntos clave de la situación que describen. Así pues, las ecuaciones usadas por los biólogos que estudian las poblaciones animales son como un boceto de la realidad y no como esa imagen especular con la que sueñan los físicos. Esto no significa que esas ecuaciones biológicas no sean vitales para nuestra comprensión de la naturaleza. Como el biólogo matemático John Maynard Smith ha señalado, «Las matemáticas sin la historia natural son estériles, pero la historia natural sin las matemáticas es un lío».

Los ecólogos, por ejemplo, habían recopilado datos que mostraban que las poblaciones animales en comunidades aisladas permanecían más o menos constantes o, como decía Maynard Smith en su clásico *The Mathematical Ideas in Biology*, publicado en

1968, fluctuaban «con periodicidad bastante regular». Pero ¿cuál era la causa? Si los modelos matemáticos fueran lo bastante buenos, responderían esta cuestión.

Se puede decir que existían dos líneas de pensamiento sobre los cambios en las poblaciones animales. Por una parte, el australiano Charles Birch creía que la mayoría de las poblaciones naturales son sensibles a efectos externos, de modo que fluctúan ampliamente como resultado de las alteraciones del entorno. Birch y sus colegas solían basar sus ejemplos en poblaciones de insectos, que justamente se comportan así. En el otro bando, otro australiano, John Nicholson, opinaba que las poblaciones son reguladas por efectos que dependen, básicamente, no del entorno, sino de la densidad de población (la cantidad de animales que viven en un espacio dado). Según este punto de vista, las poblaciones tienden a crecer cuando sus densidades son bajas y a disminuir cuando sus densidades son altas; como resultado de ello, en promedio, tienden a ser relativamente constantes. Nicholson y sus partidarios basaban obviamente sus ejemplos en poblaciones relativamente estables.

Parecía, *a priori*, que sólo uno de esos puntos de vista debía ser el correcto. Pero, como a menudo sucede en la ciencia cuando dos enfoques de un problema semejan ser ambos parcialmente correctos, pero totalmente irreconciliables, lo cierto es que muchos de los protagonistas tenían una visión del problema demasiado estrecha. Resultó que la cuestión podía ser resuelta de forma mucho más simple mediante una tercera línea de pensamiento, un

paradigma distinto. La virtud del mapa logístico era que encarnaba de forma fácil y clara este nuevo y productivo modo de pensar, como pronto tuve la suerte de descubrir.

§. III

Lo que hace es que cada vez que obtiene un valor para y , lo usa como nuevo valor de x . Y así sucesivamente. Como una realimentación. Introduce la solución de nuevo en la ecuación y resuelve ésta otra vez. Iteración, ya sabes.

Tom Stoppard,

Arcadia, Acto I, Escena 4

Imaginemos un estanque lleno de peces de colores. Durante sus aisladas vidas acuáticas, esos peces comerán, procrearán, sufrirán posibles enfermedades y traumas impredecibles como la visita del gato. Una cuestión en la que los ecólogos poblacionales están interesados es saber cómo cambiará el número de peces de una generación a la siguiente.

El mapa logístico proporciona un tipo de respuesta a esa cuestión. Para ver cómo, pensemos en el número de peces del estanque como en la fracción del máximo número total de ellos que podrían razonablemente vivir en ese entorno y llamemos x a esa fracción. Por ejemplo, si el máximo número de peces de colores que podría

sostener el estanque fuera mil y sólo hubiera doscientos cincuenta en el momento en que los contamos, x valdría $250/1.000 = 0,25$.

La premisa fundamental de muchas descripciones matemáticas simples como ésta es que la población $X_{inicial}$ de una generación determina unívocamente la población de la generación siguiente, $X_{siguiente}$. Pero ¿en qué forma, matemáticamente hablando, depende $X_{siguiente}$ de $X_{inicial}$? Si aceptamos el modelo más simple, del tipo de los propuestos por el economista y clérigo inglés Thomas Malthus (1766-1834), pionero en estos terrenos, podríamos suponer que la población se incrementa una pequeña fracción cada año, suponiendo que los peces disponen de alimento ilimitado y se reproducen libremente y sin restricciones. De este modo, la conexión entre $X_{siguiente}$ y $X_{inicial}$ sería una ecuación del tipo de $X_{siguiente} = 1,05 X_{inicial}$ donde, en este caso, existe un crecimiento anual de la población de un cinco por ciento. Si, al principio, x vale 0,25, en la generación siguiente valdría $1,05 \times 0,25 = 0,2625$, una generación después $1,05 \times 0,2625 = 0,275625$, y así sucesivamente. La población se incrementaría, pues, de manera gradual.

Pero la vida real no es así. Si la población de peces fuera muy grande, pronto se quedarían sin comida y tendrían que luchar por ella, las enfermedades se propagarían con mayor facilidad y la comunidad constituiría una jugosa presa para los predadores. El resultado sería que la tasa de crecimiento de la población disminuiría. Si, por el contrario, sólo hubiera un puñado de peces retozando en el estanque, con mucho espacio para moverse, su

población crecería rápidamente. ¿Cómo podemos modificar el mapa *malthusiano* $x_{siguiente} = a x_{inicial}$ (donde a es una constante) para hacerlo más realista? Una posible respuesta es el mapa logístico, $x_{siguiente} = a x_{inicial} (1 - x_{inicial})$, que se hizo popular entre los ecólogos que estudiaban poblaciones de peces o insectos en la década de 1950. La constante a representa la tasa de crecimiento y su valor es característico del entorno (el estanque en nuestro ejemplo). El nuevo factor $1 - x_{inicial}$ garantiza que $x_{siguiente}$ no crezca demasiado deprisa, ya que cuando $x_{inicial}$ aumenta, $1 - x_{inicial}$ disminuye, manteniendo bajo control la población de la generación siguiente (si x supera el valor de 1, la población se extingue).

¿Qué predice el mapa logístico para el comportamiento dinámico de la población de peces de colores (y otros fenómenos a los que podría ser aplicable)? En la década de 1950, los expertos en población aplicaron la ecuación no sólo a comunidades de peces, sino a insectos y otros organismos. Pero cayeron en el error de verse condicionados por la costumbre: buscaron (y encontraron) situaciones en las que las poblaciones tendían a un valor estable, a un equilibrio. Investigaron incluso qué valores de la constante a garantizaban esa estabilidad. No se preguntaron qué sucedía cuando dicha constante adoptaba valores *fuera* del margen que hacía que la población resultara estable. El mundo científico constataría poco después que la simplicidad del mapa logístico es enormemente engañosa.

Vamos a ver cómo cambia $x_{siguiente}$ cada vez que la calculamos o, dicho en lenguaje técnico, cómo varía en cada iteración. A tal efecto, elegiremos tres valores de a (en su momento veremos por qué): 2,4, 3,4 y 3,99. Observemos la *figura 10.1*, en la que se ha representado la evolución de $x_{siguiente}$ en cada caso, partiendo siempre de un valor inicial de 0,01. En el primero de ellos ($a = 2,4$), $x_{siguiente}$ se instala enseguida en un valor estable —en el contexto de nuestro ejemplo, esto significa que la población de peces de colores en el estanque se haría constante—. En el caso siguiente ($a = 3,4$), $x_{siguiente}$ oscila continuamente, arriba y abajo, entre dos valores. Nuestra población de peces se repite periódicamente cada dos generaciones. El último caso ($a = 3,99$) es muy extraño: $x_{siguiente}$ salta de un valor a otro todo el tiempo. Se trata del «caos»: la población de peces fluctúa, en apariencia sin ritmo ni motivo, de forma totalmente impredecible.

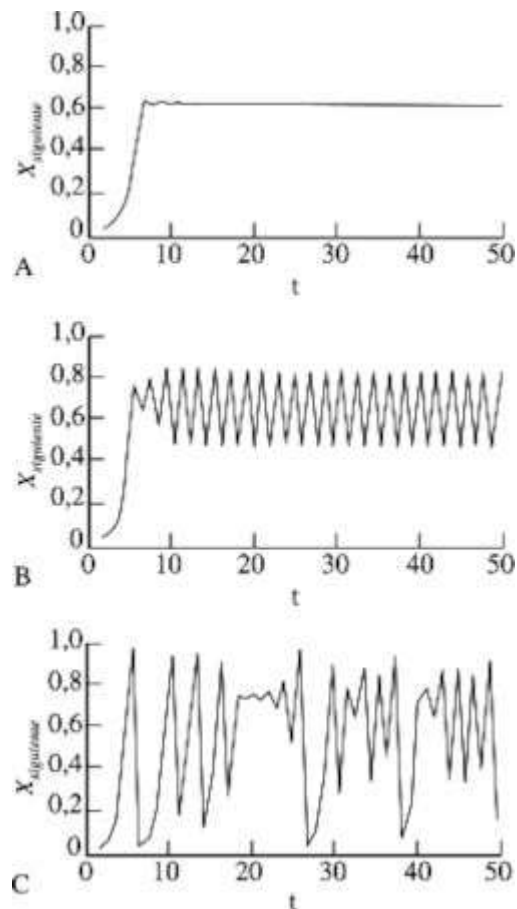


Figura 10.1. La evolución de $X_{siguiente}$, para A, $a = 2,4$; B, $a = 3,4$; C, $a = 3,99$.

A comienzos de la década de 1970, poco después de interesarme por las poblaciones animales, quedé fascinado por el mapa logístico. Quería intentar comprender matemáticamente cómo evolucionaba para *cualquier* valor inicial y para *cualquier* valor de la constante a . El camino era muy arduo y los progresos, muy lentos.

Al concluir el otoño de 1973, recién llegado a Princeton para tomar posesión de mi puesto como profesor, me acerqué a la Universidad de Maryland para dar un seminario. Llevé conmigo parte del trabajo

que había desarrollado sobre el mapa logístico y, con él, ciertas cuestiones sin resolver. En el seminario encontré a Jim Yorke, un matemático que llegaría a convertirse en un buen amigo y con el cual colaboraría en el estudio del mapa logístico.

La historia es más fácil de contar si empiezo por el final. Obsérvese la *figura 10.2*, que muestra el mapa en todo su complejo esplendor. En la parte inferior, a lo largo del eje horizontal, se representan los valores de la constante a ; en el eje vertical, el valor $x_{estable}$ al que tiende $x_{siguiente}$ al cabo de algunos miles de iteraciones, para cada valor de a . Cuando la constante es inferior a tres, $x_{siguiente}$ tiende hacia un valor único (como sucedía en la *figura 10.1 A*). Sin embargo, a medida que a crece por encima de tres, $x_{estable}$ no tiene uno, sino *dos* valores posibles: se producen oscilaciones sostenidas entre ambos (como en la *figura 10.1 B*). Conforme incrementamos la constante a , vemos emerger una serie de comportamientos periódicos en lo que podríamos denominar «una cascada de duplicaciones del periodo». Finalmente, cuando la constante se encuentra entre 3,57 y 4, el mapa se comporta de manera imprevisible. Nos hallamos en el dominio del caos, donde el valor estable es tan sensible al más mínimo cambio en el valor inicial de x , que la distribución final puede considerarse aleatoria (en la *figura 10.1 C* observamos este comportamiento).

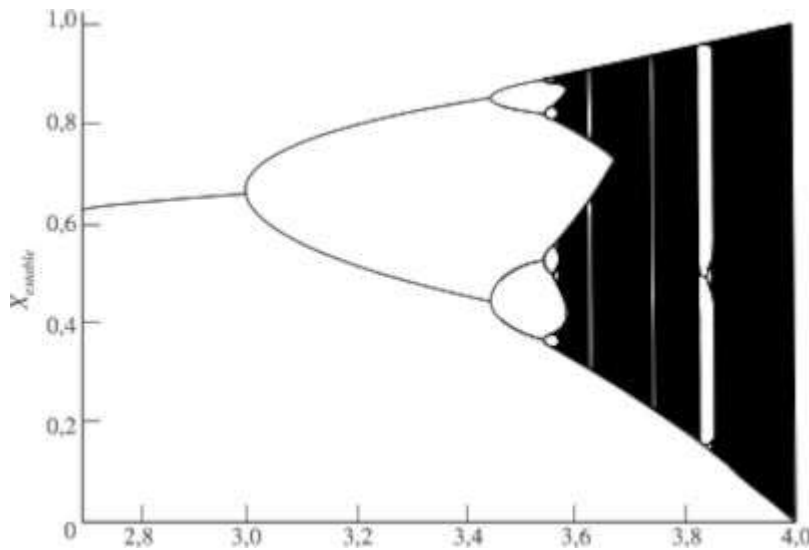


Figura 10.2. Modo en que el valor de a determina el valor estable de X .

Detengámonos un momento a pensar qué significa esto para la población de peces en el estanque. La constante a refleja cuánto cambia la población entre una generación y la siguiente. Su valor, distinto para cada situación, depende del propio pez (su fertilidad, necesidad de alimento, agresividad, visitas de los gatos hambrientos, etc.) y del entorno (la comida disponible en el estanque, sus condiciones climáticas y de salubridad, etc.). Si la constante es baja (< 3), la población será estable. Si se halla entre 3 y 3,57, oscilará periódicamente entre dos valores extremos. Y si está entre 3,57 y 4, la población fluctuará ampliamente, por lo que será imposible hacer predicciones a largo plazo, aunque la ecuación sea muy simple y totalmente determinista.

Cuando llegué al seminario de Maryland, ya conocía la parte izquierda de la *figura 10.2* (las áreas de estabilidad y la cascada de

duplicaciones del periodo). Cuando llegué al punto del discurso en el que afirmé que desconocía lo que sucedía cuando a superaba 3,57, Jim Yorke me interrumpió. «Sé lo que viene después», dijo. Tien-Yien Li y él habían investigado recientemente mapas del tipo del mapa logístico y habían descubierto su comportamiento caótico. De hecho, ambos habían acuñado el término *caos* en su sentido matemático en un artículo titulado «Periodo tres implica caos», que sería publicado en 1975.¹³¹ Varios de sus colegas les aconsejaron elegir una palabra más discreta que *caos*, pero ellos siguieron adelante y acabaron dando a esta rama científica su llamativo nombre. Yorke y Li no habían estudiado los valores de a por debajo de 3,57, con lo que no habían apreciado la cascada de duplicaciones del periodo que caracteriza el camino hacia el caos. (Toda una pléyade, casi mágica, de duplicaciones de periodo adicionales —por ejemplo, ciclos de periodo 11 que se dobla sucesivamente a 22, 44, 88...— se oculta en las complejas profundidades del caos, tanto en el mapa logístico como en algunos de sus parientes).

Juntando las piezas de ese rompecabezas que habíamos tratado de resolver por separado, Jim Yorke y yo nos dimos cuenta de inmediato de que habíamos descubierto algo importante. El comportamiento del mapa logístico no era un simple capricho matemático, sino que tenía implicaciones profundas en las predicciones hechas por modelos matemáticos sencillos. Para

¹³¹ T. Y. Li y Jim Yorke, «Period three implies chaos», *American Mathematical Monthly*, 1975, 82, págs. 985-992. El artículo demostraba que en muchos mapas —como el mapa logístico— de periodo 3 (que se repiten a ellos mismos cada tres ciclos), existirán órbitas con cada periodo, así como un número infinito de órbitas caóticas: órbitas irregulares, sin periodo fijo.

nuestra sorpresa, pronto supimos que otros habían pisado ya el mismo terreno que nosotros, comenzando por el matemático finlandés Pekka Myrberg casi veinte años antes, en 1958. Pero Myrberg —y otros pioneros posteriores en la Unión Soviética, Francia y Estados Unidos— contemplaban los patrones con la idea de que eran fascinantes fenómenos matemáticos, sin plantearse el que tuvieran que ver con una descripción del mundo natural. Por el contrario, Jim Yorke y yo explorábamos el mapa logístico en el contexto específico de los problemas ecológicos reales, tratando de comprender el origen de las fluctuaciones poblacionales, no sólo en los estanques con peces de colores, sino en otras situaciones que podían ser descritas mediante el mapa. De este modo, aunque otros habían investigado el mapa y descubierto cosas sobre su comportamiento, nosotros dos fuimos los primeros en comprender la gran trascendencia y significado de sus resultados.

Ciertamente, no llegamos a captar todos los aspectos que involucraba el mapa. Se nos escapó un importante detalle que alguien más perspicaz que nosotros fue capaz de observar. Para entender este punto, echemos de nuevo un vistazo a la *figura 10.2*, que representa el comportamiento del mapa, y, más concretamente, a la cascada de duplicaciones del periodo. Se parece a un árbol relativamente simétrico, colocado de lado y con las ramas cada vez más juntas. De hecho, a medida que crece el número de iteraciones, la distancia entre dos bifurcaciones sucesivas —medida como diferencia entre los valores de la constante a para los que dichas

bifurcaciones tienen lugar— es un número fijo. El matemático George Oster, de la Universidad de California en Berkeley, descubrió este hecho en 1976 y nosotros dedujimos una fórmula aproximada para la relación entre los valores de la constante asociados a dos bifurcaciones consecutivas que predecía un valor de 4,83.¹³² Para nosotros fue un mero detalle matemático que olvidamos apenas concluimos los trabajos.

Pero el matemático norteamericano Mitchell Feigenbaum, que trabajaba en el Laboratorio de Los Alamos en Nuevo México y había observado el mismo hecho de manera independiente —basándose en estudios numéricos en vez de en análisis matemáticos—, fue más allá. Ante todo, sus estudios arrojaron un valor más preciso para la citada relación (alrededor de 4,6692), y mostraron que este número afloraba en muchos sitios en los que un sistema hace una transición desde un comportamiento estable a uno caótico. Y lo que es más importante, Feigenbaum sugería que, si la transición a la turbulencia desde un patrón de flujo uniforme en un fluido constituía en esencia una transición al caos, debía poderse observar la cascada de duplicaciones del periodo y medir la relación que él había predicho. Los experimentadores no tardaron en confirmar este extremo y Feigenbaum obtuvo un merecido y amplio reconocimiento por sus trabajos. Esta vez me correspondía a mí estar entre los primeros en descubrir matemáticamente algo cuya verdadera importancia *descubren* otros más tarde.

¹³² El valor exacto era $2(1 + \sqrt{2})$.

Los estudios sobre el mapa logístico revolucionaron el modo en que los ecólogos explicaban las fluctuaciones de las poblaciones animales. Recordemos que, durante mucho tiempo, se había argumentado que se debían a efectos externos (según el enfoque de Birch) o a la densidad de población de los propios organismos (en el enfoque de Nicholson). Ante los resultados del mapa logístico, quedaba claro que la controversia Nicholson-Birch carecía de sentido. Ambos bandos habían errado el tiro: los efectos debidos a la densidad de población, si son lo bastante fuertes (v. *figura 10.1 C*), pueden resultar idénticos a los originados por perturbaciones externas. El problema no es decidir si las poblaciones están reguladas por efectos dependientes de la densidad (y, por lo tanto, tienden a estabilizarse) o son gobernadas por ruido externo (y, por ello, tienden a fluctuar). No se trata de adoptar una teoría u otra; en lugar de esto, cuando los ecólogos observan una población que fluctúa, deben averiguar si las fluctuaciones son causadas por sucesos ambientales externos (p. ej., cambios erráticos en la temperatura o la pluviosidad) o por la propia dinámica caótica inherente a la ecuación determinista que gobierna el desarrollo de la población.

Los ecólogos deben contemplar los fenómenos poblacionales desde una nueva óptica —alguien podría hablar de un nuevo paradigma kuhniano— que arrincona las anteriores. Es en momentos como éste cuando los antiguos supuestos son derrocados y sustituidos

por otros, cuando la ciencia resulta más apasionante para quienes tienen la suerte de estar en primera línea.

Pero esta revolución no triunfó de la noche a la mañana. Supuso una gran labor de persuasión por parte de muchos colegas, pues no resultaba fácil aceptar las profundas implicaciones que la teoría del caos tiene para la ciencia. A principios de 1976, decidí escribir un artículo divulgativo en el que sugería la gran trascendencia del caos. En él hacía una panorámica de la teoría y presentaba el mapa logístico como ejemplo del modo en que el caos puede estar agazapado tras las ecuaciones más simples, confiando en convencer a otros científicos para que analizaran el impacto del caos en sus respectivas áreas. Redacté el artículo en un estilo deliberadamente mesiánico y lo envié a la importante revista científica británica *Nature*. El equipo editorial de la revista se mostró escéptico, opinando, entre otras cosas, que el artículo era excesivamente matemático y que su contenido no les parecía de interés general. Pero una de las editoras principales, Miranda Robertson, debió pensar de otra manera y envió el manuscrito a John Maynard Smith para ser revisado. Su veredicto fue muy generoso («como si lo hubiera escrito tu madre», me dijo Miranda), así que *Nature* publicó el artículo en junio de 1976.¹³³ El artículo consiguió su objetivo de llevar el caos a una amplia audiencia de científicos y hasta la fecha ha sido citado varios millares de veces.

¹³³ Robert May, «Simple mathematical models with very complicated dynamics», *Nature*, 1976, vol. 261.

Más importante que la opinión de los científicos es la opinión de la naturaleza, que se manifiesta a través de los experimentos. Desde los primeros artículos sobre el mapa logístico, varios experimentadores han demostrado hasta qué punto la descripción que hace el mapa del comportamiento dinámico de las poblaciones animales se ajusta a la realidad —no sólo en los peces de colores, sino también en insectos y mamíferos—. En un ejemplo, los datos sobre los cambios de población del ratón campero *Clethrionomys rufocanus* desde 1922 a 1995 en Hokkaido (la isla más septentrional de Japón), presentaban de forma espectacular muchas de las características esperadas a partir del mapa logístico.¹³⁴ De igual modo, los datos clásicos sobre el lince canadiense y la liebre ártica se han visto iluminados por posteriores desarrollos basados en estas ideas. Los experimentadores han estudiado también poblaciones de insectos en condiciones controladas de laboratorio y hallado el tipo de comportamiento que cabría esperar basándose en ecuaciones similares a la del mapa logístico.

Las aplicaciones de la teoría del caos se extienden más allá de la ecología. En los últimos diez años se ha hecho evidente que el caos tiene relación con casi todas las áreas de la ciencia y la tecnología.¹³⁵ Los ingenieros mecánicos utilizan sus ideas para reducir el ruido que producen los frenos y las ruedas de los vagones de tren. Los diseñadores de barcos las usan para evitar que las

¹³⁴ Robert May, «The voles of Hokkaido», *Nature*, 1998, vol. 396.

¹³⁵ *The Nature of Chaos*, en: Tom Mullin, ed., Oxford University Press, 1993. Véase también: J. M. T. Thompson y S. R. Bishop, eds., *Nonlinearity and Chaos in Engineering Dynamics*, Wiley, 1994.

embarcaciones vuelquen a causa de una tormenta. Los ingenieros eléctricos se basan en ellas para cifrar los mensajes, para extraer información de señales ruidosas y para prevenir fallos de alimentación. Los astrónomos las tienen en cuenta a la hora de estimar la distribución de asteroides en el sistema solar. Los físicos las aprovechan para entender y predecir el movimiento de los fluidos. En definitiva, el caos se está convirtiendo en parte importante de la ciencia del siglo XXI.

§. IV

Lo impredecible y lo predeterminado evolucionan Juntos, haciendo que cada cosa sea como es. Éste es el modo en que la naturaleza se crea a sí misma, a todas las escalas, desde el copo a la tormenta de nieve.

Tom Stoppard,

Arcadia, Acto I, Escena 4

¿Por qué los científicos tardaron tanto en caracterizar el caos? Newton creó la moderna ciencia matemática a finales del siglo XVII; ¿por qué ni él ni sus sucesores estudiaron cosas tan simples como el mapa logístico y descubrieron su rica estructura?

Pienso que la respuesta es que, desde los tiempos de Newton, los estudios matemáticos sobre el cambio se centraron casi exclusivamente en la dinámica de sistemas que varían de forma

continúa con el tiempo. Desde las leyes newtonianas del movimiento y a través de los posteriores avances en la comprensión de los sistemas dinámicos debidos a brillantes matemáticos como Joseph Lagrange y Sir William Hamilton, la atención se centró en las ecuaciones *diferenciales*: ecuaciones con variables que cambian de manera continua (y que representan, por ejemplo, distancias que varían uniformemente y no a saltos discretos, como la graduación de una regla). Si cualquiera de esos matemáticos hubiese tropezado con el mapa logístico y le hubiera dedicado algún tiempo, estoy seguro de que habría descubierto el caos.

Pero no fue así. Habría que esperar hasta finales del siglo XIX, fecha en la que el caos fue vislumbrado por primera vez por el gran matemático francés Henri Poincaré, al estudiar ciertas ecuaciones diferenciales. A finales del siglo XIX, el rey Oscar II de Suecia ofreció un premio a la primera persona que pudiera demostrar que el sistema solar en su conjunto (el Sol, los planetas, los asteroides, etc.) era totalmente estable. Fue al intentar conseguir este premio cuando Poincaré abordó el «problema de los tres cuerpos»: tres objetos que interaccionan gravitatoriamente (p. ej., el Sol, la Tierra y la Luna), tratados de forma simplificada como si fueran tres puntos. Poincaré demostró que el sistema de ecuaciones diferenciales resultante podía dar lugar a órbitas de «indescribible complejidad» y concluyó que el problema planteado por el rey era irresoluble, al menos mediante las técnicas disponibles entonces. Tenía razón y, además, fue el primero en entrever el caos, aunque pocos se dieran

cuenta en aquel momento. En cualquier caso y para su satisfacción, fue el ganador del premio.¹³⁶

El estudio del caos apenas progresó en la primera mitad del siglo XX, aunque hubo algunos científicos que tropezaron con él sin llegar a apreciar su significado. Por ejemplo, los matemáticos Mary Cartwright y John Littlewood mencionaban en un artículo de los años treinta haber encontrado ejemplos de ecuaciones diferenciales relativamente simples que presentaban un comportamiento extraordinariamente complejo, lo que hoy denominaríamos caótico. Pero la opinión generalizada era que dichos ejemplos constituían complejos e intratables casos particulares; era mejor devolverlos al cajón y olvidarse de ellos. No eran la clase de cosa en torno a la que organizar todo un programa de investigación.

En realidad, la moderna teoría del caos se inició a partir de un conjunto de ecuaciones relativas a la predicción meteorológica, publicadas en 1963, un año notable no sólo para la ciencia, sino también —como Philip Larkin señalaba— para los Beatles y la revolución sexual. Las ecuaciones eran obra del gran meteorólogo Edward Lorenz, del Instituto Tecnológico de Massachusetts. A Lorenz le habían fascinado durante muchos años las veleidades del tiempo y, como muchos otros, quería averiguar si algún día sería posible predecirlo con la misma precisión que podemos establecer la llegada del cometa Halley. En aquella época, la predicción del tiempo local con un horizonte de más de unos pocos días parecía

¹³⁶ Ian Stewart, *Does God Play Dice?: The New Mathematics of Chaos*, Viking Penguin, 1997.

algo alcanzable mediante el empleo de ordenadores más potentes y una mejor información sobre las condiciones meteorológicas iniciales, obtenida a partir de satélites.

Lorenz trabajó con ciertas ecuaciones meteorológicas «de juguete», que caricaturizaban la evolución de la meteorología —ecuaciones que especificaban la variación en el tiempo de tres magnitudes meteorológicas—. Le sorprendió constatar que sus ecuaciones tenían en común una notable propiedad: sus soluciones eran extremadamente sensibles a las condiciones iniciales.¹³⁷ Si un conjunto de esas condiciones conducía a cierto resultado, otro conjunto —incluso si la diferencia entre ambos era infinitesimal— llevaba, al cabo de poco tiempo, a un resultado totalmente diferente. La razón era que sus ecuaciones sencillas se comportaban de forma caótica. Representaban algo que, hoy sabemos con certeza, es imposible predecir a largo plazo (en la práctica, más allá de unos siete a veinte días, dependiendo de los detalles de las condiciones actuales).

Las ecuaciones de Lorenz son ecuaciones *diferenciales* que representan *cambios* con respecto a un tiempo cronológico que avanza de forma continua, como la manecilla de las horas de un reloj y no como la de los segundos, que avanza a saltos. Este punto supone un fuerte contraste con el modo en que se trata el tiempo en el mapa logístico, que asemeja una serie de instantáneas tomadas en momentos concretos. Los científicos estaban mucho menos

¹³⁷ Edward Lorenz escribió un asequible informe sobre sus trabajos en *The Essence of Chaos*, University of Washington Press, 1993.

acostumbrados a los mapas discretos que a las ecuaciones diferenciales, que hasta entonces se pensaba eran dóciles y completamente predecibles. Por ello constituyó una gran conmoción el hecho de que, incluso ecuaciones diferenciales sencillas que *a priori* parecían dominadas en todos sus detalles, se comportaran de manera impredecible.

Durante casi una década, la obra de Lorenz apenas tuvo trascendencia fuera del pequeño grupo de científicos interesados por la meteorología. Uno de los motivos es que tratar con sus ecuaciones requiere una gran destreza matemática, así como una considerable habilidad con los ordenadores para representar su comportamiento de forma visual. Algunos matemáticos cuestionaron incluso que las ecuaciones de Lorenz exhibieran realmente un comportamiento caótico, sugiriendo que el aparente caos podría ser un efecto colateral de las aproximaciones numéricas empleadas al estudiarlas. No sería hasta 1999 cuando el aspirante a doctor en física por la universidad sueca de Uppsala Warwick Tucker demostró de manera rigurosa que las ecuaciones de Lorenz eran definitivamente caóticas.¹³⁸

En mi opinión, las ecuaciones de Lorenz podían haber pasado a la historia como una simple muestra de los misterios de la meteorología si no hubiera sido por la labor evangelizadora de Jim Yorke y otros a comienzos de los setenta. La tarea de convencer a sus colegas científicos de la importancia del caos fue mucho más

¹³⁸ Ian Stewart, «The Lorenz attractor exists», *Nature*, 2000, 406, págs. 948-949.

fácil una vez se conoció el comportamiento del mapa logístico —se trataba de un ejemplo enormemente simple y fácil de comprender, aunque fuese un mapa discreto del tipo de los que los científicos usaban raramente—. Pero la obra de Lorenz demostraba que el fenómeno del caos se da también en ecuaciones *diferenciales*, las cuales tratan con magnitudes que cambian de modo uniforme y continuo. Este último hecho fue el que verdaderamente llamó la atención de la comunidad científica y la hizo despertar y tomar nota. A principios de la década de 1980, se hizo habitual para los científicos contemplar con nuevos ojos sus trabajos y, mediante ordenadores cada vez más potentes, comprobar si habían obviado el caos en los fenómenos que habían estado estudiando.

Para muchos científicos, el carácter impredecible de algunas de las ecuaciones de la ciencia clásica fue toda una revelación. Desde la década de 1920, se sabía que la impredecibilidad es un componente clave de la teoría cuántica, que describe los mundos atómico y subatómico.

Los teóricos cuánticos saben que sólo cabe predecir la probabilidad de los posibles comportamientos de un electrón atómico. Pocos esperaban que esa impredecibilidad estuviera agazapada tras las sencillas ecuaciones que ellos y sus predecesores habían estado utilizando durante doscientos años.

La revelación dio pie a algunos enfoques fascinantes. A comienzos de la saga del caos, mi amigo Henry Horn, ecólogo de la Universidad de Princeton, sugería que podríamos estar ante la reconciliación

entre el libre albedrío y la predestinación del hombre que propugnan ciertas religiones. El Creador nos habría ubicado en un mundo de caos determinista, un mundo que obedece a leyes definidas sin elementos dejados al azar, pero sólo Él conoce las condiciones iniciales exactas que determinan lo que nos depara el futuro. Para nosotros, la sensibilidad del sistema a las condiciones iniciales se traduce en que aquél es impredecible, y esto lo asociamos al libre albedrío. Horn hizo su sugerencia medio en broma, medio en serio, pero actualmente goza de cierto predicamento.¹³⁹

§. V

No fueron sólo los científicos quienes se sintieron afectados por el caos a comienzos de los ochenta; el público en general también se mostró interesado. El «efecto mariposa» se convirtió en un cliché; se trata de una frase que se empezó a repetir tras una conferencia que Edward Lorenz dio en Washington capital, en 1972, titulada «¿El batir de alas de una mariposa en Brasil desencadena un tornado en Texas?». La charla no matemática de Lorenz acerca de la impredecibilidad de la meteorología terrestre llamó la atención sobre la extrema sensibilidad del sistema climático de nuestro planeta, que sólo parecía abordable en el contexto de las ideas del caos. No obstante, la noción del efecto mariposa no era nueva, como el propio Lorenz subrayó. Aparecía, por ejemplo, en «El ruido de un trueno»,

¹³⁹ Una referencia útil es «Randomness and perceived randomness in evolutionary biology», de W. C. Wimsatt. *Synthese*, 1980, vol. 43, págs. 287-329.

un impactante relato breve escrito por Ray Bradbury en 1952, mucho antes de la conferencia de Washington. En el relato, la muerte de una mariposa prehistórica altera el resultado de unas elecciones presidenciales.

La popularidad de la frase se debe probablemente a la conocida obra de James Gleick *Chaos*,¹⁴⁰ cuyo primer capítulo se titula precisamente «El efecto mariposa». Este extraordinario libro, publicado en 1988 y convertido hoy en un clásico de la literatura científica, llevó el caos no sólo al gran público, sino también a muchos científicos que no habían oído hablar de él. Bajo mi punto de vista, la obra de Gleick tiene tres grandes virtudes. En primer lugar, constituye una panorámica convincente y asequible de una nueva y difícil rama de la ciencia. En segundo lugar, hace un uso muy efectivo de un elenco de pintorescos personajes para dar vida a la historia. Y por último, y muy importante, transmite un sentido real de la naturaleza del progreso científico, con todas sus existenciales e imprevistas complejidades. Para mí, *Chaos* proporciona una visión de la ciencia más trascendente y viva que la que se puede hallar en toda la filosofía formal de Karl Popper.

El único punto débil del libro de Gleick es que la importancia relativa atribuida a los protagonistas del drama no es del todo correcta. A Jim Yorke, por ejemplo, se le atribuye demasiada poca y a los pintorescos muchachos de Santa Cruz, demasiada. Creo que su descripción de mi papel en los hechos es bastante correcta.

¹⁴⁰ James Gleick, *Chaos*, Heinemann, 1988. [Trad, esp.: *Caos: La creación de una ciencia*, Barcelona, Seix Barral, 1987].

Todas esas inexactitudes irritaron a algunos expertos, al tiempo que quienes veían su obra magnificada en él aplaudían. Por supuesto, para la mayoría de la gente, lo importante es «la trama argumental», pero los actores le dan mucha importancia a los títulos de crédito. Raramente un libro de divulgación científica ha cautivado la imaginación del público de un modo tan efectivo y con una idea tan nueva. Poco después de su publicación, el caos se convirtió en un tema popular, algo de lo que los líderes de opinión tenían que saber. Incluso Al Gore contrató a un matemático para que le enseñara los rudimentos de la teoría, tras perder las elecciones primarias como candidato demócrata a la presidencia.

Los artistas también se vieron atraídos por la idea del caos. Quedó plasmada en muchas obras de arte visual y en multitud de novelas.¹⁴¹ En demasiadas ocasiones en que la teoría aparece citada en la cultura popular, ha sido confundida con la vieja y trivial observación de que las cosas son más complicadas de lo que uno piensa. Éste es el tipo de caos del filme *Parque Jurásico*, de Steven Spielberg, cuyo guión proviene de una novela de Michael Crichton. Desde que comienza la película estoy deseando que los dinosaurios se zampen al insufrible matemático.

Pero sería un error afirmar que todas las referencias al caos en el arte son superficiales. Para concluir, permítanme regresar a la excelente *Arcadia*, que Stoppard escribió después de que el libro de Gleick cautivara su interés. He aquí las palabras de Valentine sobre

¹⁴¹ *Harriett Hawkins*, *Strange Attractors: Literature, Culture and Chaos Theory*, Prentice Hall, 1995.

la impredecibilidad fundamental de la naturaleza, una de las enseñanzas más profundas del mapa logístico:

«Nos es más fácil predecir sucesos en el límite de la galaxia o en el núcleo de un átomo que si la lluvia va a arruinar o no la fiesta que mi tía va a dar en su jardín dentro de tres domingos. [...] No podemos siquiera anticipar cuándo caerá la próxima gota de un grifo al hacerse irregular el goteo. Cada gota establece las condiciones para la siguiente; la más pequeña variación hace la predicción añicos. El tiempo es impredecible del mismo modo y siempre lo seguirá siendo». (*Acto I, Escena 4*).

Parte 11

Un cuento de hadas medioambiental

*Las ecuaciones de Molina-Rowland y el problema de los CFC**Aisling Irwin*

Cierta fotografía tomada desde una nave *Apolo* captaba el sentir de la Tierra en la década de 1970. La imagen fue tomada en un fugaz y luminoso momento y revelaba la belleza de nuestro planeta como nunca antes se había contemplado. Evidenciaba su soledad: un oasis azul flotando en la negra inmensidad del cosmos. Pero, sobre todo, el planeta parecía enormemente frágil: desde la perspectiva del espacio, sus habitantes tenían que compartir un fuerte interés en preservar su delicado hogar.¹⁴²

La imagen constituía un símbolo del sentimiento, nacido en la década anterior, de que la humanidad poseía ya la capacidad de destruir su entorno y, como consecuencia de ello, destruirse a sí misma. El mensaje de universalidad de las misiones *Apolo* exhortaba a los humanos a contemplarse a ellos mismos como «pasajeros todos de la nave Tierra».¹⁴³ Ese sentimiento fue el que llevó a decenas de millones de personas a manifestarse en protesta contra las agresiones a la naturaleza en el primer Día de la Tierra, en 1970.

¹⁴² Denis Cosgrove analiza la influencia de esta imagen, tomada desde el *Apolo 17*, en *Apollo's Eye: A Cartographic Genealogy of the Earth in the Western Imagination* (John's Hopkins University Press, 2001).

¹⁴³ Palabras de un dirigente del *New York Times* el 25 de diciembre de 1968, en las que afirmaba que los vuelos *Apolo* transformarían nuestra forma de pensar sobre nuestro sitio en el universo tan profundamente como la revolución copernicana lo hiciera en su día.

Durante ese despertar colectivo de la conciencia ambiental se publicaron unas breves líneas que iban a tener más impacto sobre nuestra forma de ver el medio ambiente terrestre que cualquier imagen cósmica.¹⁴⁴ Se trataba de una prosa sin palabras, escrita en símbolos de otro lenguaje. Profetizaba una calamidad global, confirmando que los humanos estaban dañando uno de los sistemas que dan soporte a la vida en nuestro planeta. Con sublime concisión, esas ecuaciones químicas describían la destrucción de la capa de ozono.

Las líneas eran, en parte, deudoras del sentir de la época y, a la vez, conformaban ese sentir. Desde el punto de vista político, dieron comienzo a una era en la que los *pasajeros* de la Tierra se veían obligados a negociar entre ellos para defender su hábitat. Desde el punto de vista científico, ampliaron los límites de las disciplinas empleadas en los proyectos internacionales de investigación que combinaban numerosos enfoques para intentar comprender el más complejo de los ciclos naturales. Finalmente, desde el punto de vista medioambiental, nos proporcionaron dos símbolos: el de la vulnerabilidad de la Tierra bajo el dominio del hombre y, viceversa, el de la capacidad del ser humano para evitar la catástrofe tecnológica.

La historia de esas ecuaciones se extiende a lo largo de un periodo de casi medio siglo, desde 1930 hasta mediados de la década de 1980. Es, en parte, el relato de la comprensión científica de la

¹⁴⁴ Mario J. Molina y F. S. Rowland, *Nature*, 1974, 249, págs. 810-812.

atmósfera (que antaño se pensaba simple e inerte y que hoy se sabe es un tumulto de miles de sustancias que interaccionan). Fue a lo largo de ese relato cuando las cuestiones acerca del ozono comenzaron a emerger. El proceso de darles respuesta contribuyó a forjar la idea del planeta como un sistema único: largas y proliferantes cadenas de relaciones causa-efecto que enlazan todo, desde los microbios del suelo hasta los remotos gases de la estratosfera.

Para entender cómo los científicos han podido extraer sus conocimientos sobre el ozono de entre la maraña de gases de la atmósfera, debemos tener una idea acerca del modo en que trabajan los químicos. Históricamente, el destino de los químicos ha sido buscar, entre las confusas manifestaciones de la materia, la esencia de ésta; atravesar el mutable mundo material en busca de lo inmutable; hallar lo que es permanente y predecible, encontrar las reglas de esa materia. La química es considerada a menudo el pariente pobre de la familia científica, contemplada erróneamente como una ciencia puramente descriptiva sin el *glamour* de la física o la biología. El químico parece hallarse lejos de la lucha del físico con las fuerzas y partículas fundamentales o del pensamiento puro y abstracto del matemático. Sin embargo, eran los químicos quienes poseían las herramientas con las que buscar en el crisol de las reacciones atmosféricas y hallar la que constituía la clave. Fueron capaces de expresarla usando un lenguaje simple y simbólico que ha llevado siglos crear. Los químicos supieron predecir

interacciones que tenían lugar 50 kilómetros por encima de la Tierra —sin ir hasta allí— e incluso determinar la velocidad de esas reacciones. Y, en cualquier caso, los químicos demostraron todo el poder de su ciencia al combinarla con otras disciplinas para producir modelos de la atmósfera y hacer predicciones que se verían confirmadas en las décadas siguientes.

Pocas ecuaciones han expresado mejor la relación entre el ser humano y su entorno o han producido un efecto tan dramático en la opinión pública. Ningún otro químico ha alumbrado un trabajo que, en su brevedad, haya sido considerado potencialmente como nuestra «salvación de la catástrofe medioambiental». Éstas fueron las palabras del comité que en 1995 concedió a Mario Molina, Sherry Rowland y Paul Crutzen el Premio Nobel de Química por sus estudios sobre la destrucción del ozono. Era la primera vez que el Premio Nobel reconocía una investigación sobre el impacto humano en el medio ambiente.

La existencia de algo capaz de sufrir daño en la atmósfera es una idea moderna. El aire y su inmutabilidad se habían dado siempre por supuestos. Desde la antigüedad, se pensaba que el aire era inerte: la química del mundo tenía lugar más abajo. La idea de que existe un tercer estado físico, además de los de sólido y líquido —de que alrededor y encima de nosotros podía haber diversos gases interaccionando entre ellos, con la Tierra debajo— fue una innovación del siglo XVIII. Tras este avance conceptual, los científicos han ido haciendo cada vez más compleja su visión de la

atmósfera. Las ecuaciones químicas en su forma moderna han proporcionado un modo de expresarla.

Las características de los gases atmosféricos individuales, tales como el dióxido de carbono, el oxígeno y el nitrógeno, sólo empezaron a ser perfiladas a partir de 1750. En el siglo XX, el estudio de la atmósfera condujo al descubrimiento de nuevas técnicas para desvelar sus rincones más ocultos. Hoy visualizamos esa atmósfera como una serie de capas esféricas de *aire*, progresivamente más tenues, que protegen a la Tierra de las amenazas procedentes de un espacio frío, sin oxígeno y plagado de radiación. La primera aloja la mayor parte de la actividad humana: es en ella donde vivimos, vuelan nuestros aviones y tienen lugar los fenómenos meteorológicos. Esos primeros 10-15 kilómetros son conocidos como troposfera. Más arriba, allí donde los reactores supersónicos hacen breves incursiones, está la capa siguiente, la estratosfera. Las esferas siguientes están virtualmente vacías y se desvanecen tras pocos cientos de kilómetros, en lo que constituye la frontera más próxima del espacio.

Pero el uso de una división tan simple es engañoso. La Tierra es, a la vez, espectador y actor de una compleja representación atmosférica. Miles de diferentes sustancias revolotean sobre el globo. Flotan a merced del calor y el frío, el día y la noche, las presiones crecientes y decrecientes, las fluctuaciones de la radiación solar, las estaciones y demás dinámicas diarias, anuales y a más largo plazo. Las moléculas chocan unas con otras y reaccionan

siguiendo los dictados de la posición, el tiempo, la temperatura, la luz, la presión y la presencia o ausencia de otras moléculas, y muchos de esos factores son, hasta cierto punto, desconocidos. A principios de la década de 1950, los científicos conocían catorce componentes atmosféricos. En la actualidad, han llegado a identificar más de tres mil.

Las raíces de la química moderna se hunden en una amplia variedad de campos: en las artes de la metalurgia y la elaboración de bebidas; en los enigmas de los antiguos filósofos sobre la naturaleza de la materia bruta y la diferencia entre sustancia y forma; en la obsesión mística de los alquimistas. Para explicar los fundamentos de la materia, estos últimos perseguían principios subyacentes, tales como los cuatro elementos aristotélicos (tierra, aire, fuego y agua), los siete metales, el espíritu universal y la piedra filosofal. Trataban de explicar la naturaleza de las sustancias a través de sus conexiones con los planetas, los personajes mitológicos y la teología, y las representaban por medio de símbolos, colores, imágenes y códigos y nombres secretos.

El mayor logro de la química en los últimos doscientos años ha sido desembarazarse de esos basamentos románticos. La que era una materia «confusa, misteriosa y caótica»¹⁴⁵ consiguió reconstruirse a sí misma sobre principios fundamentales menos esquivos. Sus rudimentos emergieron en la forma de una explosión de descubrimientos y revelaciones a lo largo del siglo XVIII, en un

¹⁴⁵ *William H. Brock, The Fontana History of Chemistry, Fontana, 1992.*

periodo que hoy se conoce como la Revolución química. El misterio y la oscuridad fueron sustituidos por la transparencia y la simplicidad de expresión. Las interacciones de la materia dejaron de estar asociadas a una nebulosa imaginería de animales, reyes y doncellas, y pasaron a ser expresadas mediante ecuaciones sencillas que reducían una historia química a lo fundamental: principio, desarrollo y final.

Actualmente clasificamos la materia en algo más de un centenar de elementos básicos, desde los muy conocidos, como el carbono y el oro, a otros más misteriosos, como el ununquadio y el rutherfordio, cuya breve existencia ha logrado algún científico caprichoso. La idea de que en toda materia subyacen ciertos elementos básicos proviene de la gran figura de la Revolución química y víctima de la Revolución francesa Antoine Lavoisier. Lavoisier era un ambicioso intelectual parisino cuyas acciones en la Ferme Générale, la compañía privada que recaudaba los impuestos en el *ancien régime*, financiaron su ciencia pero le costaron la cabeza durante el Terror. Definía, pragmáticamente, los elementos como sustancias que no pueden ser descompuestas en algo más simple. El concepto que había detrás era el de inmutabilidad y pureza, la idea de que un elemento es siempre el mismo, no importa su origen o el método de fabricación. Cada elemento, según él, debía tener un nombre, y si dos o más se combinaban para formar una sustancia más compleja, el nombre de ésta debía reflejar el de los elementos de partida. Los nombres, según los ideólogos de la Revolución química, tenían que

ser abstractos, carentes de significado en el lenguaje ordinario, de modo que «no indicasen idea alguna que pudiera sugerir falsas semejanzas».¹⁴⁶ En la práctica, algunos evocan el nombre de su descubridor, el color del elemento o, incluso, un planeta.¹⁴⁷

Aún quedaba por resolver la cuestión de cuál era la composición de los propios elementos básicos. Se solía suponer desde muy antiguo que, en última instancia, todo estaba compuesto de cierta popular sustancia, la materia primaria, que Platón y Aristóteles consideraban como una sustancia sin propiedades en la que cabía imprimir cualquier cualidad o característica. John Dalton, un profesor de Manchester que, junto a Lavoisier, es uno de los padres de la química moderna, propuso que los elementos fundamentales estaban compuestos por átomos. Todos los átomos de un elemento dado eran idénticos, pero diferían de los de otro elemento. Un átomo consiste en un núcleo cargado positivamente, rodeado de una nube de electrones cargados negativamente. Su identidad única se debe al número de protones con carga positiva que hay en el núcleo. Al químico, lo que le interesa principalmente es comprender el modo en que las interacciones entre átomos están gobernadas por el intercambio de los electrones más externos. Es como si los átomos estuvieran empeñados siempre en buscar la pareja perfecta con la

¹⁴⁶ Guyton de Morveau, en: *Historical Studies in the Language of Chemistry*, de M. P. Crosland (Dover, 1978).

¹⁴⁷ El metal mercurio es un ejemplo, aunque también tiene un nombre en latín, *hydrargyrum* (plata líquida), de donde proviene su símbolo, Hg. El cloro, un gas de aspecto verdoso, deriva su nombre de *khloros*, la palabra griega para el color verde. La Unión Internacional de Química Pura y Aplicada ha prohibido finalmente dar nombres a los elementos en honor de personas para evitar disputas desagradables, casi siempre de tipo nacionalista.

que formar un enlace estable a base de compartir e intercambiar electrones.

En cada tipo de átomo, dotado de una distribución de electrones diferente, la estabilidad se logra con un número y combinación de *socios* distinto. Algunos átomos, como el cloro (que representaremos mediante su símbolo, Cl), son demasiado reactivos para existir como átomos individuales y generalmente se encuentran formando moléculas diatómicas (representadas en este caso por Cl₂). Lo mismo sucede con los átomos de oxígeno (O), cuya configuración más estable es O₂, el oxígeno común. Pero los átomos de oxígeno también pueden existir en una forma menos estable: tres átomos interconectados, la forma conocida como ozono (O₃). Esta diferencia entre oxígeno diatómico y ozono triatómico se traduce en que el primero es un gas incoloro e inodoro, esencial para la respiración, y el segundo, un gas picante de color azul claro (un componente de la niebla humosa, notable por su toxicidad).

Cinco mil millones de toneladas de ozono flotan en la estratosfera, a 50 kilómetros de altura, protegiendo la vida que hay debajo de las formas menos benignas de la luz ultravioleta. El ozono permite el paso hacia la Tierra de las componentes más suaves (las de mayor longitud de onda), conocidas como UVA, las cuales sirven para fines útiles, como provocar la fabricación de vitamina D en la piel humana. Por el contrario, el ozono bloquea el paso de las formas más agresivas de la luz ultravioleta, UVB y UVC, que harían imposible la vida. Los UVB y UVC pueden debilitar nuestro sistema

inmune, reduciendo su capacidad para enfrentarse a las enfermedades. Pueden atacar la piel y los ojos, contándose el cáncer y las cataratas entre sus secuelas. Destruirían la forma de vida denominada fitoplancton, que se halla en el extremo inferior de la cadena alimenticia marina y cuya ausencia podría hacer que se derrumbaran ecosistemas enteros. Las plantas verdes —y, por lo tanto, los cultivos agrícolas— son vulnerables también a sus rayos. De hecho, la vida no pudo emerger del agua y poblar la superficie hasta que hubo una cantidad suficiente de ozono en la atmósfera (lo cual sucedió hace unos cuatrocientos veinte millones de años). El ozono surgió como parte del proceso por el que la atmósfera terrestre pasó gradualmente de ser rica en dióxido de carbono a ser abundante en oxígeno.

Mil millones de años después de que se empezara a formar el ozono, los humanos han evolucionado hasta un estado en el que son capaces de destruirlo. Por suerte, ha sido casi exactamente el mismo momento en que han empezado a comprenderlo. Entender el mecanismo natural de formación y destrucción de la capa de ozono y descubrir sus puntos débiles requirió un cierto número de pasos conceptuales.

El año en que da comienzo la historia es 1930, por tres motivos. En primer lugar, los científicos desvelaron el delicado mecanismo por el cual el ozono es producido y destruido de forma natural en la estratosfera. En segundo, el célebre ingeniero químico norteamericano Thomas Midgley anunció su invención de unos

útiles compuestos químicos conocidos como CFC (clorofluorocarburos). Y en tercero, el Premio Nobel de Física Robert Millikan (descubridor de los rayos cósmicos) señaló que la probabilidad de que la humanidad pudiera causar un daño significativo a algo tan colosal como la Tierra era mínima.¹⁴⁸ Llevaría cuarenta años relacionar los dos primeros factores para concluir que el tercero era erróneo.

Cuando la luz ultravioleta de energía media (UVB) alcanza la capa de ozono, se encuentra normalmente con moléculas de este gas. La luz ultravioleta puede romper los enlaces de la mayoría de las moléculas (es sólo cuestión de encontrar las frecuencias frente a las que dichos enlaces son vulnerables). El UVB puede romper el ozono, dividiéndolo en oxígeno diatómico y átomos de oxígeno libres. Cuando esto se produce —un fenómeno conocido como fotólisis—, el átomo libre resultante queda en un estado altamente excitado y busca nueva pareja. La luz ultravioleta puede romper también el robusto enlace de una molécula de oxígeno, pero en este caso se requiere la forma de energía más alta, el UVC. Éste rompe la molécula de oxígeno, convirtiéndola en dos átomos. En la primera reacción, se destruye ozono; en la segunda, oxígeno común. Una tercera reacción completa el ciclo. Los átomos libres de oxígeno generados en las dos primeras reacciones son criaturas agresivas, ansiosas por formar nuevos enlaces. En cuanto uno de esos átomos solitarios encuentra una molécula diatómica de oxígeno, se une a

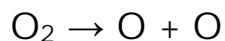
¹⁴⁸ John McNeill, *Something New Under the Sun: An Environmental History of the Twentieth Century*, Allen Lane, 2000.

ella para formar ozono de nuevo. Si, por el contrario, choca con una molécula de ozono puede robarle uno de sus átomos y dar lugar a dos moléculas de oxígeno diatómico.

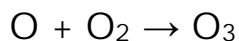
El ciclo puede ser descrito mediante algunas ecuaciones sencillas y los símbolos citados con anterioridad: O representa un átomo de oxígeno; O₂, una molécula diatómica de oxígeno común y O₃, una molécula de ozono. Utilizando una flecha para indicar un cambio químico, ésta sería la ecuación para la descomposición del ozono:



y esta otra, la de la descomposición del oxígeno ordinario:



La ecuación que completa el ciclo con la nueva generación de ozono es:



En vez de un signo igual, lo que enlaza los lados de una ecuación química es una flecha. Esto se debe a que dichos lados no son iguales en un sentido absoluto. Se trata de sustancias químicas diferentes, con distintas características (el azulado y venenoso ozono y el incoloro y vivificante oxígeno). La flecha representa el proceso

temporal durante el que se producen las interacciones químicas que dan lugar a las nuevas entidades. Pero ambos lados *son* iguales en el sentido de que se conserva el número de átomos: ninguno aparece o desaparece mágicamente. Hay tres átomos de oxígeno a cada lado de la primera ecuación (y el argumento es válido también para las otras dos).

El ciclo sigue su proceso: destruyendo y reconstruyendo ozono; cada ruptura de un enlace absorbe energía y cada creación de uno nuevo la libera en forma de calor.

La descripción de este ciclo por parte del científico inglés Sidney Chapman tiene una continuación que no se produjo hasta cuarenta años después. Las ecuaciones de Chapman no explican del todo la producción y destrucción naturales del ozono. Los cálculos basados en sus trabajos y en las tasas de las distintas reacciones químicas involucradas implicaban que el ozono debería estar presente en la estratosfera en cantidades mucho más altas que las observadas. Los científicos sabían que tenía que haber en juego otro mecanismo que favoreciera la descomposición del ozono en cuanto éste se generaba, haciendo que su concentración se mantuviera en los niveles registrados por los instrumentos. Se tardó cuatro décadas en identificar al último protagonista del ciclo natural del ozono, y cuando se encontró, resultó estar aquí abajo: en el suelo.

El descubridor de este ciclo, Paul Crutzen, ha realizado muchas aportaciones a la comprensión de la capa de ozono, la primera de las cuales data de cuando tenía veintiséis años y empezaba a

trabajar en el departamento de meteorología de la Universidad de Estocolmo. Eran los últimos años de la década de 1960 y Suecia bullía con el descubrimiento de la lluvia ácida, quizás el primer problema ambiental que se extendía a regiones enteras y que era un preludio de la destrucción del ozono. Pero Crutzen deseaba estudiar procesos naturales, así que, en cuanto tuvo la oportunidad de investigar, eligió trabajar en el ozono estratosférico.

Hacia 1970, Crutzen había encontrado que el agente que faltaba en la destrucción natural del ozono residía decenas de kilómetros más abajo. Las bacterias del suelo producen cierto tipo de óxido de nitrógeno (N_2O) en pequeñas cantidades. Crutzen observó que este óxido se difunde hacia arriba, a través de la troposfera, transformándose gradualmente en otros óxidos de nitrógeno más reactivos. Esos gases alcanzan finalmente la capa de ozono. El ozono, como hemos visto, es descompuesto fácilmente. Uno de los citados óxidos, el denominado óxido nítrico (NO), puede arrebatarse un átomo de oxígeno a una molécula de ozono y transferírsele después a un átomo de oxígeno libre, convirtiéndolo en oxígeno diatómico. El resultado neto es la transformación de ozono en oxígeno ordinario.¹⁴⁹ Crutzen había hallado el eslabón perdido en la química de la capa de ozono e introducido dos importantes conceptos que los científicos iban a emplear más tarde en esta historia: que las moléculas estables de la Tierra podían llegar a

¹⁴⁹ El ciclo puede ser representado por las ecuaciones: $\text{NO} + \text{O}_3 \rightarrow \text{NO}_2 + \text{O}_2$; $\text{NO}_2 + \text{O} \rightarrow \text{NO} + \text{O}_2$.

difundirse hasta la estratosfera y que, una vez allí, podían descomponer el ozono.

El protagonista del segundo hecho fundamental de 1930, Thomas Midgley, era un ingeniero químico norteamericano, nacido en una familia de inventores. Al final de su vida, Midgley era titular de más de un centenar de patentes y presidente de la Sociedad Química Americana.¹⁵⁰

Se había hecho famoso en 1921 al descubrir que añadiendo plomo a la gasolina se podía reducir el golpeteo del motor. Años después, tras pasar a pertenecer al departamento de investigación en refrigeración de la General Motors, anunció la invención del diclorodifluorometano, el primero de los productos químicos que serían conocidos como CFC. Por esas dos invenciones, cierto historiador medioambiental le concedió el dudoso honor de haber sido el organismo aislado de efecto más destructivo en la atmósfera en toda la historia del planeta.¹⁵¹

El invento de Midgley era un compuesto químico extraordinariamente inerte. No ardía, apenas se disolvía en agua y no era tóxico. Su arquitectura —un átomo central de carbono rodeado de átomos de flúor y cloro— era extremadamente estable. En la búsqueda atómica de la pareja ideal con la que, a base de compartir electrones, se obtenía la estabilidad, Midgley había conseguido la suprema combinación de átomos, una molécula a la

¹⁵⁰ Su muerte fue a consecuencia de un invento desgraciado: un sistema de cuerdas y poleas mediante el cual se izaba a sí mismo desde la cama por las mañanas, ya que estaba incapacitado por la polio. Cierta día, se enredó con las cuerdas y murió estrangulado.

¹⁵¹ McNeill, *op. cit.*

que no le interesaba en absoluto interaccionar con el resto del mundo. Midgley demostró esa pasividad ante una audiencia de químicos inhalando una bocanada del gas y exhalándola sobre una llama, la cual se apagó. El hecho de que saliera indemne de la demostración y no exhalara lenguas de fuego hizo que los CFC tuvieran una acogida memorable por parte del mundo científico. Aunque pasaría algún tiempo hasta que la industria los dominara, los CFC fueron considerados moléculas milagrosas, refrigerantes ideales, pues tenían un punto de ebullición entre $-40\text{ }^{\circ}\text{C}$ y $0\text{ }^{\circ}\text{C}$ (dependiendo del CFC) y eran baratos de fabricar y fáciles de almacenar. Y, sobre todo, eran seguros.

Los símbolos químicos representan un CFC de una forma más sucinta que las palabras, utilizando C para el átomo de carbono, Cl para el de cloro y F para el de flúor. Un CFC sencillo, compuesto por un átomo de carbono, tres de cloro y uno de flúor, por ejemplo, se representa como CFCl_3 . Los CFC se empezaron a usar extensivamente tras la segunda guerra mundial como propulsores en los aerosoles, como refrigerantes, en los aparatos de aire acondicionado y como líquidos limpiadores para componentes electrónicos; las emisiones crecieron desde las 20.000 toneladas anuales en la década de 1950 hasta las setecientas cincuenta mil en 1970. Su carácter inerte fue la clave tanto de su capacidad para hacer daños devastadores en la estratosfera como de la incapacidad de los científicos para reconocer su poder destructivo.

No fue hasta finales de los años sesenta cuando surgió la sospecha de que los CFC podían ser capaces de trastornar ciclos naturales a escala global. Este cambio de actitud fue motivado en parte por los desarrollos científicos, pero requirió también un cambio en los paradigmas intelectuales, en el marco de los cuales ciertas cuestiones debían ser planteadas de nuevo. Hicieron falta esos años sesenta —una era de gran inquietud social— y, en particular, el antibelicismo, los derechos civiles, el feminismo y los movimientos medioambientales. De ese activismo, de ese fermento social, emergió una nueva conciencia medioambiental, de naturaleza apocalíptica y enfoque globalizador, retratada por Rachel Carson en su libro *Silent Spring*. Era distinta del sentimiento medioambiental anterior, que había sido una reacción contra la industrialización y la añoranza romántica de una arcadia preindustrial. En Estados Unidos, el temor y el descontento de la población ante la polución de los campos con pesticidas, la muerte de los lagos y el envenenamiento de los ríos llevó al gobierno a crear la Agencia de Protección del Medio Ambiente. Dos años después, en la primera conferencia internacional sobre el medio ambiente, celebrada en Estocolmo, se escucharon las primeras quejas oficiales sobre el hecho de que la polución producida por ciertos países fuera depositada por la lluvia en otros. Norteamérica y Europa se vieron obligadas, ante el hecho innegable de la lluvia ácida, a aceptar la idea de que las sustancias químicas producidas artificialmente podían interactuar con

procesos atmosféricos naturales a una escala que excedía las fronteras nacionales.

La gente se empezó a plantear el potencial impacto ambiental, a nivel planetario, de algunas nuevas tecnologías. La primera que podía afectar a la capa de ozono eran los planes de desarrollo de aviones supersónicos. Un grupo anglofrancés había propuesto (y finalmente construyó) el *Concorde* y los soviéticos preveían también desarrollar un avión similar. En Estados Unidos, la *Boeing* tenía un plan menos avanzado para poner en el aire ochocientos aviones de este tipo entre 1985 y 1990.

Ochocientos reactores supersónicos rugiendo en la estratosfera, dejando enormes estelas de gases de escape. Sus rutas atravesarían la capa de ozono, inundándola de óxidos de nitrógeno. Apenas un año antes, Crutzen había hecho ver que los óxidos naturales del nitrógeno destruyen el ozono. Los científicos no tardaron en darse cuenta de que lo mismo podía suceder con los óxidos de nitrógeno producidos artificialmente. Cierta investigador calculó que quinientos aviones supersónicos podían reducir la cantidad de ozono en un diez por ciento en tan sólo dos años. Los científicos relacionaron el hecho con el cáncer, prediciendo que una disminución de un uno por ciento en la concentración de ozono podía causar entre cinco mil y diez mil nuevos casos al año, sólo en Estados Unidos. De esta forma, una opinión pública ya preocupada por el deterioro del medio ambiente se habituó a la idea de la capa de ozono como un delicado escudo protector alrededor del planeta

cuya integridad resultaba vital para el bienestar de todos. Pronto, la atención se centró en los planes de la agencia espacial norteamericana, NASA, para mantener una flota de lanzaderas espaciales que realizaran vuelos semanales al espacio exterior, con sus gases de escape repletos de compuestos químicos a base de cloro. Los científicos, y en particular Richard Stolarski y Ralph Cicerone, de la Universidad de Michigan, constataron que el cloro tenía los mismos efectos destructivos sobre el ozono que los óxidos de nitrógeno de Crutzen.

Desde el punto de vista químico, todas las piezas del puzzle estaban ya sobre la mesa, esperando que alguien las organizara y descubriera la destrucción del ozono por parte de los CFC. Desde el punto de vista filosófico, el mundo occidental esperaba pruebas de estar empujando a la Tierra hacia la catástrofe. Pero había un abismo entre lo que los científicos sabían por aquel entonces acerca de la destrucción del ozono y la idea de que los CFC pudieran ser los culpables. Sus indagaciones giraban en torno a la espectacular y violenta emisión de agresivos químicos en la capa de ozono por parte de futuristas aviones y naves espaciales. Los CFC eran discretos e inertes y llevaban ya en escena un par de décadas. En el teatro de la atmósfera parecían meros espectadores.

Las mentes que establecieron la conexión entre los CFC y la destrucción del ozono se habían ido forjando a través de los acontecimientos de las décadas anteriores. Sin saberlo, Mario Molina y Sherry Rowland poseían virtualmente todos los conceptos

químicos necesarios para establecer ese vínculo. Por otra parte, estaban preocupados ya por el impacto adverso de la tecnología. El activismo estudiantil había hecho sufrir en carne propia a Molina las fobias de los profanos a los demonios surgidos de los laboratorios. En 1968, cuando era investigador en la Universidad de California en Berkeley, tuvo que enfrentarse a las protestas contra las investigaciones sobre el láser en el marco de un amplio movimiento contra el posible uso de láseres de alta potencia como armas. Molina se confesaba «consternado» por el vínculo armamentista: «Deseaba participar en una investigación que fuera útil para la humanidad y no en una que tuviera aplicaciones potencialmente dañinas».

La curiosidad científica, aderezada con unos toques de conciencia medioambiental, fue lo que, según el propio Molina, le condujo a relacionar los CFC con la disminución del ozono. «Había cierto sentido medioambiental en mí, pero muy vago. Era más bien la idea de que los humanos estaban alterando su entorno sin ser del todo conscientes de las consecuencias que ello podía tener; nos sentíamos responsables, en cierto modo, de la valoración de esas consecuencias».¹⁵²

Molina y Rowland comenzaron sus trabajos sobre los CFC por una vía inesperada: una sugerencia de un científico inglés de que podían ser útiles como herramientas de investigación para los estudiosos de la atmósfera. La idea era que la capacidad de los CFC para flotar,

¹⁵² Entrevista realizada en diciembre de 2000.

eternos e imperturbables, a través de la atmósfera podía ser aprovechada por los meteorólogos para seguirles la pista a las corrientes de aire. El uso extensivo de los CFC en las décadas precedentes significaba que el ser humano los había distribuido alrededor del globo como si deliberadamente hubiese preparado un experimento científico.

Fue una de esas ideas disparatadas y brillantes que a menudo no prosperan pero que, de vez en cuando, dan en la diana. James Lovelock, un científico que trabaja por libre en su casa de campo en Devon, ha tenido varias de ellas a lo largo de su vida.¹⁵³ Otra fue el modelo Gaia, nacido en la década de 1970, según el cual la vida en la Tierra regula su propio entorno a fin de mantenerlo saludable. Gaia es la Tierra concebida como un único y gigantesco organismo, el cual, mediante una gran variedad de procesos biológicos realimentados, tiende siempre a preservarse a sí mismo. Los ejemplos de esa regulación son hoy en día abundantes; por ejemplo, se cree que las plantas y las bacterias contribuyen al control de la temperatura del planeta retirando dióxido de carbono de la atmósfera y depositándolo en el suelo. Pero la idea, con sus connotaciones de un planeta vivo y sensible, fue cuestionada por muchos científicos, en particular por su incompatibilidad con la selección natural darwiniana. Como resultado de ello, Lovelock, que es de hecho un científico muy riguroso, ha sido blanco habitual de

¹⁵³ *Homage to Gaia: The Life of an Independent Scientist*, la autobiografía de James Lovelock (Oxford University Press, 2000).

acerbas críticas.¹⁵⁴ Su propuesta de seguirle la pista a los CFC fue rechazada también por la clase científica; cierto erudito afirmaba incluso que no le veía utilidad alguna a los resultados, aun suponiendo que la misión constituyera un éxito. Así fue cómo Lovelock, pagándolo de su propio bolsillo, embarcó en la *Shackleton*, la nave de suministro de la Estación Antártica Británica. Llevó con él un instrumento muy sensible que había desarrollado, capaz de medir niveles extremadamente bajos de ciertos gases en la atmósfera. La invención tenía una década y había sido utilizada ya para detectar minúsculas concentraciones de pesticidas y otros contaminantes, lo que permitió a los científicos constatar hasta qué punto el DDT se había extendido por el planeta. El hecho fue un ladrillo más en la construcción de la conciencia medioambiental en los años sesenta.

Lovelock recorrió de norte a sur el océano Atlántico midiendo niveles de CFC y a su regreso era portador de una información enormemente valiosa. Había logrado calcular, extrapolando a partir de sus medidas, la cantidad de CFC libres en la atmósfera. Cuando estimó, basándose en los datos de la industria, la cantidad de CFC que se debían haber liberado en la atmósfera, constató que ambas magnitudes eran muy similares. ¿La conclusión? Que los CFC no se

¹⁵⁴ La selección natural de Darwin explica la actividad de un organismo en términos de lucha individual por la supervivencia; la teoría Gaia implica que los organismos puedan actuar como comunidad buscando el bien general. En *Unweaving the Rainbow*, Richard Dawkins hace una crítica de la teoría Gaia. [Trad. esp.: *Destejiendo el arco iris*, Barcelona, Tusquets Editores, col. Metatemáticas 61, 2000].

descomponen y que probablemente vagarían por la atmósfera para siempre.

Esta valiosa información fue recogida de un modo un tanto casual por Sherry Rowland, un reputado químico de la Universidad de California en Irvine. Con cuarenta y tres años, Rowland tenía en su haber una brillante carrera en química radiactiva. Había entrado en la escuela a toda velocidad, propulsado por una brillante familia, y atravesado como un meteoro el instituto y la universidad. Tras obtener el posgrado de química en la Universidad de Chicago, trabajó para el inventor de la datación por carbono radiactivo, Willard Libby. En los cinco años anteriores, Rowland había dirigido el departamento de química, cargo que abandonó en 1970 a raíz de comenzar a investigar en un nuevo tema. Por aquel entonces, se había ido interesando por el medio ambiente debido a la relevancia pública del tema y al interés que su familia tenía en él. En estas circunstancias, no es extraño que atrajera su atención el programa de una conferencia sobre la aplicación de su propia materia —la radiactividad— a asuntos medioambientales. Tras asistir a la conferencia en Salzburgo, Austria, en el tren de vuelta se encontró con un compañero que le habló de una serie de grupos de trabajo cuyo objetivo era mejorar el conocimiento de la atmósfera mediante el trabajo conjunto de meteorólogos y químicos. Rowland siempre había albergado un interés por la atmósfera desde sus tiempos de joven graduado en que trabajaba junto a Libby, por lo que acabó asistiendo a uno de los grupos, el organizado en 1972 en Fort

Lauderdale, Florida. Allí fue donde oyó hablar —a través de terceros— de los hallazgos de Lovelock y de su sugerencia de usar los CFC inertes para rastrear los movimientos atmosféricos.

Pero Rowland era químico y sabía que no hay molécula que perdure eternamente en la atmósfera, aunque sólo sea porque, cuando tarde o temprano alcanza la estratosfera, queda a merced de la radiación ultravioleta. Se preguntó cuál sería el destino final de los CFC y su curiosidad le llevó al núcleo mismo de la historia. Rowland se llevó el problema a la Universidad de California en Irvine, donde, en 1973, comenzó a trabajar en él junto a Molina, un mexicano de treinta años que acababa de concluir su doctorado. Molina era un apasionado por la ciencia que había hecho experimentos de química desde los once años en un cuarto de aseo que sus padres habían transformado para él. Abordó de inmediato el problema, cuyo sujeto se alejaba mucho de todas sus empresas científicas anteriores, y en tres meses tuvo listo el trabajo que iba a conmocionar al mundo industrializado y a cambiar nuestras vidas para siempre.

En primer lugar, Molina analizó todos los posibles destinos finales de los CFC en la troposfera: los «sumideros», tales como la oxidación o la disolución en agua de lluvia, mediante los cuales la atmósfera se deshace de la mayoría de las moléculas. No encontró nada que les impidiera alcanzar la capa de ozono. A continuación, calculó el tiempo que tardaría una molécula de CFC, recién salida de un bote de aerosol, en ascender lo suficiente para ser destruida: medio siglo, aproximadamente. Según eso, un soplo de aerosol que hubiera

servido para perfumar a Brigitte Bardot en 1970 estaría circulando treinta años después y aún le faltarían dos décadas para desintegrarse.

Molina y Rowland se dieron cuenta de que, una vez los CFC hubieran alcanzado la suficiente altura, sufrirían el inmediato ataque de los rayos ultravioleta de alta energía, capaces de romper uno de sus enlaces y de liberar un agresivo átomo de cloro. Podían haberse detenido aquí, en la descomposición de los CFC, pero decidieron seguir y descubrir el destino final de sus fragmentos. En primer lugar, tuvieron que considerar las interacciones moleculares y, seguidamente, integrar esta escena química en el teatro de la dinámica atmosférica. A medida que trabajaban con las diversas sustancias que podían interaccionar químicamente, hacían uso de investigaciones anteriores sobre cinética química, el estudio de la velocidad con la que interaccionan las moléculas y del modo en que las reacciones tienen lugar. El concienzudo trabajo de los químicos había demostrado ya que un experimento de laboratorio puede revelar cuán rápidamente se desarrolla una reacción, aunque esa reacción implique átomos de cloro que pululen en un lugar inaccesible, tal como las frías y enrarecidas regiones de la estratosfera. Los científicos habían realizado ya muchos de esos experimentos y registrado los resultados, con lo que algo que podría haberles llevado décadas de trabajo a Molina y a Rowland fue conseguido en pocos días.

Ambos introdujeron esos datos en modelos de los procesos dinámicos que controlan el movimiento del ozono. Molina, que actualmente trabaja en el Instituto Tecnológico de Massachusetts, me decía:

«La cuestión es cómo se comporta la química en el sistema real, en vez de en el laboratorio. Es fácil perderse en los detalles. Hay que sintetizar las características esenciales de los procesos del sistema para poder confiar en los resultados. Es algo que ahora sabemos muy bien: cómo funciona la atmósfera. Pero, como científicos atmosféricos, entonces estábamos en pañales».

Molina recordaba su soledad de esa época; trabajando por su cuenta en lo que a menudo eran prosaicas labores, informando regularmente a Rowland y avanzando sin descanso hacia una conclusión que cada vez parecía más emocionante. Mucho de ese trabajo lo hacía con papel y lápiz, efectuando cálculos o dibujando rudimentarias representaciones de moléculas, como hacen los químicos cuando exploran las posibles alternativas de una reacción. También llevaba a cabo sencillos experimentos para analizar con detalle lo que le sucedía a los distintos CFC cuando se hallaban en presencia de la luz ultravioleta.

En aquel tiempo, los modelos de la atmósfera se ocupaban sólo del movimiento de los gases hacia arriba y hacia abajo, representándolos como si fuesen columnas. Los modelos no les decían nada a los investigadores acerca de las variaciones debidas a los cambios de estación o de latitud. Aun así, manejados con

habilidad, esos toscos modelos arrojaban una chocante respuesta: los reactivos átomos de cloro harían lo mismo que los óxidos naturales de nitrógeno que había estudiado Crutzen: devorar ozono. Comenzarían por extraer del ozono un átomo de oxígeno, convirtiendo aquél en oxígeno ordinario, y luego transferirían ese átomo de oxígeno a otro átomo de oxígeno libre, creando una nueva molécula de oxígeno diatómico. El cloro, pues, estaría absorbiendo O_3 y O y generando O_2 en el proceso. Y, lo que era crucial: cuando el átomo de cloro terminara de realizar el ciclo, estaría listo para comenzar otra vez; de hecho, así es en la realidad: un solo átomo de cloro puede destruir miles de moléculas de ozono. Sólo se detendría cuando encontrara una sustancia distinta, una a la que se enlazara más rápidamente, dando lugar a una denominada molécula terminal. Una vez alcanzada esa forma estable, las fechorías del cloro habrían finalizado; la nueva molécula descendería en el seno de la atmósfera hasta ser arrastrada por la lluvia.

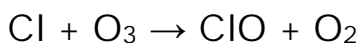
En promedio y según el modelo, un único átomo de cloro podía desintegrar cien mil moléculas de ozono antes de hallar su propio destino. Al final, la producción y destrucción del ozono alcanzaría un nuevo equilibrio, en el cual la presencia de CFC habría causado aproximadamente una disminución de un diez por ciento en los niveles de ozono estratosférico. La consecuencia era que los rayos UVB penetrarían con mayor facilidad y alcanzarían a los seres vivos de abajo, desencadenando decenas de miles de nuevos casos de

cáncer de piel al año y haciendo que los animales fuesen más vulnerables a las enfermedades por el debilitamiento de sus sistemas inmunes.

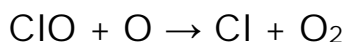
Mediante el lenguaje de la química, Molina y Rowland plasmaron la esencia de su perturbador mensaje. El CFC se descompone, liberando cloro:



El cloro ataca el ozono y produce una molécula de oxígeno:



A continuación, da origen a otra más:



Las ecuaciones son más fáciles de seguir si nos centramos en la historia de uno solo de los actores: el cloro que aparece en la primera de ellas. Durante cincuenta años ha existido en la forma representada a la izquierda de la primera ecuación —formando parte de un CFC estable—. Es abruptamente arrojado fuera de este confortable estado por efecto de la luz ultravioleta, emergiendo como un átomo agresivo y solitario a la derecha de esa misma ecuación. En la segunda, se topa con una molécula de ozono. El cloro le

arrebató al ozono uno de sus átomos y tiene con él una breve relación. El desenlace de la historia se produce cuando nuestro átomo de cloro pierde a su compañero y vuelve a aparecer de nuevo como un cloro libre y agresivo al final de la tercera ecuación.

Ambos químicos fueron conscientes de la sensacional naturaleza de su descubrimiento y de la importancia de comunicarlo lo antes posible al resto del mundo. Publicaron sus ecuaciones, junto con una descripción del trabajo realizado, en la revista científica *Nature* del 28 de junio de 1974, condensándolo todo en menos de tres páginas y aguardando ver la reacción del público. Nada sucedió.

Su trabajo pasó inadvertido para el público y para los periodistas científicos que, teóricamente, se lo debían interpretar. Molina y Rowland concluyeron que, quizás, el artículo resultaba inaccesible y sus advertencias quedaban enmascaradas por el lenguaje de las publicaciones científicas. Su artículo, «El fin estratosférico de los fluorometanos: la destrucción del ozono catalizada por el átomo de cloro», hacía una breve referencia en el penúltimo párrafo a la posibilidad de que, de las reacciones descritas por las ecuaciones, «se derivaran consecuencias importantes» que dieran lugar a «problemas medioambientales». El editorial que lo acompañaba era también muy cauto.

Así las cosas, los dos científicos dieron una conferencia de prensa durante un congreso de la Sociedad Química Americana celebrado en Atlantic City en septiembre de 1974, donde presentaron dos artículos más sobre el destino de los CFC. Molina recuerda:

«Pensamos que la única alternativa para la sociedad era involucrar a la gente. No nos sentíamos demasiado cómodos en actividades como las de tratar con la prensa y la industria, pero creímos que era importante dar aquel paso extra. No fue difícil tomar la decisión de hacer una conferencia de prensa; nos proponíamos atraer la atención del gobierno y del público hacia nuestros descubrimientos, pero éramos conscientes de que, en la comunidad científica, no todos aplaudirían nuestra decisión. Suele causar cierto resentimiento que los científicos difundan sus ideas en los medios de comunicación».

Durante la conferencia de prensa abogaron por la prohibición total de las emisiones de CFC a la atmósfera. A partir del momento en que habían decidido saltar desde el terreno científico al de la opinión pública, no había marcha atrás. A finales de 1974, sus trabajos se convirtieron en gran noticia en Estados Unidos; aparecieron en la primera página de los periódicos y constituyeron el tema de algunos programas de televisión, viéndose también refrendados por artículos publicados por otros especialistas del campo. Los dos años siguientes fueron un auténtico «torbellino», como reconocían después. Obviamente, sus comentarios suponían toda una amenaza para la industria química estadounidense, que en aquel tiempo encabezaba la producción mundial de CFC. Los fabricantes arremetieron contra los científicos, señalando que sus

pretensiones eran pura teoría y que no se podía socavar toda una industria, provocando un desastre económico de imprevisibles dimensiones, a partir de unos argumentos tan débiles. Contra esto, los científicos argumentaron que la amenaza era tan seria que no era posible esperar a que emergiera la «prueba».

«Las ideas científicas no eran fáciles de explicar al principio, así que no era nada sencillo cuando teníamos confrontaciones con la industria o con otros expertos. Ése era nuestro reto», relata Molina, que piensa que su actuación contribuyó a hacer cambiar la actitud de los científicos sobre su responsabilidad a la hora de comunicar al público sus hallazgos.

Durante dos años, «El increíble circo estratosférico y su sociedad de debate» recorrió Estados Unidos.¹⁵⁵ Molina y Rowland comparecieron en numerosas audiencias legislativas, federales y estatales junto a Ralph Cicerone, de la Universidad de Michigan, que había realizado una investigación similar. El gigante químico DuPont, mientras tanto, encabezaba la oposición científica. Uno de sus testigos era nada menos que Lovelock, quien, en 1974, compareció como tal ante el Congreso de Estados Unidos. Aunque la historia le hará justicia por sus contribuciones al esclarecimiento de la historia de los CFC, Lovelock fue en un primer momento escéptico respecto a las afirmaciones de Molina y Rowland. Años después, se vio obligado a decir en su descargo en la revista *New Scientist*: «Algunos puede que piensen que era un estúpido, pero

¹⁵⁵ Lydia Dotto y Harold Schiff, *The Ozone War*, Doubleday, 1978.

creo que me limité a hacer lo que me pareció más lógico. Me gustaba aquella gente [de la industria], me parecía un grupo de científicos simpáticos y honrados». ¹⁵⁶ De hecho, escéptico siempre sobre la capacidad del hombre para destruir la Tierra —la formidable fuerza de Gaia—, nunca estuvo convencido del todo de que una exposición algo mayor a la radiación ultravioleta fuera motivo de preocupación.

Mientras tanto, la industria química se sacó de la manga a otro científico, Richard Scorer —que había calificado la investigación de «ampulosa idiotez»—, y difundió su mensaje por toda Norteamérica. Quienes hacían campaña contra los CFC y eran pelirrojos y de tez clara solían ser acusados de partidismo, con lo que, a menudo, el debate se volvía «personal y violento», en palabras de Molina y Rowland, que, no obstante, nunca se arrepintieron de sus decisiones.

En esta etapa el debate se centraba casi exclusivamente en el uso de los CFC en los aerosoles y se hacía escasa mención a su empleo como refrigerantes o como agentes de expansión en la fabricación de plásticos como el poliestireno. Dos años después, en 1976, la Academia Nacional de Ciencias hizo una revisión general de la técnica, apoyada por nuevas investigaciones. En la citada revisión recomendaba la restricción drástica de las aplicaciones no esenciales de los CFC, salvo que en un plazo de dos años se hallase algún modo de mitigar el problema. El informe inclinó

¹⁵⁶ *New Scientist*, 9 de septiembre de 2000.

definitivamente la balanza y, en 1978, varias entidades, entre ellas la Agencia de Protección del Medio Ambiente, habían emitido regulaciones que tendían a la eliminación de los CFC. Canadá, Noruega y Suecia tomaron también medidas. A comienzos de los ochenta, crecía la respuesta internacional al problema. El Programa Medioambiental de las Naciones Unidas (puesto en marcha en 1972, tras la celebración en Estocolmo de la primera conferencia internacional sobre el medio ambiente) promovía discusiones sobre los CFC. En 1985 logró que veinte naciones firmaran el Convenio de Viena para la Protección de la Capa de Ozono.

Lo más sorprendente de esta historia es que se tomaran tantas medidas a pesar de la ausencia total de evidencias empíricas que demostraran que los CFC estaban destruyendo el ozono. Los científicos no habían conseguido registrar dato alguno que probara que el ozono estaba desapareciendo. De hecho, y tal como ellos mismos reconocían, no se había detectado en ningún lugar de la estratosfera la presencia de sustancias que contuvieran cloro.¹⁵⁷ El Convenio de Viena supuso la primera vez que las naciones acordaban atajar un problema medioambiental global antes de que nadie hubiera sentido o demostrado empíricamente sus efectos.

La explicación de este hecho se halla en parte en la conciencia ambiental de Estados Unidos de los años setenta, en la que algunas organizaciones recién creadas, como la Agencia de Protección del

¹⁵⁷ A finales de la década de 1970 existían medidas que demostraban que los CFC alcanzaban la estratosfera y posteriores registros mostraron la presencia en ella de radicales de cloro libre. Los cambios en los niveles de ozono sólo fueron registrados a mediados de los ochenta.

Medio Ambiente, deseaban mostrar sus poderes. En cuanto Ronald Reagan sustituyó a Jimmy Cárter como presidente en 1981, el panorama cambió y muchos de los planes para reducir la producción de CFC fueron archivados.¹⁵⁸ Si Molina y Rowland hubiesen presentado su trabajo sólo unos años más tarde, la respuesta inicial podría haber sido muy diferente. Colaboraba también la naturaleza siniestra de la amenaza, que iba asociada a conceptos terroríficos para el gran público, tales como rayos mortales invisibles y cáncer de piel. Finalmente, la solución no era, a grandes rasgos, costosa. No implicaba grandes cambios en la forma de vida individual y las industrias afectadas —los fabricantes de objetos de lujo, como los *sprays* desodorantes— no despertaban excesivas simpatías.

El desarrollo de la historia supuso un primer ejemplo de aplicación del principio de precaución: la idea de que no siempre se puede esperar a la *prueba* de un potencial perjuicio cuando están en juego la salud humana y la degradación del medio ambiente. Era obvio que los fabricantes pensaban de otra manera: los CFC eran inocentes mientras no se demostrara lo contrario.

«Pedían [para ellos] los mismos derechos que [tenían] los individuos», señala Molina. En realidad, la teoría de la destrucción del ozono por parte de los CFC era una base ideal para la aplicación del principio de precaución. Podía haber consecuencias desastrosas para la salud humana si se hacían oídos sordos a las advertencias

¹⁵⁸ Donald Hodel, secretario del Interior en la época de Reagan, argumentaba que era preferible llevar sombrero y gafas de sol a tratar de impedir la destrucción del ozono.

de los científicos —es decir, el coste de posponer la acción podía ser muy elevado— y la solución propuesta no era, en términos relativos, costosa o perjudicial. Un caso muy similar fue el ocurrido en torno a otro de los inventos de Thomas Midgley: la gasolina con plomo. En el Reino Unido, la decisión de eliminarla se tomó aunque no había pruebas de que pudiera dañar a los niños. El alto riesgo se unía al hecho de que existían alternativas tecnológicas y los automóviles podían ser adaptados sin grandes costes.

Las ecuaciones químicas representan sólo fragmentos de la realidad. El mérito de Molina y Rowland fue ser capaces de aislar un fragmento significativo, de distinguirlo entre la multitud. Sus ecuaciones no pretenden ser una relación completa de las actividades que tienen lugar en la capa de ozono. Por el contrario, tratan de describir las más significativas, contempladas desde la perspectiva del ser humano. El carácter incompleto de las ecuaciones reflejaba también las limitaciones del conocimiento químico de la época, algo que quedó patente once años después, cuando tres científicos británicos anunciaron en 1985 su descubrimiento del agujero de ozono. Fue el momento en que el ozono atrajo la atención internacional y los europeos abanderaron la causa contra los CFC; la segunda fase de la campaña había comenzado. El descubrimiento, en cualquier caso, conmocionó a los científicos tanto como al resto. Nadie se esperaba un gigantesco agujero en la capa de ozono sobre la Antártida, donde apareció: lo

que estaban buscando eran erosiones más ligeras en latitudes medias.

El problema fundamental para la química —un problema que ilustran vigorosamente los esfuerzos de Molina y Rowland— es que trata de describir el mundo real y material, un lugar infinitamente complicado. La química —y las ecuaciones que la representan— nunca puede ser correcta en la plena acepción de la palabra. Las ecuaciones no pueden enumerar todas las condiciones necesarias para que la reacción tenga lugar o para impedir que se produzca. Su pureza, pues, se logra a expensas de su completitud. Una ecuación afronta el riesgo de convertirse en falsa o irrelevante cuando cambian las circunstancias externas. Esto es lo que sucedió con el descubrimiento del agujero de ozono.

Los tres científicos que se toparon con el agujero sobre el polo sur se hallaban en aquellos días en sus despachos de Cambridge, trabajando para la Estación Antártica Británica.¹⁵⁹ Una de sus tareas de rutina era procesar los datos que llegaban de Halley Bay, donde se venían monitorizando los niveles de ozono en la estratosfera desde la década de 1950. Desde comienzos de los ochenta, el instrumento de medida —un dispositivo conocido como espectrofotómetro Dobson— había comenzado a registrar sistemáticamente bajas lecturas de ese gas. Brian Gardiner, Joe Farman y Jonathan Shanklin sospechaban que se trataba de un error del aparato. Los tres tenían un amplia experiencia en la

¹⁵⁹ La página web de la Estación Antártica Británica (<http://www.antarctica.ac.uk>) describe las investigaciones actuales y proporciona información sobre la vida en la Antártida.

Antártida. Habían trabajado en medio de fantásticos paisajes de ventiscas, blancos precipicios y llanuras sin fin con los pingüinos y el crujir del hielo como única compañía. Conocían las gélidas noches en tiendas de campaña y las caminatas de horas sobre el hielo para recoger las lecturas de algún lejano instrumento. Sabían que el frío extremo y las dificultades técnicas para manejar instrumentos remotos podían distorsionar las lecturas, por lo que se requería cierta cautela a la hora de interpretar los datos.

«La Antártida era, lógicamente, el último lugar de la Tierra en el que cabía esperar ver la disminución del ozono», comenta Gardiner.¹⁶⁰ Pero cuanto más riguroso era el trabajo y mayor la precisión de las medidas, más grandes se hacían los números que expresaban la magnitud de la reducción. Todos los años sucedía lo mismo: en octubre, a principios de la primavera antártica, se producía de pronto una disminución muy significativa que duraba unos dos meses, tras la cual los niveles se recuperaban gradualmente. El espectrofotómetro Dobson parecía, de hecho, estar contemplando un gigantesco vacío de ozono que emergía de las inmensidades antárticas. Para mayor desconcierto de los científicos británicos, el instrumento rival tecnológicamente más avanzado, ubicado a bordo del satélite de la NASA *Nimbus*, que orbitaba sobre la Antártida, no detectaba nada anómalo en relación con la capa de ozono.

Al cabo de tres años, Farman, Gardiner y Shanklin estaban seguros de haber comprobado sus medidas con el suficiente rigor. No podían

¹⁶⁰ Entrevista realizada en diciembre de 2000.

negar lo que los datos revelaban: que, cada mes de octubre, un tercio del ozono antártico desaparecía. Gardiner, que en la actualidad dirige la unidad de meteorología y ozono de la Estación Antártica Británica, habla de la «terrible conclusión» de que el efecto era real, de que habían detectado un problema de trascendencia planetaria. Conjuguar la precaución científica con la responsabilidad de disparar las alarmas lo antes posible constituyó una experiencia angustiosa.

En 1985 publicaron un artículo que describía un vacío enorme y repentino sobre Halley Bay, un agujero mucho más profundo que el Everest —se extendía hacia arriba entre 10 y 24 kilómetros—. Cada año, el agujero de ozono parecía rellenarse a sí mismo poco a poco hasta evaporarse de nuevo, como una cenicienta atmosférica, con la primera señal de que la larga noche antártica había terminado. El satélite *Nimbus*, que contemplaba la Tierra desde arriba con toda una variedad de ojos mecánicos, fue corregido de inmediato y confirmó los mismos datos. Resultó que los científicos de la NASA habían transmitido a los ordenadores su despreocupación por los niveles de ozono en la forma de una orden a los instrumentos del *Nimbus* para que ignoraran las lecturas de bajos niveles de ozono que parecieran erróneas. Una vez reajustado, el satélite comenzó a generar impactantes imágenes del globo en las que una oscura sombra sobre la Antártida crecía ominosamente todos los años. El vivido retrato de esa gigantesca mancha causada por el hombre conmocionó tanto al público como a la comunidad científica.

«Recuerdo muy bien la llegada de los resultados», relata el profesor Alan O'Neill, un importante especialista atmosférico de la Universidad de Reading. «Nadie tenía ni idea de lo que estaba ocurriendo».

Eso era lo que Molina y Rowland no habían sido capaces de predecir. La química es mucho más compleja que lo que sus ecuaciones podrían nunca describir; aparte de los factores que habían considerado, había muchos otros en juego.

Los posteriores esfuerzos por reconciliar las ideas de Molina y Rowland con los hallazgos de la Estación Antártica Británica demostraron ser una fuerza unificadora en la ciencia atmosférica, al revelar que la atmósfera era un sujeto de estudio más amplio que cualquier otra disciplina científica. Entre los meteorólogos, físicos, matemáticos y químicos, habituados a no rebasar los límites de sus respectivas materias, surgió una nueva mentalidad, una forma de trabajo capaz de saltar de una disciplina a otra, de aceptar datos imperfectos, resultados parciales, modelos con interrogantes y áreas en las que nadie sabe con certeza qué sucede.

En los dieciocho meses siguientes hubo que trabajar bajo presión, debido a la urgente necesidad de precisar los límites del agujero y a la conveniencia política de que los científicos, que representaban *la verdad* para el resto del mundo, alcanzasen un consenso al respecto. Las ecuaciones de Molina-Rowland eran, obviamente, un buen punto de partida, por lo que ambos científicos, junto a Crutzen y a otros importantes investigadores, se pusieron manos a

la obra para encontrar una explicación. Varias disciplinas se enzarzaron en batallas intelectuales por explicar las causas del agujero.¹⁶¹ Según Gardiner, todos pretendían llevar el problema a su terreno. El *lobby* troposférico argumentaba que era un efecto originado en la parte baja de la atmósfera, que se difundía hacia arriba. Los científicos de la mesosfera afirmaban que procedía de arriba. Los expertos en la estratosfera hablaban de efectos laterales; la reducción tenía lugar en alguna parte del globo y se desplazaba hacia el sur, hasta la Antártida. Los físicos solares buscaban una respuesta en las fluctuaciones de la actividad solar. Mientras tanto, los físicos bosquejaban diagramas de flujo para explicar que no se trataba de una reducción, sino de una redistribución del ozono alrededor del planeta, debida a los cambios de temperatura que tenían lugar al comienzo de la primavera antártica. Los químicos, por su parte, jugaban con las circunstancias inusualmente frías del invierno antártico para ver si creaban unas condiciones de partida que provocasen la explosión de un proceso químico de destrucción del ozono al iniciarse la primavera. Algunas de las teorías parecían claramente absurdas desde un principio, pero se consideró preferible analizarlas y descartarlas ordenadamente para evitar que pudieran resurgir más tarde con cierta credibilidad. Los científicos alcanzaron el consenso con sorprendente celeridad y muchos

¹⁶¹ Para más detalles sobre las teorías en liza y sobre la historia del descubrimiento de la disminución del ozono estratosférico hasta 1987, véase: John Gribbin, *The Hole in the Sky: Man's Threat to the Ozone Layer*, Corgi, 1988.

comentaristas consideran hoy el proceso como un modelo de respuesta científica a un problema medioambiental.

Vencieron los químicos. Varios equipos sugirieron que la respuesta se hallaba en la química del implacable invierno antártico. De abril a agosto el Sol evita la Antártida y fuertes vientos azotan el polo sur, aislando su aire del resto de la atmósfera. En ese mundo incomunicado y glacial, se crean nubes altas en forma de corona. Los exploradores han hablado muchas veces de esas nubes espectrales, que retienen el resplandor del Sol mucho después de que éste haya dejado de alumbrar la tierra. Las nubes están llenas de cristales de hielo, sólidos finamente dispersados en los que se halla la causa de la rápida destrucción del ozono. Nadie conoce con exactitud el mecanismo, pero se sabe que en la superficie de esas partículas de hielo sólido pueden producirse reacciones que de otro modo no tendrían lugar. Los cristales actúan como un «punto de encuentro» estratosférico en el que los gases, que atravesarían la noche antártica sin detenerse de no existir aquéllos, pueden hacer una pausa, reunirse unos con otros e interaccionar.

Los científicos pudieron demostrar que las moléculas que se detienen allí e interaccionan son las famosas «moléculas terminales» (fundamentalmente, ácido clorhídrico y nitrato de cloro) que Molina y Rowland ya habían descrito. Como vimos, el desbocado átomo de cloro, tras implicarse en miles de reacciones con el ozono, es capturado finalmente para formar una molécula más estable, momento en el que sus travesuras cesan y comienza a descender

hacia el suelo. Pero en presencia de los cristales de hielo, esas moléculas terminales se demoran y reaccionan en su superficie para formar nuevas sustancias, a veces la propia molécula de cloro (Cl_2). En cuanto la luz del Sol baña de nuevo la Antártida, sus rayos ultravioleta se dedican a romper los enlaces cloro-cloro, dejando átomos de cloro libres, los cuales vuelven a sus actividades destructivas. El invierno antártico, pues, crea unas condiciones únicas para un posterior episodio de destrucción de ozono.

Cuando los científicos anunciaron su descubrimiento del agujero de ozono en la Antártida —un fenómeno que las imágenes del *Nimbus* hacían claro y comprensible—, la noticia sacudió las conciencias de todo el mundo civilizado. Por primera vez se estaba ante la inminencia de un desastre medioambiental que la gente común podía contribuir a evitar. Todo lo que tenían que hacer era alterar ligeramente sus hábitos de compra. Ese vínculo entre lo apocalíptico y lo cotidiano captó la imaginación de la gente, difundió el concepto de consumidor preocupado por el medio ambiente e hizo que muchos leyeran la letra pequeña de los envases de aerosol. Incluso el príncipe Carlos de Inglaterra anunció que la princesa Diana se limitaría a utilizar vaporizadores ecológicos.

El descubrimiento del agujero estimuló las negociaciones internacionales. En la actualidad son más de noventa los signatarios de la última enmienda del Protocolo de Montreal, que incluye el compromiso de eliminar progresivamente toda una gama de otras sustancias sospechosas de destruir el ozono. Aunque existe

el temor de que la producción china de CFC podría arruinar los avances logrados, muchos se atreven hoy a hablar como si el problema hubiera quedado resuelto. La mayoría de los científicos predicen que el agujero quedará cerrado entre 2050 y 2075.¹⁶² El agujero de ozono antártico sigue creciendo todavía, pero el hecho se atribuye al tiempo que los CFC tardan en alcanzar la estratosfera. Tal vez a lo largo del presente siglo podamos contemplar una capa de ozono intacta y la solución a su destrucción se convierta en un ejemplo de libro de respuesta medioambiental. Molina cree que la mayor virtud de su obra ha sido crear la conciencia de que el mundo es un único sistema, lo que a su vez ha reforzado la idea de que las actividades humanas pueden afectar a la atmósfera de todo el planeta.

«En cierto sentido, constituyó un primer ejemplo: la ciencia tenía perfectamente claro que una cosa así podía ocurrir. Pero también es un precedente importante de que ese tipo de cosas pueden tener solución». Gardiner es más específico; argumenta que deberíamos aprender de las negociaciones en torno al ozono de cara a abordar temas como el calentamiento global:¹⁶³ deberíamos «atacar el problema más grave del cambio climático implementando acuerdos

¹⁶² Un estudio presentado en la Segunda Asamblea General sobre los Procesos Estratosféricos y su Papel en el Clima (SPARC), celebrada en diciembre de 2000 en Buenos Aires, era optimista al respecto.

¹⁶³ El calentamiento global se manifiesta como un incremento en la temperatura media de la Tierra que podría ser debido a la excesiva acumulación en la atmósfera de gases procedentes de la utilización de combustibles fósiles, así como a la tala de bosques y la expansión de la agricultura.

internacionales que limiten el empleo de combustibles fósiles antes de que las consecuencias sean irreparables». ¹⁶⁴

El ozono, pues, nos ha hechizado, haciéndonos creer que, si hemos podido impedir la primera catástrofe medioambiental global, podremos evitar futuros problemas del mismo modo. Igual que las naciones lograron restringir su producción de CFC, ¿podrían limitar el empleo de combustibles fósiles para mitigar el calentamiento global?

El siguiente capítulo en la historia de los problemas medioambientales es demasiado complejo para admitir una solución tan simple como la del ozono. El ozono era limpio y sencillo; se trataba de algo urgente y abrumador. Los problemas actuales más importantes —el calentamiento global, la deforestación, la disminución de la biodiversidad— son confusos, clandestinos y de manifestación lenta. En el caso del calentamiento global, los combustibles fósiles forman parte de los cimientos de culturas y estilos de vida. Como el propio Rowland le decía en 1997 a Bill Clinton, siendo éste presidente de Estados Unidos, «En el caso de los CFC, se trataba de gases fabricados por no más de una veintena de empresas en el mundo, todas ellas de origen científico, y su empleo se concentraba en aplicaciones típicas de la sociedad opulenta. La energía de los combustibles fósiles, en cambio, es usada por todos casi todos los días y en todas las actividades». ¹⁶⁵ La historia del ozono supuso un problema extraordinario con una

¹⁶⁴ Entrevista para la BBC, diciembre de 2000.

¹⁶⁵ Mesa Redonda sobre el Cambio Climático Global, 24 de julio de 1997.

solución extraordinaria.¹⁶⁶ Sobre todo, fue la sencillez de las ecuaciones del ozono lo que hizo que el mundo creyera erróneamente que era fácil encontrar soluciones que impidieran otros desastres medioambientales.¹⁶⁷

Dos visiones de nuestro planeta han dominado esta historia. La Tierra como un vulnerable paraíso, perfecto en su belleza delicada y azul. Y la Tierra violada, mancillada en el polo sur por nuestros efluvios industriales. La química relaciona ambas visiones y las ecuaciones químicas no hacen sino expresar la esencia de ese vínculo. Nuestra historia comenzaba con una Tierra indefensa, tan desvalida como la princesa en la torre. Llegaban entonces los guerreros medioambientales, dispuestos a defender su honor. Habían sido convocados por sabios científicos que se arriesgaban a ser proscritos por anunciar sus vaticinios. En el último capítulo, se unían todas las naciones y evitaban la tragedia.

Las ecuaciones de Molina-Rowland yacen en el corazón de ese cuento de hadas.

¹⁶⁶ McNeill, *op. cit.*

¹⁶⁷ En *Earth Rising: American Environmentalism in the 21st Century* (Island Press, 2000), Philip Shabecoff argumenta que el movimiento medioambiental del siglo XX alcanzó sus límites por dos razones: problemas más complejos con soluciones menos claras y contra-campañas más sofisticadas y agresivas.

Epílogo

Cómo sobreviven las grandes ecuaciones

Steven Weinberg

Es tremendamente difícil ponerse en el lugar de quienes vivieron en los siglos pasados, pero lo cierto es que muchos de sus artefactos — edificios, carreteras, obras de arte— han sobrevivido y algunos de ellos continúan en uso. De la misma forma, aunque suele ser difícil comprender el pensamiento de los científicos del pasado, quienes no conocían mucho de lo que conocemos hoy, las grandes ecuaciones que llevan sus nombres continúan con nosotros y siguen siendo útiles (las ecuaciones de Maxwell para el campo electromagnético, las de Einstein para el campo gravitatorio, la ecuación de Schrödinger para la función de onda de la mecánica cuántica y otras muchas examinadas en este libro). Esas ecuaciones son monumentos del progreso científico, al igual que las catedrales lo son del espíritu de la Edad Media. ¿Llegará el día en que no enseñemos esas grandes ecuaciones a nuestros estudiantes?

Aunque dichas ecuaciones formarán siempre parte del conocimiento científico, ha habido profundos cambios en nuestro conocimiento de los contextos en el marco de los cuales son válidas y de las razones por las que son válidas en esos contextos. Ya no pensamos en las ecuaciones de Maxwell como en una descripción de tensiones en el seno de un éter, como lo hacía Maxwell, o incluso una descripción

exacta de los campos electromagnéticos, como lo hacía su colega físico Oliver Heaviside. Sabemos desde la década de 1930 que las ecuaciones que gobiernan los campos electromagnéticos contienen un número infinito de términos adicionales, proporcionales a potencias cada vez más altas de esos campos y a la frecuencia con la que dichos campos oscilan. Esos términos adicionales son diminutos para las frecuencias de la luz visible, pero a frecuencias mucho más altas pueden conducir a que la luz sea dispersada por la propia luz. La teoría de Maxwell es una *teoría de campos eficaz*, una teoría que constituye una buena aproximación sólo para campos que sean lo suficientemente débiles y que varíen de forma suficientemente lenta.

Los términos adicionales que hay que añadir a las ecuaciones de Maxwell provienen de la interacción de los campos electromagnéticos con parejas de partículas y antipartículas cargadas que continuamente emergen y se aniquilan en el espacio vacío. En la década de 1930, los cálculos relativos a esos términos fueron desarrollados mediante la electrodinámica cuántica, la teoría cuántica del electromagnetismo que abarca electrones y antielectrones. Pero la electrodinámica cuántica tampoco es en sí misma la respuesta última. Proviene de las ecuaciones de una teoría más fundamental, el modelo estándar de las partículas elementales, en una aproximación en la que se asume que todas las energías son demasiado pequeñas para dar lugar a los cuantos de los campos W y Z , los *hermanos* del campo electromagnético en dicho modelo

estándar. Y el modelo estándar tampoco es la última palabra; pensamos que es sólo una aproximación de baja energía a una teoría aún más fundamental, cuyas ecuaciones no tendrían por qué tratar de los campos electromagnéticos o de los campos W y Z .

Las ecuaciones de la relatividad general han sufrido una reinterpretación similar. Al deducirlas, Einstein partió de una idea básica, el principio de equivalencia entre gravitación e inercia, pero también introdujo una premisa *ad hoc* de simplicidad matemática, la de que las ecuaciones debían ser del tipo conocido como ecuaciones diferenciales parciales de segundo orden. Esto significa que Einstein supuso que las ecuaciones implicaban sólo tasas de variación en los campos (primeras derivadas) y tasas de variación en esas tasas de variación (segundas derivadas), pero no tasas de mayor orden. No conozco sitio alguno en el que el gran físico explicara el motivo por el que adoptaba una premisa así. En su artículo de 1916 sobre la relatividad general admitía la existencia de «un mínimo de arbitrariedad a la hora de elegir esas ecuaciones», ya que se trataba, esencialmente, de las únicas ecuaciones diferenciales parciales de segundo orden para los campos gravitatorios que eran consistentes con el principio de equivalencia entre gravitación e inercia; pero, al menos en ese artículo, no hacía intención alguna de explicar por qué las ecuaciones tenían que ser de segundo orden. Tal vez se basaba en el hecho, tampoco explicado, de que cuando la teoría de la gravitación de Newton es descrita en términos de un campo gravitatorio, la ecuación que

gobierna ese campo (la ecuación de Poisson) es de segundo orden. O quizás intuía que unas ecuaciones tan trascendentales tenían que ser lo más simples posible.

En la actualidad, la relatividad general es ampliamente, aunque no universalmente, considerada como otra teoría de campos eficaz, aplicable sólo a distancias mucho mayores que 10^{-33} centímetros y para energías de las partículas muy inferiores a la equivalente a la masa en reposo de 10^{19} protones. Hoy nadie se tomaría en serio cualquier consecuencia de la relatividad general a distancias más cortas o para energías más altas.

Cuanto más importante sea una ecuación, más atentos debemos estar a los cambios en su significado. El caso más drástico a este respecto lo constituye la ecuación de Dirac. No sólo se ha alterado nuestra forma de ver por qué una ecuación es válida y bajo qué condiciones lo es, sino que se ha operado un cambio radical en nuestra comprensión del propio objeto de aquélla.

En 1928, Paul Dirac trataba de encontrar una versión de la ecuación de Schrödinger para la mecánica cuántica que fuese consistente con los principios de la relatividad especial. La ecuación de Schrödinger gobierna la función de onda, una magnitud numérica que depende del tiempo y de la posición en el espacio y cuyo cuadrado en cualquier instante y posición expresa la probabilidad de encontrar una partícula determinada en esa posición y en ese instante. La ecuación de Schrödinger no trata el espacio y el tiempo de manera simétrica, como requeriría la

relatividad especial. En lugar de ello, la tasa de cambio de la función de onda con el tiempo depende de la segunda derivada de dicha función con respecto a la posición (es decir, de la tasa de cambio respecto a la posición de la tasa de cambio de la función de onda respecto a la posición). Dirac observó que la versión relativista de la ecuación de Schrödinger para una partícula sin espín (la ecuación de Klein-Gordon) no era consistente con la conservación de la probabilidad, el principio por el que la probabilidad total de encontrar la partícula en alguna parte ha de ser siempre del cien por cien.

Dirac logró elaborar una versión relativista de la ecuación de Schrödinger consistente con el principio de conservación de la probabilidad —la que hoy conocemos como ecuación de Dirac—, pero que describía una partícula con un espín igual a un medio (en unidades de la constante de Planck), y no a cero. El hecho fue considerado un gran triunfo, ya que, a partir de la interpretación de los espectros atómicos realizada unos años antes, se sabía que el electrón tenía un espín de valor un medio. Por otra parte, al estudiar el efecto de un campo electromagnético externo sobre su ecuación, Dirac pudo demostrar que el electrón es un imán que posee la fuerza magnética que los experimentadores holandeses Samuel Goudsmit y George Uhlenbeck habían inferido de los datos espectroscópicos y fue capaz de calcular la «estructura fina» del hidrógeno, las minúsculas diferencias de energía entre estados que difieren sólo en su momento angular total. Cuando Dirac murió en

1984, una de las reseñas necrológicas le atribuía el haber explicado por qué el electrón tiene que tener un espín de un medio.

El problema es que no existe una teoría cuántica relativista como la que buscaba Dirac. La combinación de relatividad y mecánica cuántica conduce inevitablemente a teorías con un número ilimitado de partículas. En estas teorías, las verdaderas variables dinámicas de las que depende la función de onda no son las posiciones de una o varias partículas, sino *campos* del tipo del campo electromagnético de Maxwell. Las partículas son cuantos —paquetes de energía y momento— de esos campos. Un fotón es un cuanto de campo electromagnético, con espín de valor uno, y un electrón es un cuanto de campo del electrón, con espín igual a un medio.

Si los argumentos de Dirac eran correctos, debían ser aplicables a cualquier clase de partícula elemental. El análisis de Dirac no hacía uso de propiedades especiales del electrón que lo distinguieran de otras partículas: el que, por ejemplo, los electrones posean una masa diminuta y se encuentren orbitando en torno al núcleo en todos los átomos ordinarios. Pero, al contrario de lo que pensaba Dirac, la mecánica cuántica y la relatividad no prohíben la existencia de partículas elementales con espines diferentes de un medio, y se sabe que dichas partículas existen. Además del fotón, con espín unidad, hay partículas masivas como la W y la Z, también con espín unidad y tan elementales como el electrón. Tampoco hay nada en la mecánica cuántica relativista que prohíba la existencia

de partículas con espín nulo. De hecho, estas partículas aparecen en nuestras teorías actuales sobre las interacciones entre las partículas elementales, y los físicos experimentales dedican no pocos esfuerzos a encontrarlas.

El mayor éxito de la teoría de Dirac se considera que es su predicción correcta de la antipartícula del electrón, el positrón, descubierto pocos años después en los rayos cósmicos. Dirac había observado que su ecuación tenía soluciones con energía negativa. Para evitar el colapso de todos los electrones atómicos hacia estados de energía negativa, supuso que esos estados estaban casi todos ocupados, con lo que el principio de exclusión de Pauli (que prohíbe que dos electrones ocupen el mismo estado) preservaría la estabilidad de los electrones ordinarios de energía positiva. Los estados de energía negativa ocasionalmente vacíos serían interpretados como partículas de energía positiva y carga eléctrica opuesta a la del electrón, es decir, como antielectrones.

Pero, desde la perspectiva de la teoría cuántica de campos, no existe razón alguna por la que una partícula de espín un medio deba tener una antipartícula distinta. En algunas de las teorías actuales aparecen partículas con espín un medio que son sus propias antipartículas, aunque ninguna de ellas se haya detectado todavía. Por supuesto, la teoría cuántica de campos nos dice que una partícula con carga eléctrica tiene que tener una antipartícula distinta, pero esto es válido tanto para partículas de espín cero o uno (que no obedecen al principio de exclusión de Pauli) como para

partículas con espín un medio, y dichas antipartículas de espín entero son bien conocidas experimentalmente.

Tras los trabajos de Dirac, las controversias sobre este punto continuaron durante muchos años y podrían incluso proseguir hoy en día. En la década de 1950 se propuso construir en Berkeley un nuevo acelerador que, por vez primera, tendría la suficiente energía como para producir antiprotones. La principal objeción era que, si todo el mundo sabía que el protón tenía que tener una antipartícula, ¿para qué gastar dinero en un acelerador destinado a descubrirla? A ello se respondía que el protón parecía no satisfacer la ecuación de Dirac, ya que poseía un campo magnético considerablemente mayor que el que la teoría predecía, y si no satisfacía la ecuación de Dirac, no había razón para esperar que tuviera una antipartícula distinta. Todavía no estaba claro que la ecuación de Dirac no tiene nada que ver con la necesidad de antipartículas.

Entonces, ¿por qué la ecuación de Dirac funciona tan bien a la hora de predecir la estructura fina del hidrógeno y la intensidad del campo magnético del electrón? Resulta que la fusión de la mecánica cuántica con la relatividad especial requiere que un campo, cuyo cuanto tiene un espín de valor un medio e interacciona sólo con un campo electromagnético externo de tipo clásico, tenga que satisfacer una ecuación matemáticamente idéntica a la de Dirac, pero con una interpretación totalmente distinta. El campo no es una función de onda; no se trata de una magnitud numérica, como la función de

onda de Schrödinger, sino de un operador mecanocuántico que no tiene una interpretación directa en términos de probabilidad de encontrar la partícula en distintas posiciones. Al considerar la acción de ese operador sobre estados que contienen un único electrón, se puede calcular la intensidad del campo magnético de la partícula y las energías de los estados de dicha partícula en los átomos. Al ser la ecuación para el operador del campo del electrón matemáticamente la misma que la de Dirac para su función de onda, los resultados del cálculo resultan ser los obtenidos por Dirac.

Lo anterior es sólo aproximado. El electrón interacciona también con las fluctuaciones cuánticas del campo electromagnético —por lo que su campo magnético y sus energías en los estados atómicos no son exactamente iguales a las calculadas por Dirac— y experimenta, asimismo, interacciones débiles no electromagnéticas con el núcleo atómico. Aunque los cálculos de la estructura atómica basados en la ecuación de Dirac son sólo aproximados, se trata de estimaciones muy buenas que siguen siendo utilizadas.

Así son las cosas. Cuando una ecuación es tan afortunada como la de Dirac, nunca se convierte en un error. Puede dejar de ser válida por las razones que suponía su autor y, sacadas de contexto, llegar a significar cosas que el autor jamás hubiera imaginado. Debemos estar siempre abiertos a la reinterpretación. Las grandes ecuaciones de la física moderna son una parte permanente del conocimiento

científico, algo que podría sobrevivir incluso a las bellas catedrales de los tiempos antiguos.