El sueño de la razón

La lógica matemática y sus paradojas

Javier Fresán

El mundo es matemático

El sueño de la razón

La lógica matemática y sus paradojas

Javier Fresán

El mundo es matemático

Realización: EDITEC
Diseño cubierta: Llorenç Martí
Créditos fotográficos: gettyimages, agefotostock, Corbis

ISBN: 978-84-473-6972-0
Depósito legal: NA-1666-2011

Impreso y encuadernado en Rodesa, Villatuerta (Navarra)

Impreso en España - *Printed in Spain*

Para José Antonio Pascual
y Rosa Navarro Durán

Sumario

Prefacio

Una pareja empieza a discutir: «Es que siempre me llevas la contraria», dice la mujer. «No es cierto», responde el marido. «¿Ves como sí? Tú mismo lo confirmas», vuelve a atacar ella. «Tienes razón, cariño, no hago más que llevarte la contraria», admite el hombre, en un intento de zanjar la discusión. «¡Mira que tiene que ser grave para que lo reconozcas!», grita todavía la mujer antes de salir a dar una vuelta. Son escenas que ocurren a diario, hasta en las mejores familias. Si no las hubiera vivido nunca el filósofo y matemático Bertrand Russell, seguro que no se habría casado cuatro veces. Sus peleas, sin embargo, terminarían de otro modo: después del «Tú mismo lo confirmas», Russell habría permanecido en silencio unos segundos y, quizá tras despedirse con una frase del estilo «eso que dices es muy interesante», se habría encerrado en su biblioteca.

¿A qué? A pensar en las afirmaciones que hablan de sí mismas, en lo verdadero y en lo falso, hasta dar con una paradoja que pondría en duda que las matemáticas de dos mil años fuesen la realización más acabada del sueño de la inteligencia. La paradoja de Russell es uno de los protagonistas de este libro, pero, como no me pareció bien tomar la historia por el medio, tuve que contar antes cómo el descubrimiento de las geometrías no euclideas cambió de forma radical el método axiomático, y cómo la antinomia que puso fin a las «mañanas alegres y seguras» del filósofo inglés arraiga en una tradición que se remonta, al menos, hasta Epiménides de Creta. Por otra parte, la paradoja de Russell no habría pasado de ser una curiosidad si no fuera por las respuestas a las que dio origen. Primero nos ocuparemos de la solución de David Hilbert, uno de los hombres más brillantes de su tiempo, que durante treinta años mantuvo viva la esperanza de que las matemáticas volverían a ser seguras para siempre. Es lo que le habría gustado probar al jovencísimo Kurt Gödel, pero en su lugar descubrió que en la aritmética existen enunciados verdaderos que no son demostrables.

Desde su discreto anuncio en una conferencia celebrada en Königsberg en septiembre de 1930, los teoremas de incompletitud de Gödel han fascinado por igual a científicos y humanistas. Unos han querido ver en ellos la derrota de la razón en el combate que a priori le era más propicio; otros, la prueba irrefutable de la superioridad del ser humano frente a las máquinas. Pero sólo quienes de verdad asimilaron el formalismo de los artículos de Gödel fueron capaces de conducir la lógica hacia nuevos territorios. Reinterpretando precisamente los teoremas de incompletitud, el hombre que había descifrado la diabólica criptografía nazi, el genial Alan Turing, pudo imaginar los antepasados de nuestros ordenadores. De todo esto y de muchas cosas más trata este libro, que no se cierra con el «cero o uno» de las máquinas de Turing, sino que intenta dar un paso más allá para describir el mundo matizado de una de las encarnaciones más recientes del sueño de la razón: la lógica difusa.

Quisiera agradecer a los responsables de la editorial RBA su invitación a escribir este libro. Fueron, de hecho, las palabras «narrativas de divulgación», escondidas en alguno de los correos que intercambié con el editor responsable, las que me sugirieron empezar cada capítulo con unas pinceladas novelescas. Sin las historias de esa Sherezade siglo veintiuno que es mi amiga Laura Casielles, nunca habría sabido relacionar la lógica difusa con los postres de un restaurante japonés. También el arranque del quinto capítulo debe mucho a la fascinación por la figura de Alan Turing de Patricia Fernández de Lis. Gracias a los informes minuciosos con los que Jesús Fresán, David Garcés, Miguel Hernaiz, Victoria Ley Vega de Seoane, Javier Martínez y Luz Rello respondían al instante a mis envíos de los primeros borradores, la exposición ha mejorado mucho. A todos ellos les estoy agradecido, y también a María Aguirre Roquero, Luis Azcárate, Noel Garrido, Geno Galarza, María Ángeles Leal, Carlos Madrid, José María Mateos, Guillermo Rey, Roberto Rubio, María José Soler, Lucas Sánchez y Mikel Tamayo por sus valiosas sugerencias.

Capítulo 1

El método axiomático

Desde los griegos, quien dice matemáticas dice demostración.
Nicolás Bourbaki

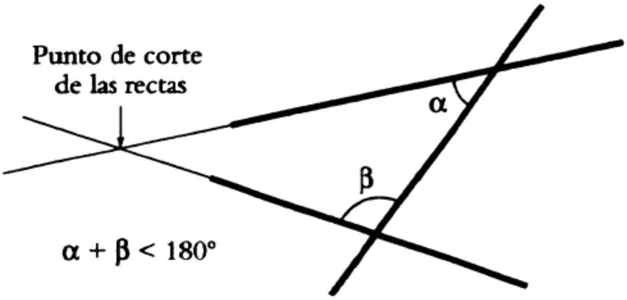
El entusiasmo con el que había desgarrado el sobre, sin buscar siquiera un abrecartas, se fue convirtiendo en decepción a medida que el abogado Taurinus avanzaba en la lectura de las dos páginas de apretada caligrafía que había recibido aquella mañana de noviembre de 1824. La carta contenía la respuesta de Carl Friedrich Gauss al anuncio de un descubrimiento de extraordinaria relevancia: la demostración del quinto postulado de Euclides.

A sus casi cincuenta años, no había rama de la física ni de las matemáticas a la que Gauss no hubiese dado un vuelco, con un sinfín de aportaciones que le valdrían el título de *princeps mathematicorum*, el príncipe de los matemáticos. Sin embargo, ninguna de sus obras atacaba de frente la cuestión palpitante del momento: ¿era verdad el quinto postulado? ¿Por un punto exterior a una recta dada se podía trazar una y sólo una paralela? Aclarar esta cuestión suponía de algún modo responder a la pregunta ¿qué forma tiene el mundo?

El origen de la historia de Euclides y de la obra que plasma sus ideas, los *Elementos de geometría*, se remonta al año 300 a.C., en torno al cual este matemático griego del que nos ha llegado poco más que el nombre había compuesto un manual de geometría que sistematizaba el corpus de conocimientos transmitidos oralmente durante los siglos anteriores por los pitagóricos y los discípulos de Platón. Pero, al contrario que a la Academia del filósofo, en cuyo frontispicio podía leerse «No entre nadie que no sepa geometría», a los *Elementos* de Euclides se acercaban los lectores para aprender la ciencia de las formas a partir de los principios más básicos. Con la doble intención de allanar el camino a los alumnos y de poner rigor y orden en la tradición científica, Euclides abrió su tratado con una serie de definiciones y de axiomas que permitían deducir, con buenas dosis de paciencia, cualquiera de los cientos de proposiciones recogidas a lo largo del libro. Tal vez ninguna otra propuesta pedagógica haya tenido consecuencias tan radicales para el pensamiento en más de dos mil años.

Euclides pintado por Rafael en La escuela de Atenas. *El matemático griego, que aparece midiendo con un compás, se encuentra rodeado por sus discípulos.*

Los diccionarios definen los axiomas como enunciados evidentes que se admiten sin necesidad de justificación. En este sentido, no forman parte de la herencia de la cultura, sino de las conclusiones a las que un hombre alejado de la civilización podría llegar sin gran esfuerzo. Euclides distingue entre nociones comunes y postulados: mientras axiomas del estilo «si dos cosas son iguales a una tercera, entonces son iguales entre sí» lo mismo sirven para hablar sobre los polígonos regulares que sobre los dioses, los postulados son específicos de la geometría. Cinco le bastan al sabio alejandrino como pilares de los *Elementos*: los tres primeros afirman que se puede trazar la recta que pasa por dos puntos, prolongar cualquier segmento rectilíneo y dibujar una circunferencia de centro y de radio arbitrarios; el cuarto establece que todos los ángulos rectos son iguales; y, según el quinto, aquel que había tenido ocupado durante meses a Taurinus, si una recta corta a otras dos de modo que los ángulos interiores del mismo lado sumen menos de 180°, entonces las dos rectas se cortarán en un punto situado en la misma mitad del plano en la que están los ángulos.



Las dos rectas se cortan en la misma mitad del plano en la que están los ángulos.

Es probable que la primera impresión del lector moderno coincida con la que tuvieron los contemporáneos de Euclides: el quinto postulado no es tan evidente como los anteriores, y para entenderlo es necesario tomar papel y lápiz. Así se explica que desde muy pronto los geómetras discutieran su condición de axioma e intentaran demostrarlo a partir de los demás. Aunque antes o después todas sus ten-

UN DIÁLOGO DE LA PELÍCULA *ÁGORA* (ALEJANDRO AMENÁBAR/MATEO GIL, 2009)

Hipatia: Sinesio, ¿cuál es la primera regla de Euclides?

Sinesio: ¿Por qué esa pregunta?

Hipatia: Sólo contéstame.

Sinesio: «Si dos cosas son iguales a una tercera, todas son iguales entre sí.»

Hipatia: Bien. ¿Y no sois ambos semejantes a mí?

Sinesio: Sí.

Hipatia: ¿Y tú, Orestes?

Orestes: Sí.

Hipatia: Quiero deciros esto a todos los que estáis en esta habitación: es más lo que nos une que lo que nos separa, y pase lo que pase en las calles somos hermanos. Somos hermanos. Recordad que las peleas son para el vulgo y los esclavos.

Cartel de Ágora, película cuya protagonista es Hipatia de Alejandría.

tativas se revelaron inútiles, por el camino fueron apareciendo expresiones equivalentes al quinto postulado que facilitaban la comprensión de sus consecuencias. Las más célebres afirman que los ángulos de un triángulo suman 180° y que por un punto exterior a una recta cabe trazar una única paralela. Bajo ésta y otras formas semejantes, la duda de si el que pasó a llamarse *postulado de las paralelas* era independiente de los anteriores o si, por el contrario, un argumento ingenioso haría posible eliminarlo de la lista de axiomas, sobrevivió a todos los comentadores griegos, árabes y renacentistas de los *Elementos*.

Cuál sería la sorpresa de Franz Adolph Taurinus aquella mañana de noviembre al descubrir que, en lugar de llenarlo de gloria por haber sido capaz de llegar más lejos que las mejores mentes de la historia, el gran Gauss le confesaba que, después de treinta años dándole vueltas a si Euclides había dicho toda la verdad, estaba seguro de que una geometría que no cumpliera el quinto postulado también era posible. Pero esa nueva ciencia no euclidea debía mantenerse en secreto hasta que todos los detalles de una serie de teoremas que parecían contradecir la imagen del mundo perpetuada durante dos milenios estuvieran listos. No habrían sido bien recibidos por los partidarios de que los triángulos y los círculos con los que estaba escrito el libro de la naturaleza eran tal y como los había descrito Euclides, y no de otra manera. Igual que Aristóteles para los escolásticos, Euclides no era un hombre, sino una rama del saber casi sagrada.

De las geometrías no euclideas a la relatividad

Así podría comenzar una novela *basada en hechos reales*, que en el siguiente capítulo presentaría a Gauss (1777-1855) midiendo el kilométrico triángulo formado por los picos de tres montes alemanes para comprobar de una vez por todas si la geometría del espacio era o no euclidea. En el transcurso de la narración, al príncipe de los matemáticos se le unirían otros personajes, como el húngaro János Bolyai (1802-1860) o el ruso Nikolái Lobachevski (1792-1856), que no tomaron tantas precauciones a la hora de dar a conocer sus descubrimientos.

En los grandes salones de la aristocracia, científicos venidos de toda Europa embelesarían a un público cautivo mostrándole maquetas de raras superficies, en las que los ángulos de los triángulos suman siempre menos de 180°. Y alguien interrumpiría el número de magia para gritar «¡Euclides ha muerto!», mientras un hombre nada revolucionario se llevaría las manos a la cabeza porque «nadie puede servir a la vez a dos señores: si la geometría euclidea es verdadera, entonces hay que echar a la geometría no euclidea fuera de la lista de las ciencias y colocarla junto a la alquimia y la astrología»[1].

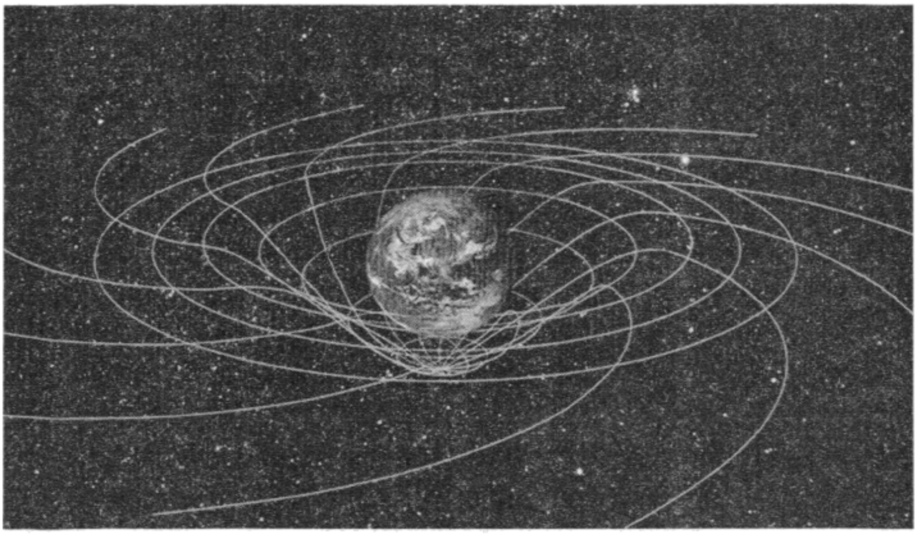
No es ésta la novela que el lector tiene entre sus manos, pero la historia que vamos a contarle también se inicia con el descubrimiento de las nuevas geometrías, y tiene un desenlace aún más inesperado, en el que surgen los ordenadores y se ponen en marcha los primeros experimentos sobre inteligencia artificial. Al margen de las ventanas que se abrían a otros mundos, la implicación más importante de la existencia de modelos no euclideos era de índole filosófica. Euclides había elegido esos axiomas y no otros porque su verdad era evidente. Sin embargo, desde el momento en el que por un punto de algunas superficies pasaban infinitas paralelas a una misma recta, mientras que en otros cuerpos de distinta forma no se podía dibujar ninguna, la pregunta sobre cuál de los axiomas era cierto perdía su sentido. ¿Por qué iba a ser más verdadero el postulado de las paralelas que sus negaciones? En realidad, lo válidos que fuesen unos y otros dependía sólo de los objetos que quisiéramos investigar.

Una de las personas que supo sacar mayor provecho del paisaje después de la batalla fue Albert Einstein (1879-1955), que, gracias a una de las geometrías no euclideas, pudo resolver un problema que le había quitado el sueño al mismísimo Isaac Newton (1643-1727). Según la Ley de la Gravitación Universal, enunciada en 1685 por el científico inglés, dos cuerpos se atraen con una fuerza que es mayor cuanto mayor es el producto de sus masas y cuanto menor es el cuadrado de la distancia que los separa. Este principio permitía explicar de forma unificada el movimiento de los planetas y la caída de los frutos de los árboles, pero dejaba sin respuesta una cuestión fundamental: ¿cómo puede ejercer la Tierra alguna fuerza sobre la Luna si las separan casi 400.000 kilómetros? La acción a distancia se contaba entonces entre los ingenios atribuidos a los alquimistas, así que en ningún caso podía ser aceptada por el pensamiento dominante como garantía del equilibrio del universo. Para evitarla, se rescató de la mitología griega el éter, una sustancia volátil que llenaría los huecos del vacío y por la que la atracción de la gravedad se propagaría de un cuerpo a otro. Sin embargo, varios experimentos pusieron en duda que existiera nada semejante.

Y en ésas llegó Einstein. Cualquiera puede imaginar qué le sucede a una sábana que sujetan dos personas cuando se deja caer en el centro una pelota de jugar a los bolos, pero era necesaria la imaginación del genial empleado de la oficina de paten-

1 Carta de Gottlob Frege (1848-1925) a David Hilbert (1862-1943).

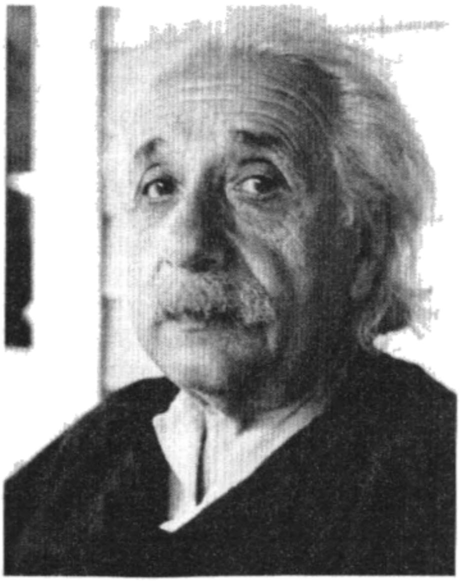
tes de Berna para pensar que lo mismo les ocurriría a los planetas en el espacio. Un cuerpo de masa tan grande como la Tierra deforma el vacío a su alrededor, y la gravedad no es otra cosa que esa curvatura del universo. Del mismo modo que si soltamos una canica en la sábana deformada por la bola, al instante se verá atraída hacia el centro, cuando se deja un cuerpo libre cerca de la superficie terrestre, la pendiente que la Tierra ha creado en el espacio lo obligará a caer. Si el cuerpo está más lejos y se mueve, como la Luna, la deformación del universo no lo hará precipitarse sobre nuestro planeta, sino que lo mantendrá en órbita alrededor de él. Pues bien, resulta que en la geometría que se obtiene al identificar la gravedad con la curvatura del espacio, el quinto postulado de Euclides no es cierto.



Representación gráfica de la deformación del espacio provocada por la fuerza gravitatoria del planeta Tierra.

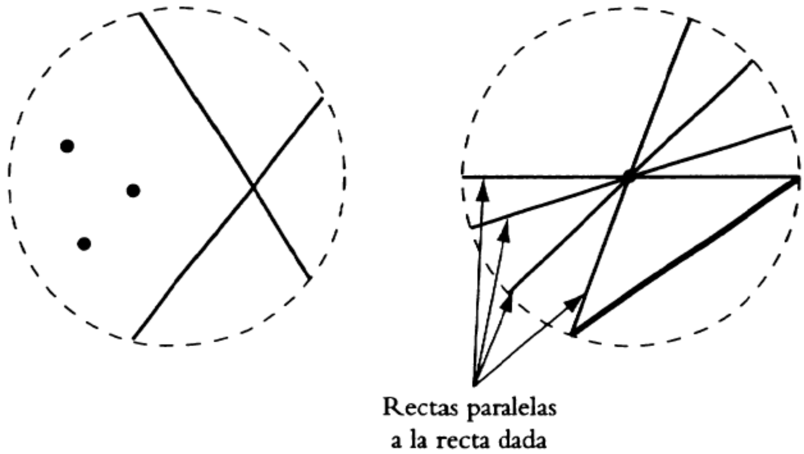
Sin embargo, a Einstein no le preocupaba en absoluto que su teoría de la relatividad hubiese desterrado el sueño de un cosmos euclideo, porque con el tiempo había comprendido que la geometría era algo formal. En el primer capítulo de *Sobre la teoría de la relatividad especial y general*, una exposición divulgativa de sus investigaciones publicada en 1920, Einstein explica que la geometría parte de un puñado de conceptos como «punto», «plano» o «línea recta», de los que tenemos una imagen definida, y de una serie de proposiciones simples, los axiomas, que nos parecen verdaderos al interpretarlos según la idea que nos hemos hecho de los objetos a los que se refieren. A partir de estos principios básicos, el resto de proposiciones se demuestran siguiendo un proceso lógico de deducción, de manera que si admitimos que los razonamientos intermedios son correctos, la verdad del resultado descansa únicamente sobre la verdad de las premisas. Así las cosas, la cuestión sobre qué forma tiene el mundo requiere saber si son ciertos o no los cinco postulados. Pero esa pregunta no sólo no puede resolverse por métodos geométricos, sino que carece de sentido.

Es inútil intentar determinar —prosigue Einstein— si es cierto que por dos puntos pasa una sola recta. Todo lo que sabemos es que la geometría trata de cosas que se llaman «punto» y «línea recta», cuya relación es la siguiente: dos «puntos» distintos determinan una única «recta». Para que tenga sentido discutir la verdad de los axiomas, hace falta establecer primero una correspondencia con la realidad: si cada vez que Euclides dice «punto» y «línea recta», nosotros pensamos en lo que todo el mundo entiende por esas palabras, entonces el axioma «se puede dibujar la recta que pasa por dos puntos» se convierte en una afirmación tangible, y podemos comprobar si es verdadera o falsa, por así decirlo, experimentalmente. Sin embargo, nada indica que la geometría sea esa traducción al lenguaje de los objetos cotidianos, sino sólo un conjunto de relaciones abstractas entre ideas.



Una de las últimas fotografías que se conservan de Albert Einstein, tomada hacia 1950.

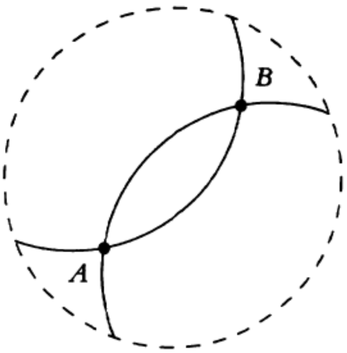
Veamos un ejemplo que aparece por primera vez en dos memorias del geómetra italiano Eugenio Beltrami (1835-1900). El espacio en el que se encuentran los objetos será a partir de ahora el interior de un círculo, sin incluir su borde, y la correspondencia que vamos a proponer es bien sencilla: cuando Euclides dice «punto», nosotros pensaremos en los puntos del interior del círculo, y cuando dice «línea recta», nos imaginaremos los segmentos que empiezan y terminan en el borde. Con esta traducción, dos «puntos» determinan una sola «línea recta» y, por tanto, el primer postulado de Euclides es verdadero. Para ver qué ocurre con el quinto axioma, recordemos que dos «rectas» son paralelas si no se cortan nunca. Tomemos un «punto» cualquiera del interior del círculo, por ejemplo, el centro, y una «línea recta» arbitraria. Al unir el «punto» con los extremos del segmento, obtenemos dos «líneas rectas» que pasan por él y que son paralelas a la de partida, ya que los hipotéticos puntos de intersección están sobre el borde, que ¡no pertenece al espacio! Por tanto, en el modelo de Beltrami, no es cierto el postulado de las paralelas.



El modelo no euclideo de Eugenio Beltrami.

Nótese que en los párrafos anteriores las palabras «punto» y «línea recta» unas veces aparecen entrecomilladas y otras no. Hemos querido distinguir de este modo los conceptos abstractos de «punto» y «línea recta», que podrían tener las más diversas interpretaciones, de los puntos y rectas reales que los originaron. A quien piense que hemos hecho trampa al presentar este juguete no euclideo, quizás un poco de biología le haga cambiar de opinión, pues en el mejor de los casos la vista humana alcanza unos pocos kilómetros de distancia. Como consecuencia de este límite, todas las rectas que se cortan más allá del borde del círculo son iguales para nuestros ojos, y lo que vemos a nuestro alrededor se ajusta razonablemente a la imagen del geómetra italiano. Al fin y al cabo, ¿qué diferencia puede encontrar un europeo entre dos rectas que se cortan en Nueva York y otras cuya intersección queda en Los Ángeles? El pequeño mundo del ser humano no es euclideo. Pero hay más cosas en el cielo y en la tierra de las que sueña su filosofía...

Lo que nos interesa señalar aquí es que la propuesta de Beltrami es sólo una elección arbitraria entre las muchas posibles. En el mismo espacio podríamos haber llamado «rectas» a los arcos de círculo, y entonces el primer postulado no se verificaría, pues son ilimitadas las maneras de unir dos puntos de esa forma. Para determinar totalmente una circunferencia se necesitan tres puntos, y es precisamente esta libertad de elegir el tercero lo que impide que el axioma se cumpla. Si algunos modelos satisfacen el primer postulado y otros no, el hecho de que por dos «puntos» pase una sola «recta» tiene que depender, por fuerza, del significado que se dé a los conceptos «punto» y «recta». Preguntarse por su verdad es tan absurdo como tratar de descubrir si es cierta la profecía «En el año *A* nacerá *B*», donde el lector puede sustituir, si lo desea, *A* y *B* por expresiones suficientemente vagas.



Espacio en el que dos rectas distintas unen los puntos A *y* B
y en el que no se verifica el primer postulado de Euclides.

Esto es a lo que nos referíamos hace un rato cuando dijimos que Einstein era muy consciente de que la geometría es una construcción formal. A pesar de todo, él no se interesaba por las relaciones lógicas entre conceptos, sino por la cuestión concreta de explicar la acción a distancia sin tener que recurrir al éter. Sus «puntos» eran los puntos del espacio, localizados mediante coordenadas que dijesen dónde están y en qué instante nos fijamos en lo que sucede en ellos. Sus «rectas» eran los caminos más rápidos entre dos puntos, que son los que recorre un rayo de luz. Si el instrumento que el físico necesitaba para hacer sus predicciones sobre la naturaleza del espacio era una negación del postulado de las paralelas, ¿por qué no iba a poder usarlo? En mayo de 1919, cuatro años después de que Einstein identificara la gravedad con la curvatura del universo, una expedición a la isla africana de Príncipe consiguió observar cómo se desviaba la luz de las estrellas cercanas al Sol, sólo visibles durante un eclipse. Eran este tipo de comprobaciones experimentales, en consonancia con los resultados teóricos, las que podían decir algo sobre la validez de la relatividad, y no el hecho de que hubiese sido necesaria una geometría no euclidea para enunciar sus leyes.

Por supuesto, Euclides no tenía en mente que sus «puntos» y sus «rectas» se pudieran sustituir por casi cualquier cosa cuando empezó a componer los *Elementos*. Le habría parecido una provocación, un mero engaño del lenguaje, porque para él los ingredientes de la geometría tenían ante todo un significado físico. Valga como prueba la forma en la que enuncia los axiomas, que dicen que, dados dos puntos, se puede dibujar la recta que los une, y no que para todo par de «puntos» existe una única «recta» que los contiene, como tenemos la tendencia a interpretar. En el salto sutil de los puntos a los «puntos» y del «se puede dibujar» al «existe» que separa un enunciado del otro, la geometría se hizo abstracta y nació la lógica matemática.

Los nuevos sistemas axiomáticos

El primer cambio que pedía esta revolución era replantearse el concepto de axioma, pues ya no tenía sentido buscar «verdades evidentes». A partir del nacimiento de las geometrías no euclideas, un axioma no sería más que un enunciado que se coloca por convenio en la base de una teoría para poder deducir teoremas a partir de él. Lo milagroso de una lengua es que permite combinar palabras según se nos antoje, y que mientras se respeten unas cuantas normas, nuestro interlocutor podrá entender una frase, aunque sea la primera vez que se pronuncia. Sin embargo, al inventar una palabra, nos vemos en la obligación de explicar al otro su significado, y es improbable que ésta sobreviva si no hay acuerdo sobre su utilidad o su belleza a la hora de nombrar las cosas. Con la lógica ocurre algo parecido: no se puede demostrar una proposición partiendo de una tabla rasa, sino que es necesario grabar primero en ella unos principios que todo el mundo comparta y unas reglas de deducción o de inferencia gracias a las cuales podremos llegar más allá de los axiomas.

Un ejemplo clásico de estas reglas es el *modus ponendo ponens*, o simplemente *modus ponens*, que en latín significa «modo que afirmando afirma» y que consiste en deducir de la implicación «Si A, entonces B» y de la verificación de A, que B es verdadera. Debemos insistir de nuevo en que el significado de las reglas de inferencia es, como el de los axiomas, puramente formal. Así, la deducción «Todos los hombres pueden volar. Ícaro es un hombre, luego puede volar» es correcta, mientras que «Si llueve, el suelo está mojado. El suelo está mojado, luego ha llovido» no puede considerarse una inferencia válida. Aunque la imagen del suelo mojado a causa de la lluvia sea razonable, y la del hombre volando, completamente absurda, la primera deducción es correcta, mientras que en la segunda se han confundido la causa y el efecto. Que llueva implica que el suelo está mojado, pero que el suelo esté mojado no implica necesariamente que haya llovido; por ejemplo, podrían haber pasado los servicios de limpieza. También existe un *modus tollendo tollens*, o *modus tollens*, esto es un «modo que negando niega», que consiste esta vez en deducir de la implicación «Si A entonces B» y del hecho de que B no se verifique, que tampoco lo hace A, como en el razonamiento «De lo que no se sabe hay que callar. Si hablo es porque sé».

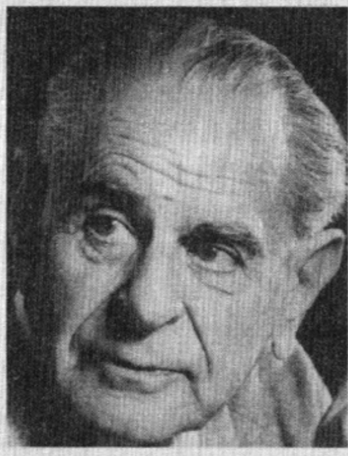
SIMBOLOGÍA LÓGICA BÁSICA

Una forma de recordar la estructura del *modus ponens* y del *modus tollens* es representarlos mediante esquemas en los que una línea separa las premisas de la conclusión. Si denotamos con $\neg A$ y $\neg B$ las negaciones de A y de B, es decir, las proposiciones que afirman lo contrario, entonces el *modus ponens* y el *modus tollens* corresponden a los siguientes diagramas:

$$\frac{\begin{array}{l} A \to B \\ A \end{array}}{B} \qquad \text{y} \qquad \frac{\begin{array}{l} A \to B \\ \neg B \end{array}}{\neg A}$$

En general, una regla de deducción es válida cuando su conclusión es verdadera con independencia de cómo se interpreten las premisas. Así, la inferencia «Si P y Q, entonces R» es correcta si sea cual sea el significado que damos a P, Q y R, siempre que P y Q se verifican a la vez, también es cierta R. Nos encontramos de nuevo ante un criterio formal, que implica, por ejemplo, que la deducción «Si cero es distinto de uno y si uno es igual a cero, entonces usted es mi padre» es válida. Como en ninguno de los mundos posibles cero es igual a uno y diferente de uno al mismo tiempo, las hipótesis no se verifican nunca y no hay nada que comprobar. De esto ya se habían dado cuenta los escolásticos, que acuñaron con gran acierto la expresión *ex contradictione sequitur quodlibet*, es decir, «de una contradicción se sigue cualquier cosa».

EL *MODUS TOLLENS* Y LA «FALSACIÓN»



El filósofo Karl Popper en la década de 1980.

Según el filósofo Karl Popper (1902-1994), el *modus tollens* es la única deducción legítima de las ciencias naturales. Cuando se trata de explicar algún fenómeno, lo que hace el método científico, que Popper llamaba «hipotético-deductivo», es enunciar hipótesis y diseñar experimentos que permitan contrastarlas. Si de una hipótesis *H* se sigue una consecuencia observable *O*, que comprobamos repetidamente en el laboratorio, entonces H se convierte en una ley científica. Pero, a menos que podamos verificar una por una todas las situaciones a las que nuestra hipótesis se aplica, nunca tendremos la certeza absoluta de que se cumple. Para estar seguros de que todos los cisnes son blancos, habría que examinarlos uno a uno en todos los rincones del planeta, pero basta con que aparezca uno solo negro, como les ocurrió a los primeros conquistadores de Australia, para refutar la hipótesis. Este principio se conoce como *falsación* y no es sino un *modus tollens*: «Si la hipótesis *H* es cierta, entonces se seguirá la consecuencia *O*. Como se observa lo contrario de *O*, deducimos que *H* es falsa».

Ahora que sabemos qué son los axiomas y las reglas de inferencia, estamos preparados para precisar los términos «teoría», «demostración» y «teorema», que han ido apareciendo con su significado más o menos intuitivo en las páginas anteriores. Una demostración, que a veces llamaremos prueba, es el proceso que permite obtener nuevos resultados aplicando a los axiomas las reglas de inferencia. En la práctica, consiste en una sucesión finita de afirmaciones, también llamadas enunciados, de las cuales la primera tiene que ser forzosamente un axioma (¡en matemáticas no hay tablas rasas!), y cada una de las siguientes puede ser un axioma o deducirse de las afirmaciones precedentes aplicando las reglas de inferencia. El último enunciado de una demostración se llama teorema, y una teoría es una colección de axiomas, de reglas de inferencia y de todos los teoremas que se pueden demostrar mediante ellas a partir de los axiomas. En algunas ocasiones, en lugar de «teoría» diremos «sistema axiomático».

Hasta el momento hemos centrado la atención en la geometría euclidea, que es la teoría compuesta por los cinco postulados de los *Elementos*, por reglas de deducción como «Si dos cosas son iguales a una tercera, entonces son iguales entre sí» y por todos los teoremas sobre círculos, triángulos y poliedros que al lector le quepa imaginar. También hemos tratado las geometrías no euclideas, que comparten con ella los cuatro primeros axiomas y añaden una negación del quinto (por ejemplo, que por un punto exterior a una recta dada se puedan trazar infinitas paralelas). Pero la auténtica protagonista de este libro es la aritmética, una teoría que se ocupa de los números que sirven para contar, que los matemáticos llamamos *oficialmente* «naturales».

Los axiomas de la aritmética

En vista de todo lo anterior, lo primero que hace falta para definir la aritmética es encontrarle unos axiomas. Esta búsqueda tuvo, a finales del siglo XIX, mayor relevancia de la que hoy podría parecer, porque si el sueño de la primera mitad de la centuria consistía en describir cómo era el mundo, durante la segunda los esfuerzos se centraron en saber con exactitud qué eran los números naturales. Otro tipo de números, como los negativos o las fracciones, se podían obtener sin problema a partir de ellos: así, −1 resulta de añadir un signo menos al número natural 1, y hace falta cuando se quieren distinguir dos direcciones, como en los termómetros o en las cuentas bancarias. Por su parte, 2/3 se forma dividiendo 2 por 3, y responde a otra necesidad: la de repartir cosas cuando el resultado no es exacto. Pero, ¿a qué había que recurrir para explicar unos números que no estaban construidos a partir de otros?

Aquí las soluciones son variadas: Georg Cantor (1845-1918) propuso definir los números naturales como aquello que mide cuántos elementos tiene un conjunto, pero, como veremos en el próximo capítulo, el remedio fue peor que la enfermedad. Esto sin duda habría alegrado a su enemigo acérrimo, Leopold Kronecker (1823-1891), para el que la cuestión de cómo describir los naturales quedaba zanjada con la fórmula «Dios creó los números naturales. El resto son obra del hombre». Menos mal que Giuseppe Peano (1858-1932) era dificilmente impresionable, porque si no, quizá nunca habría sugerido un nuevo punto de vista, que expuso por primera vez en 1889 en un libro titulado *Arithmetices principia: nova methodo exposita*, es decir, *Principios de la aritmética, expuestos según un nuevo método*. Hasta entonces —razona Peano— se había intentado primero definir los números naturales y después encontrar unos axiomas para demostrar teoremas sobre ellos. Pero, ¿por qué no hacerlo al revés? Si se empieza por la lista de axiomas, luego se pueden definir los números como aquellos objetos que los verifican, aunque además de *nuestros* números, tal vez nos aparezcan *otros*...

Portada del libro Arithmetices principia, nova methodo exposita, *de Giuseppe Peano.*

Este giro ingenioso permitió a Peano levantar el edificio de la aritmética partiendo sólo de cinco axiomas, de los cuales el quinto, que se conoce como *principio de inducción*, volvería a ser el único un poco más difícil. Los ingredientes básicos que intervienen en la aritmética son un elemento distinguido, el cero, y una operación, que a cada número natural le asigna otro, que llamaremos su sucesor, o el siguiente. Con este lenguaje, lo que propone el matemático italiano es caracterizar los números naturales como los objetos que cumplen estas propiedades:

a) Cero es un número natural.
b) Cada número natural tiene un sucesor.
c) Cero no es el sucesor de ningún número natural.
d) Dos números diferentes tienen distintos sucesores.
e) Si un conjunto A contiene el cero y cada vez que contiene un número contiene también el siguiente, entonces A contiene todos los naturales.

El primer teorema que puede demostrarse a partir de los axiomas de Peano es el que afirma que uno es distinto de cero, aunque para ello habrá primero que explicar a qué llamamos «uno». Un examen detallado de la demostración nos dará una idea de cómo se manipulan los axiomas y las reglas de inferencia. Como hemos dicho antes, la prueba de que uno es distinto de cero debe empezar por fuerza por un axioma, que en este caso será «Cada número natural tiene un sucesor» (1°). Después podemos echar mano de otro axioma o de un enunciado que se derive de los anteriores por una regla de inferencia. Elegimos de nuevo un axioma, en este caso, «Cero es un número natural» (2°). Ahora el *modus ponens* nos permite deducir de las dos primeras afirmaciones, «Cada número natural tiene un sucesor» y «Cero es un número natural», el tercer enunciado de la prueba: «Cero tiene un sucesor» (3°). Para abreviar, lo llamaremos uno y lo escribiremos 1. Llegados a este punto, podemos seguir con una reescritura del axioma c) bajo la forma equivalente «Si un número es cero, entonces no es el sucesor de ningún número» (4°), y utilizar el enunciado (3°), que ya hemos demostrado en la primera parte y que dice que «Uno es el sucesor de cero». Usando esta vez el *modus tollens*, tenemos que «Si un número es cero, entonces no es el sucesor de ningún número. Como uno es el sucesor de cero, entonces uno no es cero». Y éste es nuestro teorema: «Uno es distinto de cero» (5°).

Con la tranquilidad de que uno y cero son números distintos, lo que podríamos preguntarnos ahora es si los objetos que satisfacen los axiomas de Peano responden a esa intuición de los naturales como una serie que no termina nunca o, dicho de otro modo, si los números son infinitos. Para ello nos bastaría con saber que cada número es distinto de los anteriores. Y ahí es donde desempeña un papel fundamental el axioma de inducción, que permite demostrar teoremas sobre todos los naturales sin tener que comprobarlos para cada uno de ellos. Una forma de entender en qué consiste este principio es imaginarse los números como una hilera de fichas de dominó, entre las que se seleccionan unas cuantas, que distinguiremos del resto dejándolas caer. Con esa imagen, el principio de inducción confirma lo que el lector espera: si tiramos la primera ficha de la fila y si cada vez que cae una cae también la siguiente, entonces en el momento en el que empujemos la primera ficha, caerán todas las demás.

Una vez que hemos demostrado que existe un natural distinto de cero que se llama uno, es posible repetir el argumento para probar que todavía hay otro número diferente del cero y del uno. En efecto, «Cada número natural tiene un sucesor» (1°) y «Uno es un número natural» (2°), luego aplicando el *modus ponens* se deduce que «Uno tiene un sucesor» (3°), al que llamaremos dos. Por el axioma d), que copiaremos en la cuarta línea de la demostración, sabemos que «Dos números distintos tienen sucesores distintos» (4°). Ahora bien, nuestro teorema afirma que «Cero y uno son distintos» (5°), de modo que usando otra vez el *modus ponens*, «El sucesor de cero es distinto del sucesor de uno» (6°), pero esos números no son otra cosa que lo que hemos nombrado, por comodidad, uno y dos. Por otra parte, dos y cero son distintos, porque dos es el número siguiente a uno, y cero no es el sucesor de ningún natural.

Si volvemos a poner en marcha el razonamiento, cambiando uno por dos, demostraremos que existe otro número natural al que podemos llamar tres, y que es diferente de los que han hecho falta para definirlo, es decir, de cero, uno y dos. En general, repitiendo el proceso suficientes veces, podemos deducir que un número particular, por ejemplo, 1.729, es distinto de su sucesor y de todos los anteriores. Gracias al axioma de inducción, sabemos que para demostrar que «Todo número natural es distinto del siguiente» nos basta con ser capaces de probar que uno es distinto de cero (es decir, que la primera ficha cae) y que lo mismo vale para cualquier número concreto y para su sucesor (o, dicho de otro modo, que al tirar una ficha también cae la siguiente).

Los amables lectores que nos hayan acompañado hasta aquí, a estas alturas dudarán de que sea necesaria tanta verborrea para convencerse de algo tan elemental como que dos números naturales son distintos. Y no les falta en absoluto razón, porque ningún padre, por poco que pudiera querer a su hijo, le explicaría de este modo que no da igual tener dos caramelos que tener tan sólo uno. Sin embargo, la lógica no se ocupa de cómo razonamos en la vida diaria, sino de cómo deberíamos hacerlo para estar seguros de que la conclusión a la que llegamos es verdadera. Lo que hemos hecho ha sido vaciar las expresiones «cero», «número» y «siguiente» de todos sus significados intuitivos, hasta reducirlas simplemente a conceptos abstractos que se relacionan entre sí a través de los axiomas y de las reglas de deducción.

¿Qué se puede pedir a los axiomas?

Esta nueva concepción de los axiomas y de las demostraciones transformó las teorías que privilegiaban unas pocas verdades evidentes en sistemas democráticos en los que todos los enunciados tienen el mismo derecho a convertirse en axiomas. Pero eso es sólo a priori, porque al igual que sería insensato permitir que un bebé fuera elegido presidente del gobierno, y no por ese motivo una democracia se volvería menos democrática, si los axiomas no se escogen con cuidado, las teorías que resulten de ellos serán estériles. Euclides tenía claro cómo hacerlo, pero en cuanto desapareció la brújula de la experiencia hubo que encontrar criterios formales sobre la validez de los axiomas, como la consistencia, la recursividad o la completitud.

Para explicar qué significa que un sistema de axiomas sea consistente, nos permitiremos fantasear un poco sobre la tecnología del futuro. Nadie nos impide suponer que dentro de cien años un malvado grupo de científicos diseñará un misil infalible, que alcance cualquier objetivo y lo destruya en cuestión de segundos. O bien podríamos imaginar que, tras una investigación muy costosa sobre nuevas aleaciones, el ejército de los buenos habrá sido capaz de construir un avión a prueba de todo tipo de impactos. Por separado, ninguna de estas historias desentonaría al comienzo de una película de ciencia ficción, pero lo que los espectadores nunca podremos aceptar es que ambas hipótesis se verifiquen a la vez, porque si a alguien más malvado aún que los científicos —por ejemplo, a un lógico— se le ocurriese disparar el misil contra el avión, caeríamos en la paradoja de un proyectil perfecto frente a un blanco indestructible.

En general, decimos que un conjunto de axiomas es consistente si no genera contradicciones, es decir, si de él no pueden deducirse al mismo tiempo un enunciado y su negación. Así, los axiomas «Existe un misil perfecto» y «Existe un avión indestructible» no son consistentes, porque del primero se sigue que, cuando el misil choque contra el avión, éste se destruirá, y del segundo, que permanecerá intacto. La palabra «consistente» es un calco del inglés *consistent*, que significa coherente, vacío de contradicciones. Eso es lo mínimo que puede exigirse a los axiomas, pues en las teorías que no son consistentes cualquier proposición es verdadera, y saber hablar de todo equivale a no poder hablar de nada. El problema es que para tener la garantía de que un sistema de axiomas es consistente, con frecuencia hay que echar mano de teorías más complejas, cuya consistencia genera más preguntas de las que resuelve. Esta tortuga de caparazón gigante que se sostiene sobre otra

EN UN SISTEMA INCONSISTENTE, CUALQUIER ENUNCIADO ES UN TEOREMA

Supongamos que queremos demostrar que un enunciado Q es verdadero. Como el sistema es inconsistente, existirá un teorema P cuya negación *no-P* también será un teorema. Eso significa que podemos encontrar demostraciones de P y de *no-P*. Como ya hemos dicho, al estudiar las reglas de inferencia, la deducción «Si P y *no-P*, entonces Q» es válida, pues sus hipótesis nunca se verifican a la vez. Ahora bien, en las teorías inconsistentes P y *no-P* son teoremas, luego juntando la regla de deducción «Si P y *no-P* entonces Q» y las demostraciones de P y de *no-P*, el *modus ponens* nos permite probar que Q es un teorema. Dicho de otro modo, y por increíble que parezca, en los mundos en los que cero es igual a uno y distinto de uno al mismo tiempo, usted es mi padre (incluso si es mujer). *Ex contradictione...*

tortuga que se sostiene sobre otra tortuga, y así hasta el infinito, será una de las bestias a las que tendrán que enfrentarse los héroes de esta historia.

Para introducir el concepto de completitud, cambiaremos al género policíaco y nos serviremos de un ejemplo que nos ha inspirado el escritor argentino Guillermo Martínez. Imaginen que en una habitación cerrada se comete un crimen y que, al llegar la policía, junto al cadáver hay dos sospechosos. Cada uno de ellos sabe toda la verdad sobre el asesinato: sabe si fue o no fue él. Sin embargo, a menos que confiesen, los inspectores tendrán que encontrar huellas dactilares, restos de ADN o cualquier otra prueba secundaria que permita acusarlos ante un juez. Si esta búsqueda se demostrara inconcluyente, entonces quedarían libres, pero la verdad de lo que sucedió seguiría siempre estando ahí. Aunque la verdad existe, el método es insuficiente para alcanzarla.

Después de un día duro de trabajo, los policías salen a tomar algo para relajarse. Uno de ellos se acaba de incorporar a la comisaría, y los demás apenas lo conocen. Por lo que les cuenta de su vida, pueden deducir que nació en Salamanca, pero que enseguida se mudó a Barcelona, porque a sus padres les gustaba el mar. Sin embargo, con los datos de los que disponen, sus compañeros no consiguen ponerse de acuerdo sobre si está casado o no, y eso que no hay duda de que sólo es posible una respuesta.

Cualquiera de estas situaciones pone de manifiesto que en muchos ámbitos del día a día lo verdadero no coincide con lo demostrable. Eso es a lo que los lógicos se refieren cuando dicen que un sistema de axiomas no es completo. Lo ideal sería que todas las afirmaciones verdaderas sobre ciertos objetos pudieran demostrarse a partir de un puñado de axiomas. Pero eso ocurre raras veces: lo más habitual es que una teoría contenga enunciados que no se pueden demostrar ni refutar, que llamaremos «indecidibles». Por refutar un enunciado se entiende demostrar su negación: por ejemplo, refutar el enunciado «Todos los cisnes son blancos», al que nos hemos referido antes, significa demostrar que «Existe un cisne que no es blanco». Las teorías completas son aquellas que no contienen enunciados indecidibles, o lo que es lo mismo, los sistemas de axiomas en los que cualquier proposición, o bien puede ser demostrada directamente, o bien primero negada y luego demostrada. Los lectores atentos ya habrán advertido que, en esta segunda definición de completitud, el concepto nebuloso de verdad ha sido sustituido por el de demostración. Así se consiguieron resolver algunas de las paradojas que habían preocupado a los filósofos desde la Antigüedad.

Con la mayor parte de las teorías matemáticas ocurre como con el juez que no acierta a decidir si los sospechosos son culpables o inocentes. Por eso sorprenderá que expliquemos ahora que hay siempre un modo de escoger los axiomas para que una teoría resulte completa: consiste en incluir todos los enunciados verdaderos. Con esta axiomatización, las demostraciones sólo tienen una línea, pues lo que se quiere probar ya es un axioma. Si el paraíso de los lógicos son las teorías completas, ¿por qué no hacerlo así? Lo demostrable coincidiría con lo verdadero, y las demostraciones no podrían ser más cortas. Sin embargo, el conjunto de todas las proposiciones verdaderas es demasiado grande para que podamos tomarlas como axiomas. No nos interesa tanto la longitud de las demostraciones como poder comprobar si son correctas mediante algún procedimiento automático. Puesto que en una demostración cada línea es un axioma o se deduce de las anteriores aplicando las reglas de inferencia, para saber si una lista de enunciados demuestra algún teorema es imprescindible que seamos capaces de verificar si una proposición es un axioma. Pues bien, cuando elegimos demasiados, el tiempo que hace falta es infinito.

Diremos que un sistema axiomático es recursivo cuando esto no ocurre, es decir, cuando sea posible comprobar, en una cantidad finita de pasos, si cualquier afirmación es o no es un axioma. La recursividad pone freno a la avaricia del lógico, que quiere demostrar más y más teoremas, pues le impide añadir todos los axiomas necesarios para completar una teoría. Por supuesto, la geometría y la aritmética son teorías recursivas, como lo son, en general, todas aquellas en las que sólo haya un número finito de axiomas. Sin embargo, también existen sistemas recursivos con una infinidad de axiomas. Pero eso no importa, porque lo fundamental de los sistemas recursivos no es cuántos axiomas tengan, sino que la validez de cualquier demostración que se construya a partir de ellos pueda verificarse en un número finito de operaciones.

UN SISTEMA RECURSIVO CON UNA INFINIDAD DE AXIOMAS

Una de las posibilidades para obtener un sistema recursivo con infinitos axiomas consiste en desplegar uno de los axiomas de Peano en infinitas afirmaciones. De algún modo, «Cero no es el sucesor de ningún número» no es más que una forma resumida de decir «Cero no es el sucesor de cero», «Cero no es el sucesor de uno», «Cero no es el sucesor de dos», y así indefinidamente. Supongamos ahora que queremos comprobar si un enunciado es uno de estos axiomas. Desde luego, sólo estará en lista si empieza por «Cero no es el sucesor de...», y si lo que sigue es un número. Recordando que «uno» en realidad significa «el sucesor de cero»; «dos», «el sucesor del sucesor de cero», etc., lo único que hay que hacer es contar cuántas veces aparece la palabra «sucesor» en nuestro enunciado. Por tanto, este sistema axiomático es recursivo.

Recapitulemos. El método axiomático apareció alrededor del año 300 a.C. con los *Elementos*. Euclides consideraba que los axiomas eran verdades evidentes, de acuerdo con nuestra experiencia de las cosas físicas, pero el descubrimiento de otras geometrías distintas de la suya a mediados del siglo XIX echó por tierra esta concepción realista. Desde entonces, los axiomas son sólo unos enunciados que se eligen por comodidad como base de las investigaciones matemáticas. Cuando aplicamos ciertas reglas de deducción a los axiomas, como el *modus ponens* o el *modus tollens*, se obtienen nuevos enunciados verdaderos, que los matemáticos llamamos teoremas. Precisamente la verdad de esos teoremas descansa sobre las demostraciones, que son cadenas finitas de enunciados, de los cuales el primero es un axioma y los siguientes, o bien son axiomas, o bien se deducen de los anteriores mediante las reglas de inferencia. Una teoría es un conjunto de axiomas, de reglas de deducción y de todos los teoremas que se pueden demostrar con esos ingredientes.

La lógica es la rama de las matemáticas que se encarga de estudiar las teorías, haciendo abstracción de lo que dicen. Ante un sistema axiomático, un lógico no se interesa por su contenido, sino por tres cuestiones formales: la consistencia, la recursividad y la completitud. La primera de ellas asegura que en el seno de nuestra teoría no se producirán contradicciones, y es la garantía mínima para poder levantar el edificio de las matemáticas sobre las bases más seguras. La recursividad, por su parte, controla que no haya demasiados axiomas, porque de ser así, correríamos el riesgo de no saber determinar si una demostración es o no correcta. Finalmente, la completitud de una teoría nos dice cuándo sus axiomas bastan para deducir todas las afirmaciones verdaderas sobre el mundo al que se refieren, o dicho de otro modo, sin echar mano del concepto de verdad, cuándo puede demostrarse o refutarse cualquier proposición.

En el siguiente capítulo estudiaremos una serie de paradojas que, a finales del siglo XIX, hicieron que los cimientos de las matemáticas de más de dos mil años se tambalearan. Por fortuna, pronto se propusieron varias soluciones, a las que no bastaba con que los axiomas parecieran consistentes: ¡había que demostrar que, en efecto, lo eran! De este *programa formalista* nos ocuparemos en el capítulo 3. Y con todo ese bagaje, podremos entregarnos luego a uno de los resultados más bellos de la lógica: el teorema de incompletitud de Gödel, que establece un equilibrio entre las nociones de consistencia, recursividad y completitud.

Capítulo 2

Las paradojas

La paradoja es la pasión del pensamiento.
Søren Kierkegaard

Aunque los padres del joven Bertrand habían querido asegurarse en su testamento de que se educaría al benjamín de la familia Russell en los mismos principios por los que ellos habían luchado en la Inglaterra victoriana, la combativa abuela paterna no iba a permitir que los monstruos del ateísmo poblasen la razón de aquel niño de mirada inteligente. Se sucedieron las institutrices que, en el rigor de la casa familiar, conducirían a Bertrand por los caminos de la religión y de las lenguas, aunque también propiciaron que el joven aristócrata adquiriera un excelente dominio del francés, del alemán y del italiano, lo que le permitiría viajar cómodamente por todo el mundo años más tarde. Pero en aquellos lejanos días de su infancia, Bertrand sólo pensaba en los exóticos caracteres griegos, tan adecuados para expresar sus desdichadas reflexiones sobre sí mismo y sobre la vida que le había caído en suerte.

La melancolía no desapareció con su traslado a la academia de Old Southgate para preparar los exámenes de ingreso en la Universidad de Cambridge. ¡Y eso que Bertrand había pensado que el contacto con otros chicos de su edad le ayudaría a soportar el peso de su tristeza! Se había imaginado un ambiente idílico en el que podría leer a los grandes poetas ingleses y comentarlos con otros estudiantes, o discutir hasta el amanecer los problemas filosóficos que le intrigaban. En cambio, se encontró con un grupo de jóvenes violentos, que sólo buscaban emborracharse y perseguir atolondradamente a las mujeres, que no perdían la más mínima ocasión para reírse de la timidez de ese muchacho crecido entre algodones. Como un desalentado héroe romántico, muchas tardes Bertrand recorría las praderas camino de New Southgate para contemplar la puesta de sol mientras pensaba en el suicidio.

Si no lo hizo, no fue por falta de valor, sino porque a los once años su hermano Frank le había abierto las puertas de un paraíso en el que refugiarse, del que le quedaba aún mucho por explorar. Tan deslumbrante como un primer amor fue la

Bertrand Rusell en 1893, a la edad de veintiún años, investido «bachelor» en matemáticas en el Trinity College de Cambridge.

entrada del joven Bertrand en el jardín de los *Elementos* de Euclides, al que se acogía siempre que la hostilidad del mundo se le hacía insoportable. Pero esa felicidad la desfiguraba la idea de que aunque, según le habían contado, el sabio griego lo demostraba todo, sin embargo, lo primero que exigían del lector aquellas páginas que Frank le desgranaba cada tarde era hacer un acto de fe como los que le pedía su abuela: «Punto es lo que no tiene partes». ¿Y si las tuviera? «Se puede trazar la recta que pasa por dos puntos». ¿Y si no fuera posible? De mala gana, Bertrand tuvo que aceptar las recomendaciones de su hermano, que intentaba sin éxito hacerle ver que si no creía en los axiomas, difícilmente podrían continuar con el estudio.

Pasado el tiempo, doce años después de su llegada a Old Southgate, Bertrand volvía a estar tan bloqueado como cuando caminaba pensando en el suicidio. Entre esos dos momentos habían ocurrido muchas cosas: se había licenciado en matemáticas y en filosofía por la Universidad de Cambridge, donde una sociedad secreta formada por los mejores estudiantes, que se hacían llamar los Apóstoles, le había brindado al fin las miles de horas de conversación que años atrás pensaba que encontraría en la Academia; había viajado, publicado sus primeros libros —sobre la socialdemocracia alemana y sobre los fundamentos de la geometría—, y aún le había sobrado algo de tiempo para casarse con Alys Pearsall, la hija de una familia de cuáqueros estadounidenses. Con todo, su principal ocupación seguían siendo las matemáticas, y su objetivo, reducir incluso los axiomas de la geometría a las leyes de la lógica, para que nunca más hubiera que creer nada porque sí.

Intentando deducir de la lógica todas las matemáticas, Bertrand se había topado con una contradicción que parecía a simple vista una de esas adivinanzas del estilo «¿Se puede casar un hombre con la hermana de su viuda?». En este caso bastaba examinar con atención el significado de cada término para dar con la trampa; pero la solución de la contradicción que le preocupaba iba a exigirle mucho más esfuerzo: día a día, durante dos veranos, sentado ante una hoja en blanco, había visto pasar las largas mañanas y las lentas tardes, para ver atardecer con el papel intacto, antes de llegar Bertrand Russell a la convicción de que no existe el conjunto de todos los conjuntos que no se pertenecen a sí mismos.

La teoría de conjuntos

Para entender en qué consiste la paradoja que supuso el fin de «las mañanas alegres y seguras» de Bertrand Russell, necesitaremos primero unas pinceladas de teoría de conjuntos. En el capítulo anterior quisimos mostrar cómo la estructura básica del método axiomático aparecía ya en los *Elementos*, aunque para Euclides los axiomas fuesen verdades evidentes y no principios que se admiten por comodidad. Sin embargo, con el tiempo el lenguaje euclideo se había revelado insuficiente para transmitir las nuevas ideas matemáticas. Demostrar únicamente con palabras y con figuras los profundos teoremas del siglo XIX habría sido tan difícil como lo sería hoy, para nosotros, traducir a una lengua muerta las instrucciones de un iPhone.

Poco a poco, la escritura matemática se fue haciendo más simbólica: no sólo se disponía de una potente notación para las series, las derivadas o las integrales, sino que, gracias a las investigaciones del matemático inglés George Boole (1815-1864), también los enunciados lógicos podían escribirse en forma de ecuación. La geometría estudiaba las formas del espacio; la aritmética, los números; el análisis, las herramientas necesarias para formalizar las leyes de la física, y el álgebra, las ecuaciones. ¿Sería posible encontrar un lenguaje común a todas estas disciplinas que pusiera en evidencia la unidad de las matemáticas?

EL ÁLGEBRA DE BOOLE

George Boole fue el primero en darse cuenta de la analogía existente entre los conectores «o» e «y» de la lógica y las operaciones de sumar y de multiplicar del álgebra. También introdujo los símbolos 0 (falso) y 1 (verdadero) para los dos valores de verdad posibles. Antes de ver un ejemplo, recordemos que, al multiplicar dos números, el resultado es cero sólo si alguno de ellos es cero. Supongamos que se quiere traducir al álgebra la proposición «Todos los hombres son mortales». Boole propone llamar p al valor de verdad del enunciado «ser hombre», y q, al de «ser mortal». Con esta astucia, todo el contenido de la frase podía reducirse a la ecuación $p \cdot (1-q)=0$. En efecto, si alguien es un hombre, entonces p toma el valor de verdad 1 (verdadero). La ecuación nos dice que el producto de los números p y $1-q$ es cero. Como p es distinto de cero, sólo queda la posibilidad de que $1-q$ valga 0. Pero eso significa que q es igual a 1 (verdadero), es decir, que el hombre es mortal.

George Boole, uno de los precursores del álgebra computacional.

Reflexionando sobre un problema que a priori nada tenía que ver con esta cuestión más filosófica que matemática, Georg Cantor creyó haber encontrado una respuesta en la teoría de conjuntos entre 1878 y 1884. Intuitivamente, un conjunto no es más que una colección de objetos: hablamos del conjunto de los animales, del conjunto de los parques de París o del conjunto de lectores de este libro. Estas colecciones se pueden definir por extensión, es decir, a través de la lista de los elementos que las componen, o por comprensión, indicando cuál es la propiedad común a todos sus miembros. Así, el conjunto de los números naturales (recordemos: son los que empleamos para contar) no es otro que $\mathbb{N}=\{0, 1, 2, 3...\}$; éste es un caso de definición por extensión. Si quisiéramos estudiar sólo los números pares, escribiríamos $2\mathbb{N}=\{0, 2, 4, 6...\}$, o bien $2\mathbb{N}=\{n \in \mathbb{N} \mid n \text{ es divisible por } 2\}$, donde el símbolo $\in$ significa «pertenece», y la barra vertical $|$, «tal que». Se trata ahora de una definición por comprensión, pues consideramos en este caso el *subconjunto* de los números naturales con la propiedad de ser divisibles por dos.

Nada más iniciar sus investigaciones, Cantor cayó en la cuenta de que su nueva teoría trataba al mismo tiempo dos objetos de naturaleza radicalmente distinta: los conjuntos finitos y los infinitos. De hecho, el problema del cálculo del número de elementos de un conjunto, lo que los matemáticos llamamos su *cardinal*, tenía soluciones muy diferentes dependiendo de si el conjunto era finito o infinito. Empecemos por una situación muy simple: supongamos que se quiere saber si dos conjuntos finitos tienen el mismo cardinal, por ejemplo, si hay tantas letras en la palabra «avilantez» como colores en el arco iris. El procedimiento obvio consiste en contar los elementos que integran cada conjunto y luego comparar los resultados: como A-V-I-L-A-N-T-E-Z tiene nueve letras, mientras que rojo, naranja, amarillo, verde, azul, añil y violeta son siete colores, decimos que los dos conjuntos tienen distinto cardinal. Pero ¿qué ocurriría si intentásemos aplicar el mismo método a dos conjuntos infinitos? Todo lo que podríamos concluir en ese caso es que son infinitos, luego, o bien se acepta que desde el punto de vista del cardinal todos los conjuntos infinitos son iguales, y el juego ha terminado, o bien nos vemos obligados a modificar el método.

Volviendo a los conjuntos finitos, veamos qué sucede si en lugar de tratar las dos colecciones por separado vamos extrayendo a la vez un elemento de cada una de ellas: empezaríamos con la letra A y con el color rojo, seguiríamos con la V y con el anaranjado, así hasta llegar a la T, que corresponde al color violeta. En ese momento, uno de los dos conjuntos ya se ha terminado, pero del otro quedan todavía dos elementos, las letras E y Z, luego podemos concluir que su cardinal es mayor. Lo que hemos intentado hacer aquí, sin éxito, se llama en matemáticas establecer una biyección entre dos conjuntos y significa asignar a cada elemento de un conjunto X un elemento de otro conjunto Y *uno a uno*, es decir, de modo que se cumplan los siguientes requisitos:

1. No existen dos elementos de X a los que les corresponda el mismo elemento de Y.
2. Todos los elementos de Y están emparejados con algún elemento de X.

Con esta terminología, dos conjuntos tienen el mismo cardinal cuando existe una biyección entre ellos. Es fácil comprobar que entre dos conjuntos finitos con distinto número de elementos no se puede definir una biyección, pues necesariamente varios elementos de X irían a parar al mismo término de Y, o bien algún elemento de Y se quedaría sin emparejar.

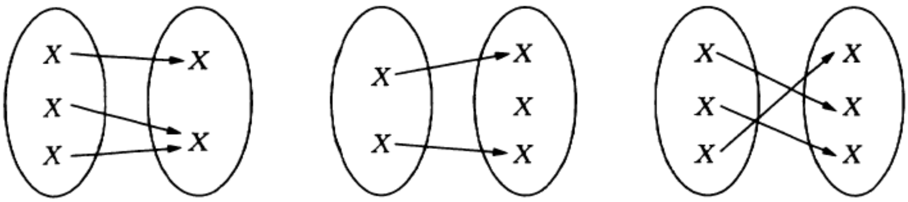


Fig. 1 Fig. 2 Fig. 3

Tres ejemplos de correspondencias entre conjuntos finitos, de los cuales sólo el último (fig. 3) es una biyección, pues en la fig. 1 dos elementos van a parar a la misma imagen, y en la fig. 2 hay un elemento del conjunto de llegada que se ha quedado sin emparejar.

La ventaja de este nuevo enfoque es que podemos generalizarlo a los conjuntos infinitos. Diremos, por tanto, que dos conjuntos infinitos tienen el mismo cardinal cuando exista una biyección entre ellos. La primera consecuencia tal vez sorprenda a los lectores: hay la misma cantidad de números pares que de números pares e impares todos juntos. ¿Cómo es posible? Según nuestra definición, para demostrar este resultado tan poco intuitivo, bastaría con definir una biyección entre los números naturales y los números pares. Proponemos la siguiente: hacemos corresponder al 0 el 0; al 1, el 2; al 2, el 4, y en general, a cada número n, dos veces ese número. Enseguida se comprueba que con este método números distintos se envían siempre a números distintos, y que cualquier número par aparece antes o después emparejado con su mitad. Como se verifican las propiedades 1 y 2, ¡hay la misma cantidad de números pares que de números!

Permítannos reformular aún el resultado: «En un hotel con infinitas habitaciones siempre hay sitio para nuevos huéspedes, aunque el hotel las tenga todas ocupadas». En efecto, en los hoteles que cuentan con un número finito de habitaciones, cuando todas están ocupadas, en el mejor de los casos el recepcionista nos dará la dirección de otro hotel donde pasar la noche. Esto nunca ocurre en los hoteles infinitos: como hay el mismo número de habitaciones que de habitaciones pares, podríamos usar la biyección que hemos construido para cambiar al huésped de la primera habitación a la segunda; al de la segunda, a la cuarta, etc., y liberar así todas las impares. No sólo hemos hecho sitio para el viajero que no había reservado, sino también para infinitos clientes en su misma situación. Convendría que los empresarios tomaran nota...

Lejos de ser una mera curiosidad de los números pares, la existencia de estos hoteles que no se llenan nunca es la característica esencial de los conjuntos infinitos, como observó Richard Dedekind en su artículo «¿Qué son y para qué sirven los números?», publicado en 1888. Un conjunto es infinito si puede ponerse en biyección con una parte de sí mismo que no contenga todos los elementos. Está claro que esto nunca ocurrirá si partimos de un conjunto finito, pues al prescindir de algunos elementos el resultado no podrá ponerse en biyección con el total (como dijimos más arriba, dos conjuntos finitos, de m y de n elementos respectivamente, sólo pueden ponerse en biyección si $m = n$). Sin embargo, los números naturales son infinitos porque una parte estrictamente contenida en ellos, los números pares, tiene el mismo cardinal que todo el conjunto. La nueva definición, por tanto, coincide con el razonamiento basado en los axiomas de Peano que nos permitió demostrar en el capítulo anterior que los números naturales eran infinitos. De hecho, se trata del conjunto infinito más pequeño que nos es dado imaginar. Por eso todos los conjuntos en biyección con los naturales reciben un nombre especial: son los conjuntos numerables, y su cardinal se denota con la primera letra del alfabeto hebreo, la orgullosa *aleph*, con un subíndice que indica que se trata del menor cardinal infinito: $\aleph_0$.

¿Qué significa que un conjunto sea numerable? Como hemos visto, es sólo un modo abreviado de decir que X se puede poner en biyección con los naturales. Así, a cada número natural n le corresponderá un elemento del conjunto que llamaremos x_n, de manera que, por un lado, si n y m son distintos, entonces x_n y x_m son distintos, y, por otro, todos los elementos de X pueden escribirse como x_n para algún n. Cuando salíamos de excursión con el colegio, la maestra solía numerarnos para asegurarse de que todos regresábamos. Antes de montar en el autobús, recitábamos a voz en grito la serie de los números: ¡eeel uuuno!, ¡eeel dooos!, ¡el treees! Todos teníamos un número, y ninguno estaba repetido. También a los conjuntos numerables se les puede hacer gritar su posición en la lista: a la llamada «¡eeel uuuno!» responderá x_1, y cuando anunciemos «¡eeel dooos!» aparecerá x_2. Los conjuntos numerables son los que pueden ponerse en fila. Vimos que los números pares son numerables porque hay un modo evidente de ordenarlos: 0, 2, 4, 6, 8, 10... Lo mismo ocurre con los positivos y con los negativos, pues empezando con el 0 podemos ir saltando a cada lado: 0, 1, −1, 2, −2...

¿Pero acaso no pueden colocarse en fila los elementos de cualquier conjunto? Entonces todos los conjuntos serían numerables, y nada habríamos avanzado con respecto al punto de partida, cuando nos limitábamos a contar ingenuamente. Esta vez el lector puede estar tranquilo, pues uno de los grandes descubrimientos de Georg Cantor fue la existencia de conjuntos que no son numerables. Veamos el ejemplo más sencillo: vamos a considerar el conjunto formado por las sucesiones infinitas de 0*s* y de 1*s*, es decir, por objetos de la forma 0100100010... o 1100101001... Demostraremos que, suponiendo que se trata de un conjunto numerable, llegamos enseguida a una contradicción. En efecto, si fuera un conjunto numerable, podríamos escribir todos sus elementos en una lista como la siguiente:

$$\begin{array}{lccccccc}
\text{Primer elemento} & \rightarrow & \boxed{a_0} & a_1 & a_2 & a_3 & \dots & \dots \\
\text{Segundo elemento} & \rightarrow & b_0 & \boxed{b_1} & b_2 & b_3 & \dots & \dots \\
\text{Tercer elemento} & \rightarrow & c_0 & c_1 & \boxed{c_2} & c_3 & &
\end{array}$$

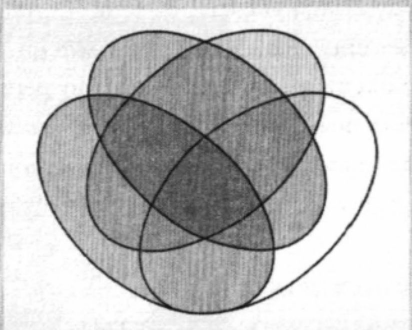
Recordemos que en ella los símbolos a_n, b_n y c_n sólo toman los valores 0 y 1. Vamos a construir un elemento que, a pesar de pertenecer al conjunto de las sucesiones infinitas de 0*s* y de 1*s*, no aparece en nuestra lista. Para hacerlo, nos fijaremos en los términos de la diagonal que hemos destacado con un recuadro. Examinemos, pues, a_0: si vale 0, comenzamos nuestra sucesión por un 1, y si vale 1, por un 0; esto nos determina el primer término. Continuemos con b_1: si vale 0, entonces el segundo término de nuestra sucesión será un 1. Si, por el contrario, es igual a 1, entonces escribiremos un 0. En general, para determinar el término *n*-ésimo de nuestra sucesión miramos el elemento correspondiente de la diagonal y escribimos el símbolo contrario. Así se obtiene una sucesión cuyos únicos términos son 0*s* y 1*s*, y que, por tanto, forma parte del conjunto. Por ejemplo, si el comienzo de la lista fuese

$$\begin{array}{lccccccc}
\text{Primer elemento} & \rightarrow & \boxed{0} & 1 & 0 & 0 & \dots & \dots \\
\text{Segundo elemento} & \rightarrow & 1 & \boxed{1} & 0 & 0 & \dots & \dots \\
\text{Tercer elemento} & \rightarrow & 0 & 0 & \boxed{1} & 1 & \dots & \dots
\end{array}$$

entonces el objeto que estamos construyendo empezaría por 100... Como consiste en modificar los elementos de la diagonal, este método de construir una sucesión de 0*s* y de 1*s* a partir de la hipotética lista se llama argumento diagonal. Lo que queremos ver es que la sucesión que se deduce del argumento diagonal, a pesar de formar parte del conjunto, no aparece en ninguna posición de la hipotética lista que contenía todos sus elementos. En efecto, no puede ser la primera sucesión, porque el primer término es distinto; tampoco la segunda, pues hemos variado el segundo término; ni la tercera, ni la cuarta: cada sucesión de la lista tiene al menos un elemento distinto de la que hemos construido, precisamente el que aparece en la diagonal. Habíamos supuesto que el conjunto de las sucesiones de 0*s* y de 1*s* era numerable, es decir, que todos sus elementos podían colocarse en fila, y hemos llegado así a una contradicción. ¡Eso demuestra que no es numerable!

Si hemos querido dedicar algunas páginas a explicar en detalle los conceptos básicos de la teoría de conjuntos no ha sido sólo para que en la sección siguiente podamos formular con precisión la paradoja de Russell. La prueba de que el conjunto de las sucesiones de 0*s* y de 1*s* no es numerable, que hasta el momento el lector podría considerar una pura exhibición de virtuosismo, nos permitirá, allá por el capítulo 5, demostrar que hay tareas que ni siquiera los ordenadores pueden llevar a cabo, límites al sueño de la razón que da título al libro. Y esperamos, de paso, haber convencido a los lectores de que el mundo de los conjuntos infinitos está lleno de ínsulas extrañas.

LA TEORÍA DE CONJUNTOS EN LA ESCUELA



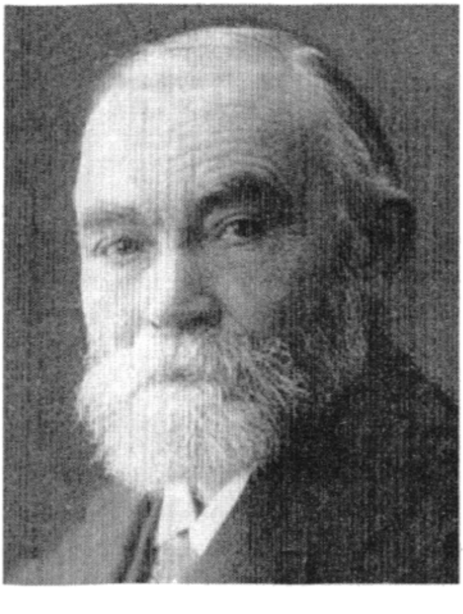
Los diagramas de Venn son el modo más común de representar los conjuntos.

Durante los años setenta, un grupo de discípulos extremistas de la sociedad matemática francesa Bourbaki, que por lo general no eran matemáticos, quisieron llevar la teoría de conjuntos a la escuela primaria. Si el lector sufrió en sus propias carnes este exceso antipedagógico seguro que recordará que los números naturales se explicaban como los cardinales de los conjuntos finitos: 0 es el cardinal del conjunto vacío, y para sumar 2+3 basta unir un conjunto de 2 y otro de 3 elementos; da igual que el resultado se llame 5, lo importante es que 2+3=3+2, ya que no importa el orden en el que mezclemos los conjuntos. Como cuenta Pierre Cartier, secretario de Bourbaki durante aquella época, el desenlace natural de dicha política educativa fueron los niños que volvían del colegio llorando: «Mamá, yo no quiero ser un conjunto».

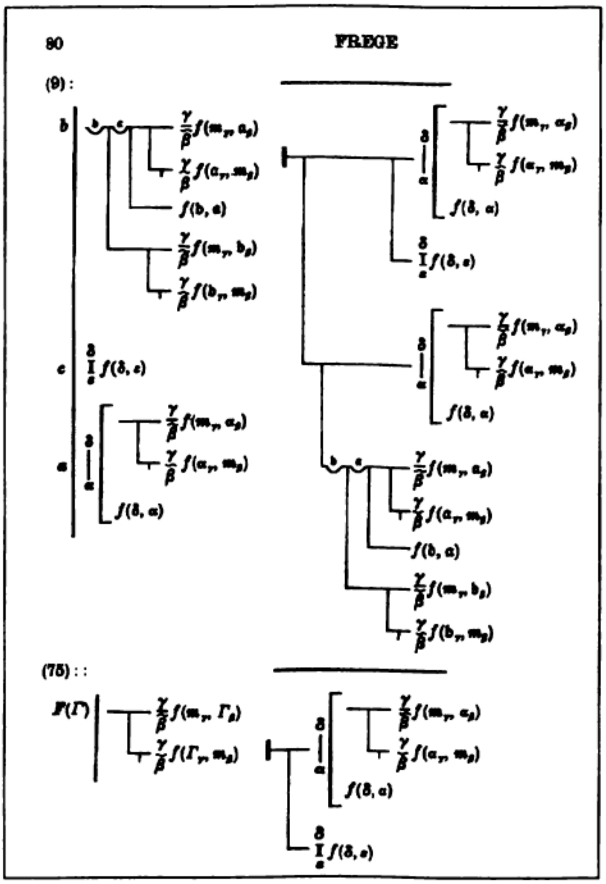
La paradoja de Russell

Bertrand Russell había conocido la teoría de conjuntos en 1896. Al principio le costó aceptarla, porque el autor del libro gracias al cual supo de su existencia estaba entre quienes creían que las ideas de Cantor no eran rigurosas, sino pura teología disfrazada, todo lo contrario de lo que él perseguía por aquella época. Sin embargo, más tarde se dio cuenta de que muchas de las acusaciones eran infundadas, e incluyó el punto de vista del matemático alemán en la última versión de *Los principios de las matemáticas*, publicado en mayo de 1903. Mientras leía la literatura más reciente para incluir aquí y allá algunas notas, Russell descubrió la obra de Gottlob Frege, que había anticipado gran parte de sus descubrimientos veinte años antes. No siempre era fácil reconocer que se trataba de las mismas ideas, porque el complicado simbolismo de Frege, similar al de una partitura de música contemporánea, nada tenía que ver con la notación transparente que Russell había aprendido de Peano.

Estudiando con detalle *Ideografía (Begriffsschrift)*, el libro en el que Frege había expuesto por primera vez sus investigaciones, Russell comenzó a pensar en el conjunto de todos los conjuntos que no son miembros de sí mismos. El conjunto de



Gottlob Frege es el padre de la lógica matemática.



Una página de la Ideografía *de Gottlob Frege.*

todos los gatos ciertamente no es un gato, pero el conjunto de todas las cosas imaginables es a su vez algo imaginable. De estos conjuntos decimos que «se pertenecen» o que «son miembros de sí mismos».

Como el lector convendrá en que se trata de una propiedad un tanto confusa, nos gustaría eliminar de golpe todos los conjuntos de este tipo. Para ello, vamos a llamar *R* (*R* de Russell) al conjunto de todos los conjuntos que no son miembros de sí mismos: en *R* se encuentra el conjunto de los gatos, el de las mesas y, por lo general, todas las colecciones con el buen gusto de no pertenecerse, de manera que estaremos a salvo mientras no crucemos la frontera que separa *R* de los demás conjuntos.

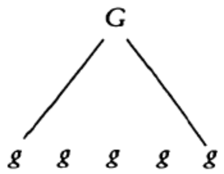


Fig. 1

$x \; x \; x \; x \; X \; x \; x \; x \; x$

Fig. 2

Diferencia entre el conjunto de todos los gatos, que no es un gato (fig. 1)
y el conjunto de todas las cosas imaginables, que vuelve a ser algo imaginable (fig. 2).
(Fuente: Umberto Eco, Vertige de la liste, *París: Flammarion, 2009, pág. 396.)*

La paradoja surge al preguntarnos en qué lado de la frontera está el propio R, porque cualquier respuesta implica la contraria. En efecto, supongamos que el conjunto R es miembro de sí mismo; entonces R satisface la propiedad que hemos intentado eliminar, luego R no puede pertenecer al conjunto de todos los conjuntos que no son miembros de sí mismos. Pero, ¿qué conjunto es ése? ¡R de nuevo! Por tanto, si R se pertenece a sí mismo, entonces R no se pertenece a sí mismo. Hasta ahora no hay problema, porque podría ocurrir que R no fuera miembro de sí mismo y que, partiendo de esta hipótesis, no se llegara a contradicción alguna. Sin embargo, veamos qué sucede al suponer que R no se pertenece. De ser así, R cumpliría la propiedad que define el conjunto de todos los conjuntos que no se pertenecen, luego R estaría incluido en él. Es decir, si R no es miembro de sí mismo, entonces R es miembro de sí mismo. Juntas, ambas conclusiones violan un principio básico que se remonta al filósofo Parménides, que había mostrado en su poema didáctico *Sobre la naturaleza* que no hay vías intermedias entre el ser y el no ser. En su versión matemática, el principio afirma que un elemento pertenece o no pertenece a un conjunto. Como cualquier tercera posibilidad está excluida, los matemáticos nos referimos a él como el axioma del tercio excluso.

Para explicar su paradoja en términos más simples, Russell imaginó un pueblecito cuya ley obliga al barbero a afeitar a todas las personas que no se afeitan a sí mismas y a nadie más. Hemos cambiado la propiedad «ser miembro de sí mismo» por la de «afeitarse», de modo que el barbero cumple ahora el papel del conjunto R. La paradoja nace de la pregunta: ¿Quién afeita al barbero? Porque si él mismo se afeitase, entonces pertenecería al grupo de personas a las que la ley les impide afeitar. Pero si no lo hiciera, entonces tendría la obligación legal de afeitarse a sí mismo. Haga lo que haga, el barbero terminará en la cárcel, donde es posible que un lógico intente convencerlo de que unos pocos años de condena son siempre mejores que descubrir una contradicción que pone en duda la validez de las matemáticas de dos milenios.

Otra versión de la paradoja sustituye al barbero por un ceñudo bibliotecario al que han encargado poner orden en una biblioteca tan grande que hace falta un catálogo que aglutine todos los catálogos. Alguien propone que un buen criterio sería separar los catálogos que no se citan a sí mismos de los que sí lo hacen. Como esta lucha contra el narcisismo bibliográfico le hace gracia, el bibliotecario se pone pronto manos a la obra. Después de trabajar día y noche durante varios años, todas las estanterías han sido examinadas una a una, y sólo queda decidir dónde poner el volumen en el que tanto esfuerzo ha invertido. Si se cita a sí mismo, entonces no puede incluirlo en el catálogo de todos los catálogos que no se citan a sí mismos. Si, por el contrario, el catálogo no fuese una de sus propias entradas, entonces debería formar parte del catálogo de todos los catálogos que no se citan a sí mismos. Si pertenece, no pertenece, y si no pertenece, entonces pertenece. Sólo en ese instante comprendió el bibliotecario que su esfuerzo había sido inútil, pues el criterio nunca permitiría una clasificación completa.

Poco después de descubrir la paradoja, Russell le escribió una carta a Frege, que corregía por entonces las pruebas de la segunda parte de su obra magna, *Los fundamentos de la aritmética*. En ella se incluía un axioma gracias al cual era posible formar el conjunto de todos los objetos que satisfacen una propiedad P, pero Russell había descubierto que ese axioma, aplicado a P = «ser miembro de sí mismo», conducía a una contradicción, pues el conjunto R de todos los conjuntos que no se pertenecen violaba el axioma del tercio excluso. Consternado por la noticia, pero sin perder de vista su rigor característico, Frege añadió un epígrafe en el que confesaba que «nada más triste puede suceder a un escritor científico que ver cómo, después de haber terminado su trabajo, uno de los fundamentos de su construcción se tambalea». Luego proponía modificar su axioma, pero la nueva opción era inconsistente con el resto del sistema, por lo que habría que esperar aún algunos años para que se resolviese la paradoja de Russell.

Entre 1906 y 1908, Russell encontró una solución simple a su paradoja, con la que sentaría las bases de la teoría de tipos. Antes se había preocupado por el problema ontológico que planteaban descripciones como «el mayor número natural» o «el actual rey de Francia», que, siendo gramaticalmente correctas, no se referían, sin embargo, a nada. El caso del «conjunto de todos los conjuntos que no se pertenecen a sí mismos» es aún peor: no es que no exista, sino que ni siquiera la descripción

RUSSELL SOBRE FREGE

En una carta dirigida al historiador de la lógica Jean van Heijenoort el 23 de noviembre de 1962, Russell hablaba de Frege en los siguientes términos:

> «Cuando pienso en actos de gracia y de integridad, me doy cuenta de que no conozco ninguno comparable a la dedicación de Frege a la verdad. Estaba Frege culminando la obra de toda su vida, la mayor parte de su trabajo había sido ignorada en beneficio de hombres infinitamente menos competentes que él, su segundo volumen estaba a punto de ser publicado y, al darse cuenta de que su hipótesis fundamental era errónea, reaccionó con placer intelectual, reprimiendo todo sentimiento de decepción personal. Era algo casi sobrehumano y un índice de aquello de lo que los hombres son capaces cuando están dedicados al trabajo creador y al conocimiento, y no al crudo afán por dominar y hacerse famosos».

que lo define es válida. Sería como hablar de «el Francia de actual rey» o «mayor el natural número».

Según la versión más simple de la teoría de Russell, a cada objeto matemático se le puede asignar un número en función de su complejidad: los elementos tienen tipo 0, los conjuntos de elementos son de tipo 1, los conjuntos de conjuntos de elementos tienen tipo 2, y así sucesivamente. Por ejemplo, si pensamos en los números naturales, el número 8 tiene tipo 0, el conjunto P de todos los pares y el conjunto I de todos los impares son de tipo 1, y al considerar el conjunto $\{P, I\}$ hemos saltado al tipo 2, porque sus elementos son ahora de tipo 1. Una vez que se ha asignado un tipo a cada objeto, existe una regla inquebrantable: únicamente se puede afirmar la pertenencia de un objeto de tipo n a otro de tipo $n+1$. La expresión «el número 8 es par» es correcta, porque 8 es de tipo 0, y P, de tipo 1. Sin embargo, no tiene sentido preguntarse si el conjunto P de los números pares es o no un número par, pues se trata de una relación de pertenencia entre objetos del mismo tipo. Ahora bien, esto era precisamente lo que se describía al hablar del conjunto de todos los conjuntos que no se pertenecen. En el lenguaje de la lógica, «ser miembro de sí mismo» es conceptualmente incorrecto, y así desaparece la paradoja: es cierto que dada una propiedad P se puede considerar el conjunto de los objetos que la cumplen, pero lo mínimo que se puede pedir a P es que esté bien definida.

Ernst Zermelo, el primer axiomatizador de la teoría de conjuntos.

Al mismo tiempo que Russell publicaba «La lógica matemática basada en la teoría de tipos» en el *American Journal of Mathematics*, Ernst Zermelo (1871-1953) proponía una nueva solución a la paradoja, menos conceptual que la de Russell, pero infinitamente más práctica para los «trabajadores matemáticos». Hoy sabemos que una de las grandes dificultades a la hora de poner en marcha una teoría reside en definir su objeto de estudio. Por todas partes se oye hablar de las ciencias de la información, pero ¿qué es la información? Hay quien definiría a los biólogos como los estudiosos de la vida, pero ¿qué es la vida? Son las mismas preguntas que se planteó Zermelo con respecto a la teoría de conjuntos. Según la idea intuitiva de Cantor, los conjuntos no eran más que colecciones de cosas que cumplen una cierta propiedad, pero eso permitía construir el conjunto de todos los conjuntos que no se pertenecen a sí mismos. Sin una definición precisa, no había esperanza de llegar muy lejos. Lo que hizo Zermelo fue sustituir la noción ingenua de conjunto por una lista de axiomas, entre los cuales se incluía uno que impedía formar el conjunto de la paradoja de Russell. A partir de entonces los conjuntos serían los objetos que verifican la lista de axiomas.

La paradoja del mentiroso

Al haber comenzado este capítulo con un análisis de la paradoja de Russell, no quisiéramos que nadie se llevara al engaño de creer que las paradojas son producto de la edad contemporánea. La etimología del propio término, *para-doxa*, que se refiere a aquello que está fuera de la opinión común, apunta por una vez de modo inequívoco a sus raíces griegas. En sentido amplio, una paradoja es el absurdo al que conduce un razonamiento que se inicia con hipótesis plausibles y que continúa con deducciones lógicas en apariencia válidas; así que, cuando Russell se preocupó por el conjunto de todos los conjuntos que no son miembros de sí mismos, tenía a sus espaldas toda una tradición literaria y filosófica. Sin embargo, hasta finales del siglo XIX no parecía posible que las paradojas cruzaran los límites de las humanidades para atacar también el reino de la razón más pura. Los filósofos se habían servido de las paradojas para negar la ilusión de los sentidos, y los poetas, como único vehículo para decir la verdad sobre el amor; pero los matemáticos las temían tanto como a una caja de Pandora que, al abrirse, podría destruirlo todo en un instante. Por eso, el descubrimiento de las contradicciones en las que desembocaba la teoría de conjuntos, en una época en la que la obra de Cantor comenzaba a aceptarse como un lenguaje universal para las matemáticas, supuso una crisis de fundamentos de la que la ciencia tardaría varios años en recuperarse.

Una de las paradojas más antiguas es la de Aquiles y la tortuga, con la que el filósofo presocrático Zenón de Elea, discípulo de Parménides, quiso mostrar que no existe el movimiento, atacando así a los defensores de una concepción atomista del espacio y del tiempo. La ventaja que Aquiles deja a la tortuga para que ambos compitan en igualdad de condiciones —explica el griego— supone una brecha insalvable, pues cuando el atleta haya corrido hasta la posición inicial de la tortuga, ésta ya se habrá desplazado un poco. Cuando Aquiles alcance la nueva posición de la tortuga, tampoco podrá atraparla, porque ella ya se habrá movido un poco. Del espacio que los separa quedará siempre una fracción, por mínima que sea, que impida la victoria del de los pies ligeros.

En otra formulación equivalente, Zenón afirma que una flecha nunca alcanza la diana, porque cuando haya recorrido la primera mitad de la distancia hasta el objetivo, tendrá que recorrer la otra mitad; cuando haya recorrido la mitad de ésta, le quedará aún una cuarta parte; cuando haya cubierto la mitad de esa cuarta parte, le quedará la octava todavía, y así hasta el infinito. Sin embargo, en la práctica Aquiles vence a la tortuga, y la flecha alcanza la diana…

Quizá las paradojas clásicas más intrigantes sean las antinomias, afirmaciones verdaderas y falsas a la vez. Entre ellas destaca la paradoja del mentiroso, atribuida generalmente a Epiménides de Creta, aunque es posible que el filósofo, del que se decía que se había quedado dormido durante 57 años en una cueva bendecida por Zeus, no fuera consciente de que la estaba enunciando. En un verso de su poema *Cretica*, Epiménides ataca a los «cretenses, siempre mentirosos» que no creían en la inmortalidad de Zeus. Pero, siendo él cretense, su afirmación, dicha sobre sí mismo, se convertía en «Yo siempre miento».

Supongamos que Epiménides está mintiendo; entonces lo que dice no puede ser verdad, y como dice que está mintiendo, por fuerza dice la verdad. Si, por el contrario, Epiménides dice la verdad, entonces lo que dice debe ser verdad, y como dice que está mintiendo, por fuerza miente. Según la leyenda, el poeta Filitas de Cos habría muerto de agotamiento al no encontrar una respuesta a la paradoja.

En realidad, «Yo siempre miento» no es una paradoja en sentido estricto, pues su negación no es «Yo siempre digo la verdad», como hemos insinuado en el razonamiento precedente, sino «Yo no miento siempre», o bien «Yo algunas veces digo la verdad». Sin embargo, al poner en boca de Epiménides la afirmación «Esta frase es falsa», se obtiene una auténtica paradoja. En efecto, supongamos que la frase es verdadera: entonces debe ocurrir lo que dice, luego es falsa. Pero si la frase es falsa,

LA ISLA DE LOS CABALLEROS Y DE LOS ESCUDEROS

Un lógico llega a una isla cuyos habitantes se dividen en dos tipos: los caballeros dicen siempre la verdad, y los escuderos siempre mienten. Al encontrarse con los habitantes *A*, *B* y *C*, el lógico le pregunta a *A* si es caballero o escudero, pero la respuesta es tan confusa que no tiene más remedio que repetírselo a *B*: «¿Qué ha dicho *A*?». *B* le responde: «*A* ha dicho que es escudero». Pero justo en ese instante interviene *C* para prevenir al lógico: «¡No creas a *B*, que está mintiendo!».

Con estas dos informaciones, el lógico es capaz de identificar a *B* y a *C*. En efecto, según *B*, el habitante *A* ha dicho «Yo soy escudero», que es otra versión de la paradoja del mentiroso: «Yo siempre miento». Por tanto, la única salida no contradictoria es que *B* haya mentido al transmitir la información de *A*, luego *B* es un escudero. Entonces *C* decía la verdad al avisar al lógico, de lo que se deduce que *C* es un caballero. A menos que sigamos preguntando, nos falta información para determinar qué es *A*.

puesto que no es otra cosa lo que ella afirma de sí misma, por fuerza debe ser verdadera. Si es verdadera, es falsa; si es falsa, es verdadera. Esto viola el *principio de bivalencia*, según el cual una frase es verdadera o falsa, y el *principio de contradicción*, que afirma que ambas situaciones no pueden darse al mismo tiempo.

Cada época ha reinterpretado a su manera la paradoja del mentiroso. Cervantes, por ejemplo, la reelabora en el capítulo LI de la segunda parte del *Quijote*, «Del progreso del gobierno de Sancho Panza, con otros sucesos tales como buenos», para

En la obra magna de Miguel de Cervantes también don Quijote plantea una paradoja a su escudero.

presentarla como ejemplo de las difíciles decisiones que Sancho Panza deberá tomar al frente de la ínsula de Barataria. Antes, en el capítulo XVIII, había explicado don Quijote que entre las ciencias que tiene que saber un caballero andante están las matemáticas, porque «a cada paso se le ofrecerá tener necesidad de ellas». Será así cuando a Sancho Panza se le presente el caso del dueño de una finca separada por un río que obligaba a todo aquel que quisiera atravesarlo a jurar primero a dónde iba. Si decía la verdad, le dejaba pasar, pero si mentía, debía ser ahorcado al momento. Desde que la ley comenzó a aplicarse, los jueces dejaban pasar a casi todos libremente, hasta que un buen día apareció un hombre que juró que iba allí a morir en la horca. Al reparar los jueces en el juramento, dijeron: «Si a este hombre le dejamos pasar libremente, mintió en su juramento, y conforme a la ley debe morir; y si le ahorcamos, él juró que iba a morir en aquella horca, y, habiendo jurado verdad, por la misma ley debe ser libre».

De poco sirve a nuestros propósitos que, puesto que había las mismas razones para ahorcarlo que para no hacerlo, Sancho Panza aconsejara dejar libre a aquel hombre, «pues siempre es mejor hacer el bien que el mal». Lo que sí nos interesa añadir es que las dos paradojas que más fortuna histórica han tenido, la de Aquiles y la tortuga y la del mentiroso, son en realidad muy diferentes. Por un lado, el argumento de Zenón para refutar la victoria de Aquiles sobre la tortuga se basa en un concepto erróneo del infinito. Suponiendo que la ventaja sea de un metro, lo que explica el filósofo presocrático es que Aquiles debe recorrer la distancia

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\frac{1}{32}+\text{etc.}$$

para alcanzar a la tortuga, pues primero debe recorrer la mitad (1/2); luego la mitad de la mitad, es decir, un cuarto (1/4); después la mitad de la mitad de la mitad, o sea un octavo (1/8), y así sucesivamente. Como hay infinitos sumandos, necesariamente esta distancia es infinita, de modo que Aquiles nunca vivirá lo bastante como para recorrerla y vencer a la tortuga. Lo que ignoraba Zenón es que la suma de infinitos números no tiene por qué ser infinita, siempre que éstos se hagan cada vez más pequeños con cierta rapidez. De hecho, un bello argumento geométrico debido a Nicolás de Oresme (1323-1382) muestra que la suma de Zenón no sólo no es infinita, sino que el resultado es 1, exactamente la ventaja que Aquiles había dejado a la tortuga. Por lo tanto, la paradoja de Zenón no es más que una idea equivocada sobre las sumas infinitas.

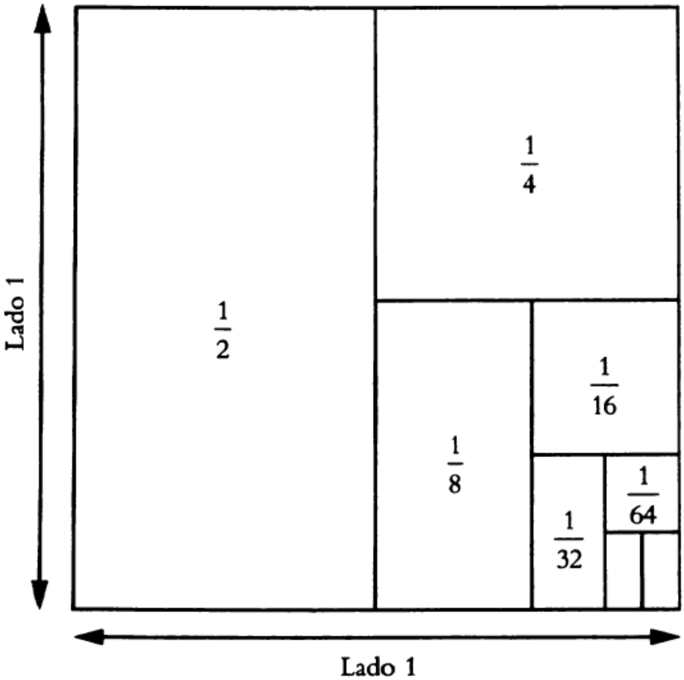


Gráfico con el que Nicolás de Oresme demostró en el siglo XIV que la suma que interviene en la paradoja de Aquiles y la tortuga no vale infinito.

No ocurre lo mismo con la paradoja del mentiroso: «Esta frase es falsa» es un enunciado del que no se puede decidir si es verdadero o falso, porque cualquier respuesta implica la contraria. Como observó el lógico griego Crisipo de Soli, quienes formulan la paradoja del mentiroso «se desvían totalmente del significado de las palabras; únicamente producen sonidos, sin expresar nada». La primera reacción natural consiste en atribuir la contradicción al hecho de que el enunciado hable de sí mismo, pero la autorreferencia por sí sola no basta para explicar la paradoja, porque los enunciados «Esta frase es verdadera» o «Esta frase pertenece al libro *El sueño de la razón. La lógica matemática y sus paradojas*» también son autorreferentes y, sin embargo, no plantean problema alguno.

Otra solución un poco más retorcida era plantearse si el concepto de verdad no sería, como el de conjunto, fácil de utilizar pero difícil de definir. Fue el punto de vista de Alfred Tarski (1902-1983), que en 1933 publicó un artículo de más de doscientas páginas, escrito en polaco, en el que daba la primera definición formal de la verdad. Pese a la extensión del texto, Tarski no se proponía dar a la palabra «verdadero» un nuevo significado, sino sólo capturar matemáticamente la noción aristotélica de verdad como correspondencia entre lo que se afirma sobre la realidad y lo que la realidad es. De la misma forma que «la nieve es blanca» si y sólo si la nieve es blanca, una proposición *P* es verdadera en una teoría si y sólo si al interpretar *P* en la estructura a la que se refiere la teoría, *P* es verdadera. Ahora bien, ¿en qué estructura había que interpretar una sentencia como «Esta frase es falsa»? Como veremos en el capítulo 4, sólo Kurt Gödel conseguiría responder a la pregunta.

Después de todo, la paradoja de Russell, la de Aquiles y la tortuga y la del mentiroso tenían solución, pero por el camino fueron apareciendo muchas más. En 1905, un profesor de un instituto de Dijon, Jules Richard, había descubierto una paradoja en relación con el argumento diagonal de Cantor. Un año después, un joven bibliotecario de la Bodleian Library de Oxford, y no precisamente el que pasaba las noches componiendo el catálogo de todos los catálogos que no se citan a sí mismos, había simplificado la paradoja de Richard imaginando qué pasaría si sólo se pudieran usar quince palabras para describir un número natural. Como la cantidad de expresiones formadas por quince palabras es finita, únicamente podremos describir así una cantidad finita de números. Entre todos los que somos incapaces de describir con este método, habrá uno que sea el menor; llamémoslo *n*. Pero entonces, *n* es «el menor número que no podemos describir con menos de quince palabras», y esta descripción tiene... ¡doce palabras!

¿Cómo estar seguros de que las paradojas no continuarían reproduciéndose igual que un virus? El infinito, la autorreferencia y los conceptos demasiado vagos eran fuente de contradicción, pero ni todos los enunciados autorreferentes daban lugar a ellas, ni parecía posible eliminar el infinito de las matemáticas, ni disponíamos de una brújula que apuntase hacia los conceptos imprecisos. En el próximo capítulo abordaremos la estrategia con la que el matemático más brillante de su generación, David Hilbert, quiso erradicar las paradojas por completo.

Capítulo 3

El programa de Hilbert

Dios existe porque las matemáticas son consistentes,
y el Diablo existe porque no podemos demostrarlo.
Atribuido a André Weil

«¿Quién de nosotros no se alegraría si pudiera levantar el velo tras el que se oculta el porvenir, dejando caer su mirada sobre los futuros avances de nuestra ciencia y sobre los secretos de su desarrollo?»

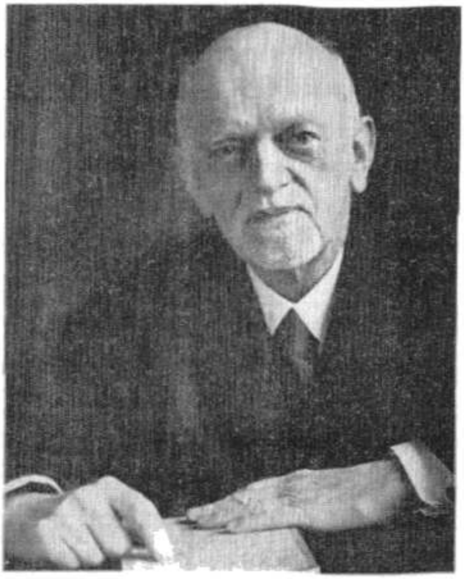
Empezaba un nuevo siglo, y miles de personas recorrían los pabellones de la Exposición Universal de París bajo el rotundo sol de agosto. Mientras tanto, David Hilbert había tomado la palabra en el anfiteatro Chasles de la Sorbona para hablar por primera vez en un Congreso Internacional de Matemáticos no de lo que había sido demostrado, sino de lo que quedaba aún por descubrir. Nadie ponía en duda que Hilbert fuese el mejor matemático de su generación, pero la conferencia había sido relegada a una de las secciones secundarias del congreso, donde la acompañaron un estudio sobre los antiguos geómetras japoneses y la proposición de adoptar una lengua científica común a todos los países. Por supuesto, a Hilbert lo habían invitado a dar una de las conferencias plenarias de la reunión de París, pero el matemático alemán se había retrasado tanto a la hora de elegir el tema, que los organizadores finalmente tuvieron que excluirlo del programa. Al verlo encaminarse hacia el atril con sus inconfundibles anteojos, el público se preguntaba con gran expectación qué habría estado tramando David Hilbert durante todo ese tiempo.

«La historia nos enseña la continuidad del desarrollo de la ciencia. Sabemos que cada época tiene sus propios problemas, y que dependerá de la generación siguiente, ya sea resolverlos, ya sea rechazarlos por inútiles y sustituirlos por otros nuevos.» Hilbert estaba convencido de que el único motor de progreso de las matemáticas era la resolución de problemas. Por eso, al dirigirse al auditorio de la Sorbona, el líder de la escuela de Gotinga insistió mucho en qué significaba realmente resolver un problema, es decir, en la importancia de encontrar un argumento que, partiendo de un número finito de hipótesis formuladas en términos exactos, llegara a la conclusión tras una cantidad finita de deducciones rigurosas. Para ilustrar estas ideas, Hilbert escogió las veintitrés cuestiones que, a su juicio, marcarían el rumbo de los exploradores matemáticos del siglo XX, aunque no tuvo tiempo de comentarlas todas. Gracias al testimonio de sus amigos, los también matemáticos Hermann Minkowski (1864-1909) y Adolf Hurwitz (1859-1919), sabemos cuánto esfuerzo le costó a Hilbert seleccionar los problemas de los que se ocuparía en París. Sin embargo, en ningún momento dudó de la necesidad de incluir uno de ellos. El segundo problema de la lista se preguntaba, con aparente inocencia: ¿Son los axiomas de la aritmética no contradictorios?

EL PROBLEMA DEL CARDINAL DEL CONTINUO

En el capítulo anterior vimos que uno de los grandes descubrimientos de Georg Cantor consistió en demostrar que no todos los conjuntos infinitos tienen el mismo tamaño. En efecto, el argumento diagonal pone de relieve que hay menos números naturales que sucesiones infinitas de 0*s* y 1*s*. El primer problema de la lista de Hilbert pedía responder afirmativa o negativamente a la pregunta de si existe algún conjunto cuyo cardinal sea mayor que el de los números naturales, pero menor que el de las sucesiones de 0*s* y 1*s*. Gracias a los trabajos de Kurt Gödel (1940) y del matemático de la Universidad de Stanford Paul Cohen (1963), hoy sabemos que esta cuestión no puede demostrarse ni refutarse partiendo de la axiomatización habitual de la teoría de conjuntos.

Cuando Hilbert pronunció su conferencia, el 8 de agosto de 1900, habían surgido ya las primeras paradojas de la teoría de conjuntos, pero a Russell le faltaba casi un año para descubrir la contradicción que haría saltar todas las alarmas. En poco tiempo, la difusión de la paradoja del conjunto de todos los conjuntos que no son miembros de sí mismos conmocionaría los círculos matemáticos europeos: en Inglaterra, Whitehead sentenciaría el fin de las «mañanas alegres y seguras»; en Alemania, Frege añadiría el resignado apéndice a sus *Fundamentos de la aritmética*, y en Francia, Henri Poincaré, enemigo de la lógica matemática, repetiría victorioso: «La lógica formal no es estéril: produce contradicciones». Si de algún matemático se esperaba una réplica brillante, era de David Hilbert, al que muchos veían como una auténtica reencarnación de Euclides, después de que hubiera abandonado el estudio de la teoría de números para publicar, en 1899, una axiomatización de la geometría que inauguraba el punto de vista moderno. Sin embargo, Hilbert no se molestó en encontrar una respuesta que pasara a la historia, como las de Whitehead, Frege y Poincaré: no era necesario cuando uno sabía cómo eliminar las paradojas de las matemáticas.



David Hilbert era la persona más indicada para poner fin a las paradojas.

El programa formalista

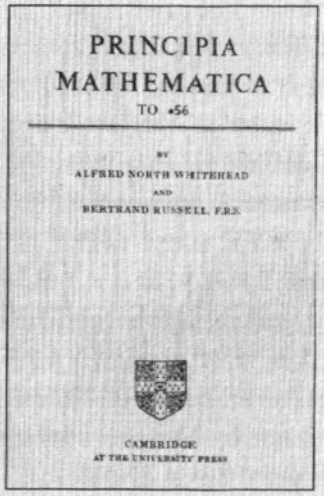
La solución propuesta por Hilbert consistía en dos etapas: en primer lugar, había que formalizar completamente la aritmética, lo cual significaba traducir todo su contenido a un sistema formal. Este proceso debía realizarse con el máximo rigor posible, pero los lógicos no podían detenerse ahí: a la primera etapa tenía que seguir una segunda, en la que se demostrase que esa formalización era, en efecto, consistente. Al contrario de lo que sucedía con la mujer del César, no bastaba que las matemáticas parecieran consistentes, sino que también tenían que serlo, y, para demostrarlo, Hilbert proponía un conjunto de técnicas que llamó *metamatemáticas*. El lector se estará preguntando, con razón, qué diferencia hay entre los sistemas axiomáticos que hemos considerado hasta ahora y los sistemas formales que Hilbert buscaba para la aritmética. Aunque los dos conceptos se parecen mucho, hay un rasgo fundamental que distingue a los sistemas formales: en ellos cualquier afirmación se ha traducido a una serie de símbolos de un lenguaje artificial, que aparecen desprovistos de significado. Lo que pretendía Hilbert se entiende muy bien a la luz de su correspondencia, en la que explica, por ejemplo, que la geometría no cambia si en lugar de «punto», «recta» y «plano», escribimos «amor», «ley» y «deshollinador». Como consecuencia, para un formalista, «capítulo tres» y «capítulo 3» son dos expresiones diferentes, cuya única relación consiste en el hecho sintáctico de que comienzan por la misma palabra.

La base de un sistema formal a la manera de Hilbert es un conjunto de *símbolos primitivos* L que representan el alfabeto de nuestro lenguaje. A partir de ellos, se pueden generar *fórmulas*, que no son más que cadenas finitas de símbolos construidas de acuerdo con una serie de reglas gramaticales. Si, por ejemplo, el lenguaje contiene paréntesis de apertura y de cierre, una de estas reglas podría ser que, por cada paréntesis de apertura, debe aparecer también otro de cierre más a la derecha. Además de especificar el alfabeto, para definir un sistema formal son necesarios unos axiomas y unas reglas de deducción. Los *axiomas* son fórmulas como todas las demás, con la única diferencia de que nosotros les hemos concedido un papel privilegiado. Como indicamos en el primer capítulo, la elección de los axiomas es una de las tareas más difíciles a la hora de poner en marcha un sistema formal: si elegimos demasiados, corremos el riesgo de que se mezclen con las demás fórmulas y ya nunca sepamos distinguirlos; pero si seleccionamos pocos, habrá fórmulas que no se puedan demostrar ni refutar en la teoría. Las *reglas de deducción*, por su parte, son procedimientos que nos permiten obtener nuevas fórmulas partiendo de las existentes. Los axiomas y las reglas de deducción, también llamadas *de inferencia*, se combinan en las *demostraciones formales*, que son cadenas de fórmulas en las que cada una de ellas o bien es un axioma, o bien se obtiene a partir de las anteriores aplicando las reglas de deducción. Como de costumbre, la última fórmula de una demostración se denomina *teorema*.

Por tanto, el primer requisito del programa de Hilbert consistía en describir un alfabeto, unos axiomas y unas reglas de deducción formales para la aritmética. Éste es el empeño al que Bertrand Russell y Alfred North Whitehead consagraron los tres gruesos volúmenes de los *Principia mathematica*, publicados entre 1910 y 1913. En realidad, la propuesta de Russell y Whitehead, que enseguida empezó a conocerse como logicismo, iba más allá del programa formalista: ambos matemáticos no se contentaban con formalizar la aritmética, sino que querían reducirla a la lógica, es decir, definir todos los conceptos de la teoría de los números naturales partiendo de nociones puramente lógicas, y deducir también de estos principios todos los teoremas de la aritmética. Uno de los grandes éxitos de los matemáticos del siglo XIX había sido construir cualquier clase de números a partir de los naturales, de modo que si Russell y Whitehead conseguían su propósito, las matemáticas y la lógica irían siempre de la mano, por un camino libre de contradicciones (o ésa era, al menos, su esperanza).

LAS MATEMÁTICAS NO SON RENTABLES

La obra magna de Russell y de Whitehead fue publicada por la Cambridge University Press. Sin embargo, la prestigiosa editorial no estaba dispuesta a pagar más de trescientas libras por los gastos de edición, lo cual suponía la mitad del coste presupuestado. La Royal Society de Londres, de la que Russell era miembro, se había comprometido a aportar las trescientas libras restantes, pero finalmente sólo fueron doscientas, de modo que Russell y Whitehead tuvieron que pagar cien libras de su propio bolsillo. «Buen balance» —bromearía Russell tiempo después—: «nosotros ganábamos así menos cincuenta libras cada uno por diez años de trabajo.»



En una versión simplificada, el sistema formal que los *Principia mathematica* de Russell y Whitehead proponían para la aritmética está compuesto por los símbolos primitivos 0 (el número cero), s (la función sucesor), $\neg$ (la negación), $\vee$ (la disyunción «o»), $\exists$ (existencia), = (igual) y los paréntesis de apertura y de cierre, a los que luego se añaden variables x, y, z de tipo 0, que representan, por tanto, números naturales, así como variables A, B, C de tipo 1, es decir, conjuntos de números naturales, y así sucesivamente, a medida que sean necesarios nuevos niveles. Es posible que el lector atento haya echado en falta otros símbolos que deberían formar parte del lenguaje: por ejemplo, del mismo modo que hemos incluido el cuantificador de existencia $\exists$ gracias al cual se pueden formalizar afirmaciones como «Existe un número natural con la propiedad P», habría que añadir otro símbolo que significara «para todo», como en «Para todo número natural es cierta la afirmación P». De hecho, ese cuantificador universal existe, y su uso está muy extendido en las matemáticas: «para todo» se escribe $\forall$. Podríamos, en efecto, haber añadido el símbolo $\forall$ al lenguaje, pero en realidad no es necesario, pues «Para todo número natural es cierta la afirmación P» dice lo mismo que «No existe ningún número natural para el que no sea cierta la afirmación P»; luego el símbolo $\forall$ se puede reconstruir a partir de los símbolos de negación y de existencia.

Lo mismo ocurre con la conjunción «y»: existe el símbolo $\wedge$ para representarla, pero es redundante si se dispone ya de $\vee$ y de $\neg$. Para comprobarlo, recurriremos a tres operaciones de la teoría de conjuntos: el complementario, la unión y la intersección.

Dado un conjunto A contenido en otro conjunto B, se llama *complementario* de A en B al conjunto formado por los elementos que pertenecen a B, pero no a A. Por ejemplo, el complementario de las vocales $\{a, e, i, o, u\}$ en el alfabeto son las consonantes. Pasemos ahora a la unión y a la intersección: dados dos conjuntos X e Y, se define su *intersección* $X \cap Y$ como el conjunto de los elementos que pertenecen a X y a Y al mismo tiempo. Por ejemplo, si X fuera el conjunto de los números pares 0, 2, 4, 6, 8, 10... e Y fuese el conjunto de los múltiplos de tres 0, 3, 6, 9, 12, 15..., para calcular la intersección habría que buscar los elementos comunes, que son 0, 6, 12, 18..., es decir, los múltiplos de seis. Por otro lado, la *unión* $X \cup Y$ es el conjunto al que pertenecen todos los elementos de X y todos los elementos de Y. Siguiendo con el ejemplo anterior, la lista de números de la unión de X e Y empezaría por 0, 2, 3, 4, 6, 8, 9...

La gran similitud que existe entre los símbolos que representan la intersección de dos conjuntos ($\cap$) y la conjunción de dos afirmaciones ($\wedge$) por un lado, y la unión de dos conjuntos ($\cup$) y la disyunción de dos afirmaciones ($\vee$), por otro, no es en absoluto casual. Si se asocian a las propiedades P y Q los conjuntos de números que las cumplen, digamos, X e Y, entonces los números que satisfacen P y Q simultáneamente son los elementos de la intersección $X \cap Y$, y los números que verifican P o Q, es decir, al menos una de las dos propiedades, son los miembros de la unión $X \cup Y$. El complementario de un conjunto corresponde, por su parte, a la negación de un enunciado. Para representar el complementario, la unión y la intersección de dos conjuntos, son muy útiles unos diagramas creados por el matemático y filósofo británico John Venn en 1880. Sirviéndonos de ellos, podemos demostrar que la conjunción de las propiedades P y Q equivale a la negación de la disyunción de las negaciones de P y de Q, o dicho de otro modo: $P \wedge Q = \neg(\neg P \vee \neg Q)$ lo cual permite reconstruir $\wedge$ a partir de $\vee$ y de $\neg$.

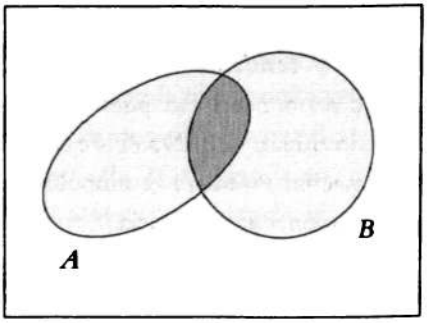


Fig. 1: Intersección de dos conjuntos, correspondiente a la conjunción $P \wedge Q$.

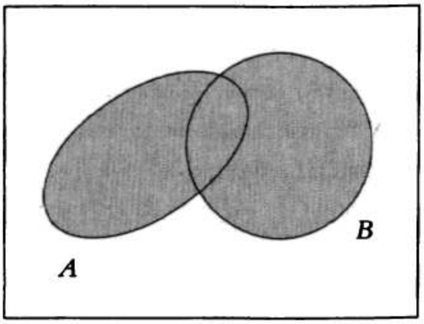


Fig. 2: Unión de dos conjuntos correspondiente a la disyunción $P \vee Q$.

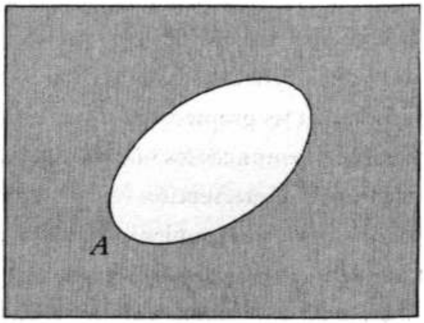


Fig. 3: Complementario de un conjunto correspondiente a la negación $\neg P$.

Diagramas de Venn en los que se muestran las operaciones de intersección (fig. 1), unión (fig. 2) y complementario (fig. 3) en la teoría de conjuntos.

Hecha esta advertencia sobre cómo representar «para todo» y la conjunción de dos enunciados, veamos cómo se traducen al sistema formal de la aritmética algunos de los axiomas de Peano. Recordemos que el primero afirmaba que «Cero es un número natural». Éste no hace falta traducirlo, pues ya hemos incluido el símbolo 0 en nuestro lenguaje. Vayamos con el segundo: «Cada número natural tiene un sucesor». Antes de nada, tenemos que observar que en este axioma intervienen dos variables: el número natural en cuestión, que representaremos por x, y su sucesor, que vamos a llamar y. Recordando que el sucesor de un número se escribe poniendo la letra s delante del número, la relación entre x e y queda expresada a través de la fórmula $y=sx$ es decir, «y es igual al sucesor de x». El siguiente paso consiste en darse cuenta de que «cada número natural» es lo mismo que «para todo número natural», y de que, en este contexto, «tiene» significa «existe». Esto transforma el axioma en «Para todo número natural x existe un número natural y tal que $y=sx$». Si tuviéramos a nuestra disposición el símbolo $\forall$, el trabajo ya estaría terminado, y el axioma se leería $\forall x \exists y (y=sx)$, donde hemos usado los paréntesis para encerrar la propiedad que cumplen los números x e y. Como no es el caso, hay que hacer to-

EL CUARTO AXIOMA DE PEANO

Traduzcamos al sistema formal de la aritmética el cuarto axioma de Peano, que afirma que «Dos números diferentes tienen distintos sucesores». Como antes, lo primero que hay que hacer es identificar las variables que intervienen, que en este caso son dos números naturales x e y. Lo que dice el axioma es que no puede suceder al mismo tiempo que x e y sean distintos y que sus sucesores coincidan, es decir: «No existen números x e y tales que:

1. x sea distinto de y;
2. el sucesor de x sea igual al sucesor de y».

Si el símbolo de conjunción formara parte del lenguaje, el axioma se escribiría del siguiente modo:

$$\neg\exists x \neg\exists y(\neg(x=y)\wedge(sx=sy)).$$

Como no es así, tenemos aún que expresarlo en función de la negación y de la disyunción. Teniendo en cuenta que negar dos veces un enunciado equivale a afirmarlo, el cuarto axioma de Peano se convierte en:

$$\neg\exists x \neg\exists y(\neg((x=y)\vee(\neg(sx=sy)))).$$

davía una última traducción: puesto que «Para todo número natural x, existe un número natural y tal que $y=sx$» dice lo mismo que «No existe ningún número natural x tal que no exista ningún número natural y tal que $y=sx$», el segundo axioma de Peano se escribe: $\neg\exists x\neg\exists y(y=sx)$. Tras esta explicación detallada, el lector puede comprobar que el tercer axioma de Peano, «Cero no es el sucesor de ningún número natural» corresponde a la expresión $\neg\exists x(sx=0)$.

Del lenguaje al metalenguaje

Gracias al proceso que acabamos de describir, se ha conseguido vaciar la aritmética de significado hasta reducirla a su esqueleto formal. Los axiomas ya no describen nada, sino que son únicamente cadenas de símbolos abstractos, y las demostraciones se han convertido en ejercicios de combinatoria. Pero todavía es posible enunciar proposiciones con significado: podríamos decir, por ejemplo, que «El segundo axioma de Peano es más largo que el tercero», que «El cuantificador de existencia aparece dos veces en el segundo axioma de Peano» o que «La fórmula $\neg(0=1)$ es un teorema de la aritmética». Lo importante es que ya no se trata de expresiones formalizadas en el lenguaje L, sino de frases escritas en español que se refieren a las fórmulas de L. Sus protagonistas ya no son los números, sino las proposiciones que hablan sobre los números, de modo que, al enunciarlas, hemos atravesado la frontera de las matemáticas para entrar de lleno en los dominios de las ***metamatemáticas***. El salto es idéntico al que se produce cuando, en una novela, de pronto uno de los personajes comienza a escribir otra novela. Igual que la literatura se convierte a veces en metaliteratura, las matemáticas pueden transformarse en metamatemáticas.

Una de las contribuciones esenciales de Hilbert fue distinguir con claridad los niveles lingüísticos a los que pertenecían los distintos enunciados. Imaginemos una clase de inglés en la que el profesor explica en español los matices de significado de alguna palabra. En ese momento hay dos lenguas en juego: el inglés, que es el idioma que los alumnos quieren aprender, y el español, que les sirve de herramienta. Lo mismo ocurre en una frase como «La fórmula $\neg\exists x\neg\exists y(y=sx)$ es más larga que la fórmula $\neg\exists x(sx=0)$», que combina cadenas de símbolos del lenguaje L con las expresiones «fórmula» y «ser más larga», que no forman parte de L, sino de un ***metalenguaje*** que empleamos para referirnos al sistema formal, por así decirlo, desde fuera. Los términos «cero», «sucesor» o «igual» están autorizados en el lenguaje L, donde se escriben 0, s e =, respectivamente, pero las palabras «fórmula», «demostración» o «verdadero» son propias de un metalenguaje que L no sabe interpretar. Por tanto, cuando se formaliza la aritmética, todas estas afirmaciones dejan de tener sentido dentro de la aritmética.

Pero, ¿qué tiene que ver esto con las paradojas? No olvidemos que el objetivo final del programa de Hilbert consistía en erradicarlas de las matemáticas. Como indicamos en el capítulo anterior, muchas paradojas respondían a fenómenos de autorreferencia, que son posibles en las lenguas naturales, pero que no tenían por qué serlo en el lenguaje artificial de los sistemas formales. Mientras que, al enunciar la paradoja de Russell en español, nos pareció muy razonable que hubiese dos clases de conjuntos —los que eran miembros de sí mismos y los que no—, un sistema formal hubiese detectado de inmediato que la relación de pertenencia aplicada a dos variables del mismo tipo infringía las reglas gramaticales. Más extremo es todavía el caso de la paradoja del mentiroso «Esta frase es falsa». Para que haya que tomársela en serio, no sólo la autorreferencia debería estar admitida en el sistema formal, sino que también la propiedad de «ser verdadero» tendría que poder expresarse en el lenguaje, además de en el metalenguaje. Hilbert tenía la esperanza de que las dos situaciones nunca se produjeran a la vez si se formalizaba convenientemente la aritmética.

Sin embargo, no era suficiente con tener una esperanza, y es aquí donde entraba en juego la segunda fase del programa de Hilbert, que proponía poner fin a la crisis de los fundamentos de las matemáticas demostrando *metamatemáticamente* la consistencia de la aritmética formalizada. Sólo así los matemáticos del futuro podrían tener la certeza de que nunca volverían a encontrarse contradicciones. Por si fuera poco, no todos los métodos estarían permitidos en esta demostración metamatemática: había que servirse sólo de los más seguros, que Hilbert bautizó —sin explicar nunca demasiado bien a qué se refería— con la palabra alemana *finit*, que luego se transformaría en *finitarios*. Estos métodos finitarios debían eliminar cualquier razonamiento que no fuera tangible: no se aceptaban, por ejemplo, las demostraciones por reducción al absurdo, un argumento que ya Euclides había utilizado para probar que existen infinitos números primos o que la raíz cuadrada de dos no se puede obtener dividiendo dos números naturales. El primer paso de una demostración por reducción al absurdo consiste en negar el enunciado que se desea probar. Si, por ejemplo, se trata de demostrar que existen infinitos números primos, entonces la hipótesis de partida será que sólo hay una cantidad finita. A partir de esta suposición, se empiezan a hacer deducciones lógicas correctas hasta que se llega a una afirmación absurda como, por ejemplo, que un teorema de la aritmética que ha sido demostrado independientemente no se verifica. Todo el razonamiento intermedio es válido, luego la única explicación posible del absurdo es que la hipótesis de partida fuese falsa, y es así como hemos demostrado lo que pretendíamos. Muchas veces, a la hora de probar que existe un determinado objeto matemático, pongamos, la solución de una ecuación, es más fácil ver que, si no existiera, entonces llegaríamos a un absurdo, que construirlo («aquí está, ésta es mi solución»). Lo mismo podría suceder con las *metamatemáticas*: quizá no fuésemos capaces de probar un enunciado como «La fórmula *P* es demostrable» encontrando explícitamente una demostración de *P*, pero sí razonando que, en caso de que no existiera, se derivaría una contradicción. Sin embargo, a Hilbert estos procedimientos no le parecían lo bastante seguros.

POINCARÉ CONTRA HILBERT

Henri Poincaré.

Henri Poincaré (1854-1912), al que algunos historiadores han llamado «el último universalista», detestaba a quienes querían reducir las matemáticas a relaciones formales entre símbolos. Cuando, en 1899, Hilbert publicó sus *Fundamentos de la geometría*, Poincaré escribió una larga reseña en la que criticaba al matemático alemán por pretender que «las matemáticas funcionen como una pianola». Algunos años más tarde, sin tener aún muy clara la distinción entre el lenguaje y el metalenguaje, Hilbert intentaría demostrar la consistencia de la aritmética usando el principio de inducción, es decir, el quinto axioma de Peano. Siempre alerta, Poincaré señalaría el círculo vicioso en el que Hilbert caía al tratar de demostrar la consistencia de la aritmética recurriendo al axioma más importante de la aritmética. De nada sirvió que Hilbert se defendiera diciendo que su método no era de inducción, sino de *metainducción*, porque Poincaré estaba en lo cierto, algo que el matemático alemán terminaría reconociendo con la ayuda de su alumno Hermann Weyl (1885-1955).

David Hilbert no era el único que rechazaba los métodos no constructivos. Junto al logicismo y al formalismo, se había desarrollado otra respuesta a las paradojas de la teoría de conjuntos, que suponía eliminar de raíz cualquier uso del infinito. Para los intuicionistas, todos los objetos matemáticos eran producto de la mente humana, luego su existencia equivalía a la posibilidad de construirlos. Los seguidores de esta corriente distinguían el infinito *potencial*, que es el que corresponde a los conjuntos que pueden ampliarse tanto como se quiera, del infinito *actual* de las totalidades acabadas. Ellos admitían que los números naturales eran potencialmente infinitos, pues a cualquier conjunto finito de la forma $\{0, 1, 2, \ldots, n\}$ se le podían añadir aún nuevos números, pero no que se pudiera hablar de todos los números naturales a la vez. Los intuicionistas tampoco aceptaban el axioma del tercio excluso, según el cual si un enunciado no es verdadero, entonces su negación lo es. Al rechazar dicho principio, los seguidores de esta corriente tuvieron que renunciar también a todos los teoremas de las matemáticas en cuya demostración se había empleado. De hecho, el propio fundador de la teoría, el matemático danés L.E.J. Brouwer (1881-1966), se vio obligado a rechazar muchos de sus brillantes resultados anteriores, en los que usaba el axioma del tercio excluso.

Otro ejemplo representativo de las técnicas que los intuicionistas querían eliminar de las matemáticas era el axioma de elección que Ernst Zermelo había propuesto para la teoría de conjuntos. Dada una colección de conjuntos, finita o infinita, este principio permitía elegir un elemento de cada uno de ellos, formando así un nuevo conjunto. A quienes no admitían el infinito actual, difícilmente podía gustarles este modo de elegir elementos como por arte de magia, sin atenerse a ninguna regla explícita.

En una serie de artículos publicados entre 1904 y 1927, David Hilbert fue precisando cada vez más los detalles de su estrategia para reemplazar todas las demostraciones de las matemáticas por pruebas realizadas mediante métodos finitarios y culminar su programa demostrando la consistencia de la aritmética de la forma más rigurosa y segura posible. Con lo que no contaba el líder de la escuela de Gotinga era con que un joven austríaco, que había comenzado a estudiar física en la Universidad de Viena, pero que pronto se había sentido más atraído por las matemáticas, mientras trataba de hacer avanzar el programa formalista iba a descubrir que el sueño de Hilbert era imposible, y lo que es peor, ¡lo iba a demostrar usando los métodos finitarios!

Capítulo 4

Los teoremas de Gödel

Si surgiesen conflictos de opiniones, dos filósofos no discutirían más que dos contables: bastaría con que cogiesen papel y lápiz, se sentaran y se dijeran el uno al otro: «Calculemos».

Gottfried Leibniz

Las calles de la ciudad de Königsberg lo habían visto todo. Como la atravesaban siete puentes, sus habitantes se habían preguntado desde que se construyeron si sería posible recorrerlos todos una sola vez volviendo al punto de partida. Nadie había sido capaz de hacerlo, pero tampoco de demostrar que era imposible, hasta que en 1735 el matemático suizo Leonhard Euler inventó la teoría de grafos para dar una respuesta negativa a la pregunta.

Cuarenta años después, Immanuel Kant se paseaba por esos mismos puentes intentando fijar los límites de competencia de la razón pura. Sólo este azar, al que se sumaría el nacimiento de David Hilbert, tuvo que haber bastado para que una sociedad en defensa de la filosofía empírica, que colaboraba con el Círculo de Viena, organizase en Königsberg la Conferencia sobre la Epistemología de las Ciencias Exactas entre el 5 y el 7 de septiembre de 1930.

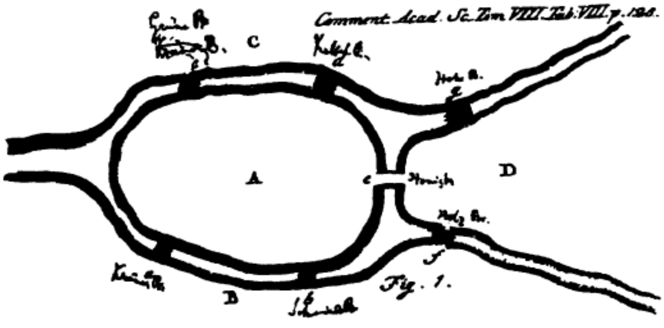


Gráfico de la solución original de Leonhard Euler al problema de los puentes de Königsberg.

El objetivo del encuentro era decidir hasta qué punto se había conseguido resolver, durante las primeras décadas del siglo, la crisis en los fundamentos de las matemáticas que había despertado Russell con su paradoja. Como conferenciantes plenarios se eligió a quienes más habían contribuido a desarrollar en los últimos tiempos las tres soluciones principales a la crisis: el logicismo, que sostenía que todas las matemáticas son reducibles a la lógica; el formalismo, cuyo gran éxito consistía en distinguir el lenguaje del metalenguaje, y el intuicionismo, que pretendía expulsar el infinito de las matemáticas. El resto del programa se reservó para que los participantes presentaran sus descubrimientos más recientes y para que charlasen distendidamente en los cafés de la ciudad, sin duda peores que los de Viena, pero agradables.

Al lógico austríaco Kurt Gödel lo habían invitado a exponer su tesis doctoral, que dejaba aún la puerta abierta a unas matemáticas todopoderosas. Sin embargo, en los meses trascurridos entre el inicio brillante de su carrera y el congreso de Königsberg, Gödel había avanzado en sus investigaciones hasta convencerse de que el sueño de los lógicos de su generación era imposible. Nada parecía indicarlo mientras pronunciaba su conferencia, pero en los últimos minutos de la mesa redonda que cerró el encuentro al día siguiente, por fin se atrevió a anunciar que «tenía

Fotografía de la Universidad de Königsberg hacia 1900, conocida popularmente como Albertina.

UN DIÁLOGO DE *LOS CRÍMENES DE OXFORD* (ÁLEX DE LA IGLESIA/JORGE GUERRICAECHEVARRÍA, 2008)

Sheldon: Oh, olvidaba que estoy hablando con el defensor de la lógica universal. Usted y la policía creen que se puede demostrar la verdad. A partir de unos axiomas y con un razonamiento válido se llega a una conclusión válida, ¿no es cierto?

Martin: Tan cierto como que hoy es miércoles.

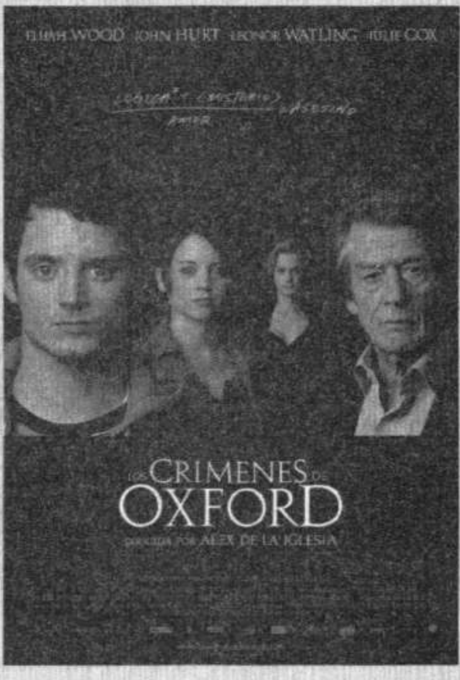
Sheldon: ¿Y si yo dijera «Todos los británicos son mentirosos»? ¿Verdadero, falso, o imposible de demostrar?

Martin: De acuerdo, hay algunos enunciados matemáticos que no se pueden afirmar o negar a partir de los axiomas: enunciados indeterminados.

Sheldon: Exacto. El teorema de incompletitud de Gödel. Incluso en su mundo de pureza matemática hay cosas que no se podrán demostrar.

Martin: Sí, lo sé, pero éste no es el caso.

Sheldon: Hay una grieta. Hay un abismo, ¿sabe?, entre lo verdadero y lo demostrable. Nunca sabremos si tenemos todos los datos acerca de un fenómeno, y la falta de un dato lo cambiaría todo.



ejemplos de proposiciones verdaderas por su contenido que no podían demostrarse a partir de los axiomas». Como el final de un cuento cuyo protagonista se cuelga del clavo que había aparecido en la primera página, las palabras de Gödel pillaron tan desprevenidos a los asistentes que apenas hubo discusión, y ni siquiera fueron recogidas en las actas.

No a todos los presentes se les escapó que aquel discreto joven de gafas circulares estaba a punto de cambiar los derroteros de la lógica con su comentario más bien incomprensible: entre ellos se encontraba John von Neumann que, gracias a su legendaria rapidez mental, vio de inmediato a qué podía referirse Gödel y le pidió más detalles tras la clausura de la conferencia. Von Neumann había estudiado con Hilbert en Gotinga, entre otras muchas universidades, y aunque había publicado varios artículos a la zaga del maestro, pronto empezó a dudar de que los métodos finitarios propuestos por el formalismo sirvieran para demostrar la consistencia de las matemáticas. En su juventud, Von Neumann había obtenido algunos resultados en esa dirección que lo animaron a trabajar sin pausa. Una noche soñó que había superado el último escollo, se despertó sobresaltado y continuó pensando en el problema hasta el día siguiente, pero a la hora de acostarse quedaban todavía cabos sueltos. Esa noche soñó de nuevo que había dado con la solución, pero al tratar de redactarla encontró un fallo en su argumento, y decidió al fin dedicarse a otros asuntos.

Además de sus contribuciones a la lógica, John von Neumann llevó a cabo importantes trabajos en mecánica cuántica.

Tras llegar a Königsberg como la gran estrella invitada, de pronto John von Neumann veía cómo un actor secundario le robaba el protagonismo dando a conocer lo que él podría haber soñado la tercera noche. De vuelta a casa, el antiguo colaborador de Hilbert descubrió que, si las investigaciones del austríaco eran correctas, entonces la consistencia de la aritmética no podría demostrarse dentro de la propia aritmética. Así se lo comunicó el 20 de noviembre de 1930, con tan mala suerte que tres días antes Gödel había enviado a la revista *Monatshefte für Mathematik und Physik* el manuscrito «Sobre proposiciones formalmente indecidibles en los *Principia mathematica* y sistemas afines I», donde aparecía también el nuevo resultado. En lugar de enfurecerlo, el incidente despertó la admiración de Von Neumann, de tal modo que cuando se publicó el artículo en la primavera de 1931, paró sus clases en Berlín para explicarlo, y veinte años más tarde seguiría recordando aquel momento como «un hito que siempre se divisará desde remotas distancias en el espacio y en el tiempo».

También David Hilbert se encontraba en Königsberg, pero no en la Conferencia sobre la Epistemología de las Ciencias Exactas, sino en un encuentro de la sociedad de científicos alemanes, que lo había invitado a pronunciar la conferencia «La lógica y la comprensión de la naturaleza», el día después del anuncio de Gödel. Aunque Hilbert y él no llegaron nunca a entablar conversación, se sabe que el lógico austriaco permaneció en Königsberg varios días después del encuentro, así que no es del todo improbable que estuviera entre el público que escuchó a Hilbert proclamar con más vehemencia que nunca que en matemáticas no existen problemas irresolubles: «No debemos creer en quienes hoy día, con aires filosóficos, profetizan el fin de la cultura y aceptan el *ignorabimus*[1]. Porque para nosotros no hay *ignorabimus,* y en mi opinión tampoco existe en ninguna de las ciencias naturales. Frente al necio *ignorabimus*, nuestro eslogan será: "Debemos saber, ¡sabremos!"». El eco de su voz rotunda aún resonaba cuando Hilbert supo que los sucesos de Königsberg ponían en peligro su programa.

Los teoremas de incompletitud

Antes del anuncio de Gödel en la conferencia, el estado del programa de Hilbert daba razones para la esperanza: el primer requisito, formalizar la matemática, pare-

1 Abreviación de la sentencia latina *ignoramus et ignorabimus*, es decir, «no sabemos y no sabremos nunca», que el fisiólogo alemán Emil du Bois-Reymond acuñó en 1872 para expresar su pesimismo sobre los límites del conocimiento científico.

cía haberse completado con éxito en los *Principia mathematica* de Russell y de Whitehead, y varios lógicos trataban de demostrar la consistencia de los sistemas formales clásicos, empezando por la aritmética. Aunque en la introducción de su tesis doctoral Gödel ya había sugerido la posibilidad de que existiesen «sentencias verdaderas que no pueden deducirse en el sistema en cuestión», su objetivo no era poner fin al sueño de Hilbert, sino probar la validez del programa. Sin embargo, el espíritu intelectual de la época apuntaba en otra dirección: como consecuencia de las investigaciones geométricas de Gauss, se había concluido que es imposible dibujar un mapa perfecto de la Tierra. Évariste Galois (1811-1832), por su parte, había demostrado que casi ninguna ecuación algebraica se puede resolver con métodos sencillos, y Werner Heisenberg (1901-1976) acababa de establecer un nuevo límite para la ciencia con su principio de incertidumbre, según el cual no es posible medir al mismo tiempo y con idéntica exactitud la posición y la velocidad de los electrones.

Con sus teoremas, Gödel pondría en conocimiento de todos las limitaciones intrínsecas del método axiomático: si en el primer capítulo explicamos que los atributos que hacen absolutamente irresistible un sistema formal son la consistencia (que no dé lugar a contradicciones), la recursividad (que los axiomas puedan reconocerse entre el resto de enunciados) y la completitud (que lo verdadero coincida con lo demostrable), el lógico austriaco mostraría cómo en el caso de la aritmética estas tres condiciones son incompatibles. De acuerdo con las investigaciones de Gödel, ninguna axiomatización recursiva y consistente de la aritmética puede ser completa, es decir, siempre existirán algunas propiedades verdaderas de los números que no podamos demostrar a partir de los axiomas. Éste es el contenido del *teorema de incompletitud* de Gödel, al que los expertos suelen referirse como primer teorema de Gödel, pues todavía le quedaron fuerzas para demostrar un segundo teorema, que afirma que el enunciado «La aritmética es consistente» es un ejemplo de estas proposiciones indecidibles. Ése era el resultado que Von Neumann había conseguido deducir tras la reunión de Königsberg.

Para demostrar el primer teorema de incompletitud, Gödel modificó la paradoja del mentiroso hasta convertirla en una sentencia indecidible que no reflejaba contradicción alguna. De hecho, uno de los encantos indudables del teorema es este modo de vivir peligrosamente, a un paso de las paradojas, pero sin caer en ellas. Recordará el lector que en el segundo capítulo vimos que una de las formulaciones de la antinomia de Epiménides era «Esta frase es falsa». En efecto, si la afirmación se supone verdadera, entonces ella misma dice que es falsa, mientras que si se considera falsa, por fuerza es verdadera. Ahora bien, ¿qué ocurriría si nos interesásemos por lo demostrable en lugar de por lo verdadero? Llamemos G (G de Gödel) a la proposición «Esta frase no es demostrable» y supongamos que nuestro sistema axiomático es consistente. Si G fuera falsa, como lo que dice G es «No soy demostrable», entonces G sería demostrable, pero en un sistema consistente ningún enunciado falso puede ser demostrable, ya que al instante se obtendría una contradicción. Si G no es falsa, entonces es verdadera, luego tenemos una frase verdadera que dice: «No soy demostrable». Por tanto, en el momento en el que suponemos que el sistema es consistente, podemos encontrar una sentencia verdadera pero indemostrable; dicho de otro modo, «consistente» implica «incompleto».

En el momento en el que suponemos que el sistema es consistente… Pero ¿qué sistema? Al lector benévolo que, tras hacerse esta pregunta al final del párrafo anterior, haya pensado que era culpa suya haberse perdido entre tanta autorreferencia y no saber de qué sistema estábamos hablando, con alegría le diremos que acaba de plantearse la cuestión crucial, para la que no había respuesta antes de Gödel. Nuestro razonamiento muestra que la afirmación «No soy demostrable» tiene que ser verdadera, pero no se trata de un enunciado matemático, como nos gustaría, sino metamatemático, pues no se refiere a los objetos de estudio de ninguna teoría, sino a las propias teorías. La genialidad de Gödel consistió en traducir algunas expresiones del metalenguaje al lenguaje de la aritmética, gracias a un sistema de codificación basado en los números primos. Tras esta *gödelización* de la metamatemática, los números naturales llevaban una doble vida: por un lado eran ellos mismos, los de siempre, pero, por otro, representaban el papel de alguna fórmula, lo cual permitía que un enunciado como «No soy demostrable», que a priori sólo tenía sentido en el metalenguaje, se convirtiese en una relación numérica.

A la espera de una explicación más detallada sobre el código de Gödel, nos contentaremos con saber que, utilizándolo, se podía encontrar en la aritmética un enunciado equivalente a «No soy demostrable». Si la aritmética tenía un conjunto de axiomas S recursivo y consistente, entonces existía una fórmula G_S verdadera pero indemostrable (hemos usado aquí el subíndice S para indicar que la sentencia construida depende de los axiomas, de modo que si los cambiásemos obtendríamos otra distinta). Frente a la soñada omnipresencia de los lógicos, Gödel planteaba una elección entre dos senderos que se bifurcan: el de la completitud y el de la consistencia-recursividad. Y lo que es peor, la aritmética no sólo era incompleta, sino también incompletable. Cuando al inicio del libro presentamos el ejemplo del inspector que acababa de incorporarse a la comisaría, alguien podría habernos repro-

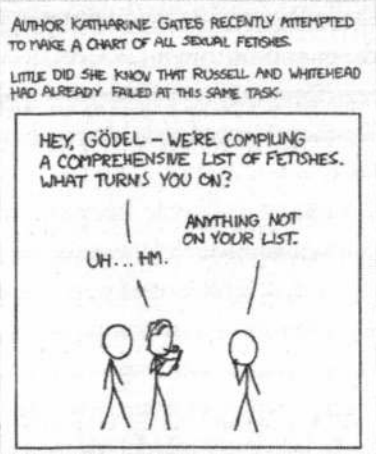
«CUALQUIER COSA QUE NO ESTÉ EN TU LISTA»

Randall Munroe (n.1984) trabajaba para la NASA hasta que en el año 2005 descubrió su gran talento para hacer reír a los demás con humor científico. Comenzó entonces a dibujar la serie *xkcd*, «un webcómic de amor, sarcasmo, matemáticas y lenguaje». Se trata de unas viñetas de trazos muy simples que a menudo incorporan referencias a resultados de la física, las matemáticas o la informática. Kurt Gödel ha tenido varias apariciones estelares, pero ninguna tan genial como en la viñeta titulada *Fetiches*, reproducida abajo. En ella se ven tres personajes, y sobre la viñeta aparece la leyenda:

> «La escritora Katharine Gates intentó recientemente hacer una lista de todos los fetiches sexuales. Ignoraba por completo que Russell y Whitehead ya habían fracasado en la labor».

Uno de los personajes dice:

—Hey, Gödel, estamos haciendo una lista exhaustiva de fetiches. ¿Qué es lo que te pone?
—Cualquier cosa que no esté en tu lista —contesta Gödel.



Randall Munroe durante una charla en el Massachusetts Institute of Technology (fuente: Petehurne).

chado que sus compañeros hubieran sabido si estaba casado o no tan sólo con un poco más de conversación. Hay sistemas incompletos que dejan de serlo al añadirles un puñado de axiomas. Pero éste no es el caso de la aritmética: además de exhibir la sentencia indecidible G_S, Gödel demostró que no sirve de nada incorporarla como axioma, pues al aplicar el método a $T = S + G_S$, que vuelve a ser un conjunto de axiomas recursivo y consistente, se obtiene otra proposición verdadera pero indemostrable G_T. Cortarle a la hidra una de sus infinitas cabezas nunca nos salvará de la incompletitud.

Hemos prometido que íbamos a explicar cómo es posible traducir a la aritmética la proposición indecidible «No soy demostrable», pero antes de hacerlo avanzaremos algunos pasos hacia el segundo teorema de incompletitud. En el primer capítulo dijimos que en los sistemas axiomáticos inconsistentes cualquier proposición es un teorema. Por tanto, la existencia de al menos una fórmula que no sea un teorema es un criterio inconfundible para saber cuándo una teoría es consistente. Si somos capaces de encontrar una proposición no demostrable, automáticamente nos libraremos de las contradicciones. ¡Con una sola basta! Así que, ¿por qué elegir una muy complicada teniendo a mano la más simple: $0 = 1$? Al empezar el libro, indicamos cómo se deducía el teorema «Cero es distinto de uno» de los axiomas de Peano. No cuesta mucho esfuerzo convencer al lector de que, aunque elijamos otros axiomas, cualquier teoría sensata que hable de los números distinguirá el cero del uno. En resumidas cuentas, decir que la aritmética es consistente es lo mismo que decir que la fórmula $0 = 1$ no es demostrable.

Nos encontramos de nuevo ante un enunciado del metalenguaje, pero en virtud de la *gödelización* podemos transformarlo en una fórmula sobre los números, que llamaremos Con_S (*Con* de consistencia y S por el sistema de axiomas). Con esta traducción, lo que dice el primer teorema de incompletitud es que Con_S implica G_S, ya que si la aritmética es consistente (es decir, si Con_S es verdadera), entonces G_S es verdadera. Llegado este punto, conviene recordar cómo funciona una de las reglas de deducción más potentes, el *modus ponens*, que permite deducir de las demostraciones de la implicación lógica «Si A entonces B» y del enunciado A una demostración de B. Supongamos por un momento que la consistencia de la aritmética pudiese probarse dentro de la aritmética. Entonces Con_S sería demostrable y, al ponerla junto a la demostración del primer teorema de incompletitud, $Con_S \rightarrow G_S$, deduciríamos por *modus ponens* una demostración de G_S. Pero esto es absurdo, pues ¡G_S es indemostrable! La única conclusión posible es que para probar la consistencia de la aritmética es preciso salir de la aritmética, y eso es lo que dice el segundo teorema de incompletitud, que el propio Gödel consideraba un «corolario sorprendente» de sus investigaciones.

De acuerdo con el programa de Hilbert, para demostrar la consistencia de las matemáticas había que comenzar por la aritmética. Sin embargo, el segundo teorema de Gödel señalaba que, si existía una prueba de la consistencia de la aritmética, necesariamente debería usar técnicas mucho más complicadas que los métodos finitarios defendidos por los formalistas. Seguramente el lector se habrá percatado de que el título de Gödel «Sobre proposiciones formalmente indecidibles en los *Principia mathematica* y sistemas afines I» anunciaba una segunda parte; la razón es que el artículo sólo contiene un esbozo del segundo teorema de incompletitud. Aunque todo lo indicado allí es correcto, Gödel nunca terminó de escribirla, lo cual encaja con la imagen de «explorador que deja los detalles para los demás» que presentan sus biógrafos. De hecho, fueron David Hilbert y su colega Paul Bernays (1888-1977), al que Gödel había explicado todas las sutilezas de la demostración durante un viaje al otro lado del Atlántico, quienes publicaron la primera prueba completa del segundo teorema de incompletitud en 1939. No deja de ser un síntoma de la salud moral de la ciencia del momento que fuese el propio Hilbert quien completara los detalles del teorema que echaba por tierra su trabajo de más de veinticinco años.

Aún así, la recepción de los teoremas de incompletitud no fue todo lo buena que éstos se merecían. Algunos matemáticos creyeron que la proposición indecidible «No soy demostrable» era una mera curiosidad que nunca afectaría a su trabajo, y hubo quienes, al no comprender el matiz que separaba lo verdadero de lo demostrable, acusaron a Gödel de haber reproducido la paradoja del mentiroso. Entre ellos se encontraba el sexagenario Ernst Zermelo, a pesar de que él sabía mejor que nadie lo duro que resulta luchar por una idea, pues su axioma de elección le había valido innumerables críticas. En general, la comunidad matemática no estaba preparada en aquella época para entender un trabajo que incorporaba técnicas muy novedosas dentro de un dominio que había sido siempre minoritario. Cuánta razón tenía Thomas Kuhn al señalar en su estudio sobre *La estructura de las revoluciones científicas* que «en la ciencia lo nuevo sólo surge con dificultad, puesto de manifiesto por la resistencia, sobre el fondo que proporciona lo esperado». Por suerte, la traducción al inglés del artículo de Gödel y una exposición divulgativa de sus teoremas propiciaron que, a partir de la década de los sesenta, los teoremas de incompletitud comenzaran a reconocerse como el avance más importante de la lógica desde los tiempos de Aristóteles.

Kurt Gödel fotografiado en el Instituto de Estudios Avanzados de Princenton, Nueva Jersey.

CIUDADANO GÖDEL

Tras huir de la Alemania nazi, Kurt Gödel se instalaría definitivamente en la Universidad de Princeton, en 1940. Cuando siete años después obtuvo la nacionalidad estadounidense, tuvo lugar una de las anécdotas más conocidas del personaje. Como todos los solicitantes, Gödel debía dar cuenta de su conocimiento de la legislación estadounidense en un examen sobre la Constitución. En la práctica, la prueba no era más que un trámite, pero él quiso preparársela a conciencia y, mientras lo hacía, creyó descubrir algunas contradicciones lógicas:

—Usted tenía hasta ahora la nacionalidad alemana.
—Perdone, señor, austriaca —corrigió Gödel.
—Ah, ya, el maldito dictador. Afortunadamente, eso no es posible en América.
—Al contrario —interrumpió Gödel—. ¡Yo sé cómo!

Pero antes de dejarle hablar, el juez, al que Albert Einstein ya había advertido de que Gödel no era un candidato como los demás, tomó las riendas de la situación y condujo el examen hacia preguntas más rutinarias: «Tampoco es necesario meterse en honduras». Más o menos por aquella misma época, algunos lógicos habían empezado a sentar las bases de una teoría, la lógica deóntica, cuyo objetivo es precisamente evitar que surjan contradicciones durante la incorporación de nuevas leyes a los códigos.

La *gödelización*

El 21 de junio de 1851, Adolf Anderssen, el mejor ajedrecista del momento, se reunió en uno de los restaurantes más antiguos de Londres con Lionel Kieseritzky, que daba clases de ajedrez en un club de París, para jugar una partida que en los siguientes años empezaría a conocerse como «la inmortal». Impresionado por la estrategia de Anderssen, que había sacrificado su alfil, la reina y las dos torres para hacer jaque mate, Kieseritzky quiso comunicar de inmediato la descripción del juego a su club parisino. Pero en lugar de comenzar con «Blancas: el quinto peón por la izquierda se mueve dos casillas hacia delante. Negras: el peón de la misma columna se coloca frente a él. Blancas: el tercer peón por la derecha avanza dos casillas. Negras: el peón del primer movimiento captura la última ficha...», los primeros símbolos del mensaje eran algo parecido a «e4 e5 / f4 exf4...». Toda la información de la partida apenas ocupaba tres renglones, ¡y más le valía a Kieseritzky!, porque de haber seguido el primer método, pagar el telegrama le habría costado mucho esfuerzo en el Café de la Régence, donde jugaba al ajedrez por cinco francos la hora.

Los jugadores de ajedrez habían encontrado un modo extremadamente conciso de condensar toda la información de sus jugadas. Para ello se servían en primer lugar de un método de traducción antiguo, la geometría analítica de Descartes, gracias a la cual cada casilla del tablero podía identificarse con dos coordenadas: una letra de la *a* a la *h* para representar la columna, y un número del 1 al 8 para indicar la fila. Salvo los peones, que iban sin identificar, cada ficha del juego se le hacía corresponder con la inicial del nombre: A de alfil, C de caballo, D de dama, R de rey y T de torre. Se añadían luego otros símbolos, como x de captura, + de jaque, o ++ de jaque mate. De acuerdo con esta convención algebraica, la secuencia «e4 e5 / f4 exf4» decía lo mismo que «Las blancas mueven un peón a la casilla e4, y las negras responden moviendo otro peón a la casilla e5. Después las blancas mueven un peón a la casilla f4, que las negras se comen con el peón que estaba en la casilla e5».

Con este ejemplo hemos querido hacer hincapié en la utilidad que los sistemas de codificación tienen en muchos ámbitos fuera y dentro de las matemáticas, donde transforman las más complejas expresiones en símbolos fáciles de manejar. En el capítulo anterior vimos cómo las propiedades de los números naturales, escritas en el lenguaje cotidiano, podían traducirse al simbolismo de los *Principia mathematica*. Por ejemplo, el axioma «Cero no es el sucesor de ningún número» se convertía en la fórmula $\neg\exists x(sx = 0)$ a través de este sistema. Sin embargo, Gödel necesitaba ir más allá: para demostrar su teorema de incompletitud no le bastaba con reducir a fórmulas la aritmética, sino que debía ser capaz de condensar cualquier fórmula, ¡incluso cualquier demostración!, en un solo número. Entonces el lógico recordó que en los seminarios de historia de la filosofía a los que había asistido durante sus años de aprendizaje en la Universidad de Viena, el profesor Theodor Gomperz había comentado la edición de Louis Couturat de los manuscritos inéditos de Leibniz publicada en 1903.

Como sus predecesores más geniales, Leibniz dedicó grandes esfuerzos a tratar de poner fin a la *confusio linguarum* con la que Dios había castigado la soberbia humana de querer construir una torre que alcanzase el cielo. Para ello había imaginado una lengua universal que reducía todos los pensamientos humanos, con independencia del idioma en que se formulasen, a un catálogo de ideas primitivas, a las que se asignaba un número primo. A partir de este inventario, igual que se forman los números compuestos, podían calcularse los caracteres de las ideas derivadas, de modo que siempre fuera posible «extraer las nociones primitivas que las componen». Si a los conceptos de agua y de quietud les correspondían, pongamos, los números 3 y 5, entonces la idea compuesta de lago podría reflejarse por medio del producto $3 \cdot 5$. Recíprocamente, si nos dijeran que el concepto de lago admite el *carácter* 15, descompondríamos 15 en factores primos y, buscando en la enciclopedia las ideas primitivas asociadas a los números 3 y 5, concluiríamos que un lago no es más que agua quieta. Así, para saber si una afirmación del tipo «*A* es *B*» es verdadera, bastaría con comprobar que el carácter de *B* divide al carácter de *A*, y «si surgiesen conflictos de opiniones, dos filósofos no discutirían más que dos contables». El ambicioso programa de Leibniz, descubierto dos siglos después de la muerte del filósofo, nunca ha llegado a realizarse, pero sugirió a Gödel cómo traducir el metalenguaje a la aritmética.

Recordemos que los números primos son aquellos que sólo son divisibles por 1 y por sí mismos: por ejemplo, 5 es primo, porque ni 2, ni 3, ni 4 lo dividen; pero 6 no lo es, ya que el resultado al dividir por 2 es 3. Los primeros números primos son 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31... y por un argumento de reducción al absurdo, como los que tanto odiaban los intuicionistas, se puede demostrar que la lista continúa indefinidamente. La mayor parte de los esfuerzos de la física de la segunda mitad del siglo XX se han concentrado en identificar las partículas elementales de la materia, aquellas que no pueden dividirse en otras más sencillas. Pues bien, los matemáticos saben desde tiempos de Euclides que las partículas elementales de la aritmética son los números primos. En efecto, al elegir un natural n cualquiera hay dos posibilidades: o bien n es primo, y entonces ya hemos terminado, o bien existe algún número distinto de 1 y de n que lo divide. Si, por ejemplo, n vale 23, estaríamos en el primer caso, pero si n es igual a 30, entonces podemos dividir por 2.

Supongamos, por tanto, que el número de partida no es primo; entonces podremos descomponerlo como un producto: $n = a \cdot b$ (en nuestro caso, $30 = 2 \cdot 15$). Hemos obtenido así dos números a los que aplicar de nuevo el proceso: si los dos son primos, entonces hemos terminado, pero si alguno de ellos no lo es, volvemos a escribirlo como producto de sus factores. Siguiendo con el ejemplo, 2 es primo, luego no hay nada que hacer, pero 15 aún se descompone como $15 = 3 \cdot 5$, de modo que $30 = 2 \cdot 3 \cdot 5$. Como 2, 3 y 5 son números primos, el juego ha terminado. En general, o bien encontramos un factor primo, o bien los términos que aparecen son cada vez más pequeños, lo cual nos garantiza que el proceso se detendrá antes o después. Hemos demostrado así el ***teorema fundamental de la aritmética***, que dice que cualquier número se descompone como un producto de factores primos que tal vez se repitan. Por ejemplo: $77.220 = 2 \cdot 2 \cdot 3 \cdot 3 \cdot 3 \cdot 5 \cdot 11 \cdot 13$, y en ese caso se emplea la abreviatura $77.220 = 2^2 \cdot 3^3 \cdot 5 \cdot 11 \cdot 13$, donde los exponentes indican el número de veces que aparece cada primo.

El teorema fundamental de la aritmética dice algo más fuerte: no sólo existe una descomposición de este tipo para cualquier número natural, sino que, además, es única salvo por el orden de los factores; es decir, quizá podamos escribir 77.220 de otra manera, por ejemplo, como $77.220 = 5 \cdot 2^2 \cdot 11 \cdot 3^3 \cdot 13$, pero en la nueva descomposición intervendrán los mismos números primos elevados a los mismos exponentes.

En el capítulo anterior vimos que el alfabeto de la aritmética está compuesto por los ocho símbolos: 0 (el número cero), s (la función sucesor), $\neg$ (la negación), $\vee$ (la conjunción «o»), $\exists$ (existencia), $=$ (igual) y los paréntesis de apertura y de cierre, y también disponemos de variables x, y, z que representan los números que vamos a estudiar. Como primera fase de la codificación, Gödel propone hacer corresponder a cada símbolo un número del 1 al 8, y a las variables x, y, z, los tres primeros primos mayores que 8, tal y como se indica en la siguiente tabla:

0	s	$\neg$	$\vee$	$\exists$	$=$	(	)	x	y	z
1	2	3	4	5	6	7	8	11	13	17

Una vez que se ha asignado un número a las «ideas primitivas» de la aritmética, codificar una fórmula es muy sencillo: en primer lugar, se cuenta el número de símbolos que intervienen en ella (con repeticiones) y se elige esa misma cantidad de números primos; el tamaño de la fórmula no importa, porque hay infinitos primos. A continuación se eleva cada primo al exponente correspondiente al símbolo, de acuerdo con el diccionario anterior, y se multiplican todos entre sí. Veámoslo con un ejemplo, que vale más que mil explicaciones.

El tercer axioma de Peano afirma que «Cero no es el sucesor de ningún número», que expresamos como $\neg \exists x (sx = 0)$. Si seguimos al pie de la letra las instrucciones de la *gödelización*, lo primero que hay que hacer para transformarlo en un número es contar los símbolos que aparecen en la fórmula; son nueve: $\neg$, $\exists$, x, (, s, x, $=$, 0 y). Elegimos, por tanto, los nueve primeros primos, a saber: 2, 3, 5, 7, 11, 13, 17, 19 y 23. De acuerdo con el diccionario, la negación $\neg$ va asociada al número 3, luego tenemos que elevar el primo 2 a la potencia 3, es decir: 2^3. Del mismo modo, el cuantificador de existencia $\exists$ queda reflejado en el número 5, de manera que habrá que elevar el primo 3 a la potencia 5, o sea: 3^5. Repitiendo el proceso, se obtienen 5^{11}, 7^7, 11^2, 13^{11}, 17^6, 19^1 y 23^8, y al multiplicarlos todos, la fórmula se transforma en:

$$2^3 \cdot 3^5 \cdot 5^{11} \cdot 7^7 \cdot 11^2 \cdot 13^{11} \cdot 17^6 \cdot 19^1 \cdot 23^8.$$

El método que acabamos de describir permite codificar cada fórmula en un número, que llamaremos número de Gödel, pero nada nos impide hacer lo mismo con las demostraciones. Recordemos que una demostración no es más que una sucesión finita formada por, digamos, n fórmulas, luego es posible codificar primero cada una de las fórmulas, elegir después n números primos, elevarlos al número de Gödel de cada una de ellas y hacer luego el producto. De esta manera, cada demostración de la aritmética queda reducida a un número.

El punto crucial es que la *gödelización* es un proceso reversible. Quienes estén algo familiarizados con la química sabrán que uno de los asuntos de mayor interés de esta ciencia consiste en saber en qué reacciones se puede dar marcha atrás para volver al estado de partida. Por ejemplo, al quemar un combustible, éste se convierte en vapor de agua y en dióxido de carbono, el famoso gas de efecto invernadero. Sin embargo, es imposible recuperar el combustible original a partir de dichos gases; de lo contrario ¡se habrían solucionado los problemas energéticos del planeta! Otras reacciones químicas, en cambio, son reversibles, como la que se obtiene al pasar vapor de agua sobre una lámina caliente de hierro. Partiendo de la piedra imán y del hidrógeno resultantes de esta reacción, se pueden recuperar el hierro y el vapor de agua.

Éste es el escenario que queremos imitar en la aritmética, pues los números nunca podrían llevar la doble vida que Gödel estaba dispuesto a inventar para ellos si, al representar uno de los papeles, olvidasen para siempre el otro. Gracias al teorema fundamental de la aritmética, en el laboratorio de la *gödelización* todas las reacciones son reversibles. Vamos a ver por qué: supongamos que tenemos el número

304.496.379.203.017.490.604.020.678.113.081.132.612.291.772.080.917.708.
404.389.616.093.394.253.015.558.500.327.468.465.234.375.000,

que nos hemos tomado la molestia de escribir para que el lector se haga una idea de cómo son los números de Gödel más pequeños... El teorema fundamental de la aritmética nos asegura que es posible descomponerlo en factores primos. Si le da pereza hacerlo a mano, lo cual es más que razonable, puede visitar la página web http://www.wolframalpha.com y escribir el número (sin los puntos que facilitan la lectura) precedido de la palabra «factor», en el recuadro principal. Con cantidades aún mayores, el ordenador podría tardar mucho tiempo, pero lo que importa aquí es que el teorema fundamental de la aritmética nos garantiza que tal factorización existe siempre, y que, además, es única. Por suerte, Internet aún considera que el número 304.496...375.000 es pequeño, así que en menos de un segundo nos ha devuelto la descomposición:

$$2^{3} \cdot 3^{5} \cdot 5^{11} \cdot 7^{3} \cdot 11^{5} \cdot 13^{13} \cdot 17^{7} \cdot 19^{13} \cdot 23^{6} \cdot 29^{2} \cdot 31^{11} \cdot 37^{8}.$$

Todo lo que queda por hacer es fijarse en los exponentes y recuperar los símbolos a través del diccionario. Así obtenemos la fórmula $\neg \exists x \neg \exists y (y = sx)$, que dice que no existe ningún número x tal que no exista ningún número y con la propiedad de que y sea el sucesor de x. Reformulando un poco la proposición, mis lectores se convencerán de que podemos escribirla como «Cada número natural tiene un sucesor», que es el segundo axioma de Peano.

Por supuesto, no todos los números naturales son el número de Gödel de alguna fórmula, pero incluso si nos diesen uno que no correspondiera a ninguna expresión de la aritmética, sabríamos detectarlo enseguida. Por ejemplo, $15 = 3 \cdot 5$ no es el número de Gödel de ninguna fórmula, ya que la *gödelización* obliga a que aparezcan los primeros primos, sin ningún salto, y en la descomposición de 15 no interviene el 2. El número $1.536 = 2^{9} \cdot 3$ tampoco corresponde a ninguna expresión de la aritmética, porque aunque en este caso los primos aparecen en orden, ninguno de los símbolos del alfabeto corresponde al exponente 9.

Recapitulando: el sistema de codificación que hemos descrito en estas páginas permite asociar a cada fórmula (también a cada demostración) de la aritmética un número que codifica toda su estructura. Además, esta reacción matemática es reversible en el sentido de que, factorizando cualquier número natural N, podemos decidir:

1. Si N es el número de Gödel de alguna fórmula o no.
2. En el caso de que lo sea, de qué fórmula.

GÖDEL EN LA LITERATURA

En la novela *Las nuevas confesiones*, de William Boyd, el protagonista acaba de rodar la obra maestra del cine mudo, pero el lanzamiento pasa sin pena ni gloria porque coincide con los primeros cortometrajes sonoros. Sólo Kurt Gödel, en una aparición fugaz, sabe reconocer el gran talento del director de cine.

En otra novela publicada diez años más tarde, *En busca de Klingsor*, del escritor mexicano Jorge Volpi, la novia del personaje principal, un físico llamado Francis Bacon, irrumpe en un seminario de Gödel en el Instituto de Estudios Avanzados, y comienza a gritarle porque le está siendo infiel. Cuando la acción se ha desplazado a las últimas filas, «el profesor Gödel anuncia que no puede continuar con la clase y comienza a llorar, irrefrenablemente». Su gran conflicto —pone el autor en boca de Von Neumann— no son las proposiciones formalmente indecidibles, «sino su amor desgarrado y turbulento por una prostituta: su propia esposa». Mientras el retrato de *Las nuevas confesiones* es verosímil, la escena descrita por Volpi es tan cruel como disparatada.

El escritor William Boyd ha incluido a Kurt Gödel en su novela Las nuevas confesiones.

La prueba de los teoremas de incompletitud

Aunque nos hayamos detenido un buen rato en el genial método de codificación que la lectura de Leibniz inspiró a Gödel, no deberíamos olvidar en ningún momento que se trata sólo de una herramienta para alcanzar nuestro objetivo: probar que en cualquier axiomatización recursiva y consistente de la aritmética existen proposiciones verdaderas pero indemostrables. Al principio del capítulo indicamos cuál era el esquema de la demostración: había que sustituir el concepto de verdad por el de demostrabilidad en la paradoja del mentiroso para obtener la afirmación que dice de sí misma «No soy demostrable». Si no admitimos contradicciones, la frase debe ser por fuerza verdadera, luego indemostrable. La dificultad mayor, como señalé entonces, consistía en encontrar el equivalente aritmético de esta proposición metalingüística, que no habla de los números, sino de las teorías matemáticas. Pues bien, ahora tenemos a nuestra disposición todos los métodos para traducirla. En lo que sigue trataremos de explicar los pasos más importantes de la demostración de Gödel de la forma más sencilla posible.

El juego consiste en traducir a la aritmética la frase «No soy demostrable», de modo que la primera pregunta que hay que plantearse es: ¿Qué quiere decir que una proposición sea demostrable en el sistema axiomático de la aritmética? Significa que exista una demostración que termine con nuestro enunciado, es decir, una cadena finita de fórmulas en la que cada una de ellas o bien sea un axioma, o bien se deduzca de las anteriores a través de las reglas de inferencia permitidas. Para saber si una cierta sucesión de fórmulas, que llamaremos Z, demuestra el enunciado X, debemos comprobar que Z está construida según la norma anterior, y que su última fórmula es precisamente X. La idea clave es asociar, mediante el proceso de *gödelización*, a X y a Z sus números de Gödel, que representaremos con las letras minúsculas x y z. Lo que más nos gustaría es disponer de un mecanismo D (D de demostración) que tomase los números naturales x y z, y al cabo de un tiempo contestara si la sucesión de fórmulas correspondiente al número z es o no una demostración de la fórmula de número de Gödel x. Por tanto, la sentencia $D(x, z)$ sería verdadera si Z demostrase la fórmula X, y falsa en caso contrario.

Por poner un ejemplo muy elemental, recordemos que el número de Gödel del segundo axioma de Peano es $2^{3} \cdot 3^{5} \cdot 5^{11} \cdot 7^{3} \cdot 11^{5} \cdot 13^{13} \cdot 17^{7} \cdot 19^{13} \cdot 23^{6} \cdot 29^{2} \cdot 31^{11} \cdot 37^{8}$. Como los axiomas se caracterizan por ser ellos mismos su propia demostración, si en $D(x, z)$ sustituimos los valores de x y z por este número tan agradable, el resultado es verdadero, pues la sucesión de fórmulas de número de Gödel z, compuesta en este caso únicamente por el segundo axioma de Peano, es una demostración de la fórmula *gödelizada* por x; ¡de nuevo el segundo axioma de Peano! Sin embargo, si introdujésemos como valor de z el número $2^{3} \cdot 3^{5} \cdot 5^{11} \cdot 7^{7} \cdot 11^{2} \cdot 13^{11} \cdot 17^{6} \cdot 19^{1} \cdot 23^{8}$, el mecanismo D($x$, z) nos diría «falso», ya que la fórmula correspondiente no es una demostración del segundo axioma de Peano. El hecho de que la fórmula de número de Gödel x sea demostrable significa que existe un número z tal que la sucesión de fórmulas correspondiente a z es una demostración de la fórmula asociada a x, o dicho de otro modo, un z tal que el enunciado $D(x, z)$ es verdadero. Como consecuencia, la fórmula $\exists z\, D(x, z)$, que abreviaremos Dem(x) (Dem de demostrable), afirma que la fórmula de número de Gödel x es demostrable. En resumen, si existiera D, gracias a la *gödelización*, todas las sutilezas de la demostrabilidad podrían reducirse a una simple relación entre los números naturales x y z. ¿Y cuál es la teoría que se ocupa de estas relaciones? ¡La aritmética!

Como el lector ya habrá imaginado, lo más laborioso del artículo de Gödel consistía en demostrar que, en efecto, existe un mecanismo con estas propiedades. Para hacerlo, el lógico austriaco necesitó cuarenta y seis etapas, de las cuales daremos sólo una indicación. Supongamos que recibimos el número natural z que codifica alguna sucesión de fórmulas. Gracias al teorema fundamental de la aritmética, podemos descomponer z en sus factores primos:

$$z = p_1^{k1} \cdot p_2^{k2} \cdot p_3^{k3} \cdot \ldots \cdot p_n^{kn}.$$

No es nuestra intención despistar a los lectores con esta notación, que por fuerza es un poco barroca. Lo único que hemos hecho hasta ahora ha sido descomponer el número z en una serie de factores primos que aparecen elevados a unos exponentes. Como z codifica una sucesión de fórmulas, cada uno de los exponentes será el número de Gödel de una de ellas. Este proceso nos permite identificar la *gödelización* de cada una de las fórmulas de la lista, que hemos llamado $k_1, k_2, k_3\ldots$ hasta k_n.

Repitamos todavía una vez más la musiquilla que viene sonando a lo largo de todo el libro: una demostración es una sucesión de fórmulas en la que cada una de ellas o bien es un axioma o bien se obtiene de las anteriores mediante las reglas de inferencia permitidas. Por tanto, lo que hay que comprobar es lo siguiente:

– Primer paso: La sucesión de fórmulas de números de Gödel $k_1, k_2\ldots k_n$ tiene la estructura de una demostración, es decir, el enunciado correspondiente a cada uno de estos números, o bien es un axioma, o bien se deduce de los anteriores mediante una de las reglas de inferencia permitidas.
– Segundo paso: La última fórmula de la sucesión es la que se quiere demostrar.

Empecemos por esta última etapa, que es la más sencilla: nos han dado una fórmula de número de Gödel x, y queremos saber si la cadena de enunciados termina con esa fórmula, el requisito más ingenuo para que se trate, en efecto, de una demostración. Ahora bien, los cálculos anteriores nos han permitido identificar los números de Gödel de cada una de las fórmulas de la lista, y el que corresponde a la última no es otro que k_n, así que ¡basta con ver si los números x y k_n son iguales! Nadie dirá que saber si dos números son o no son iguales no es una tarea simple.

Vamos ahora con la primera etapa de esta carrera de obstáculos consistente en examinar las fórmulas de números de Gödel k_1, k_2... hasta k_n, para ver si se comportan como deberían. Es aquí donde resulta imprescindible la hipótesis de que el sistema de axiomas de la aritmética sea recursivo, algo que hasta ahora parecía más bien caprichoso. Recordemos que un conjunto de axiomas S es recursivo cuando se puede comprobar, en una cantidad finita de pasos, si una proposición es o no es un axioma. Por tanto, tenemos a nuestra disposición una fórmula $A(x)$ (A de axioma) que lee el número x y decide si la proposición correspondiente es o no un axioma. Así las cosas, basta con que calculemos $A(k_1)$, $A(k_2)$... hasta $A(k_n)$, que indican cuáles de los enunciados del candidato a demostración son axiomas. La primera fórmula, correspondiente al número de Gödel k_1, tiene que ser por fuerza un axioma, ya que no hay nada antes a partir de lo cual pueda deducirse. Por tanto, si por casualidad $A(k_1)$ fuese falso, ya habríamos terminado: z no es el número de Gödel de una demostración. Supongamos, sin embargo, que todo funciona bien por el momento.

Entre las siguientes fórmulas, reflejadas en los números $k_2, k_3...k_n$, algunas serán axiomas y otras no. Para las que no lo sean, habrá que comprobar que se deducen de las anteriores usando las reglas de inferencia permitidas. En su minucioso trabajo, Gödel demuestra que, por cada regla de deducción, existe una fórmula I (I de inferencia) que toma los primeros s números k_1, k_2... hasta k_s y responden «verdadero» si la fórmula de número de Gödel k_s se deduce de las fórmulas de números de Gödel k_1, k_2... hasta k_{s-1} (el inmediatamente anterior) aplicando la regla de deducción correspondiente. Por ejemplo, $I(k_1, k_2, k_3, k_4)$ será verdadero si la cuarta fórmula de la cadena se deduce de las tres precedentes aplicando la regla de inferencia que se ha codificado a través de la fórmula I. Así, a las fórmulas que no sean axiomas podemos aplicarles este proceso, y si para cada una de ellas, al menos una de las respuestas de las diferentes reglas de deducción es «verdadero», entonces la primera etapa ha concluido con éxito, y z es el número de Gödel de una demostración. Como es muy fácil perderse en los detalles técnicos, puntualizaremos qué es lo importante: lo que hay que retener es que hemos demostrado que existe un proceso $D(x, z)$ que decide si la sucesión de fórmulas representadas por z es una demostración del enunciado de número de Gödel x. Para hacerlo, basta con traducir a relaciones entre números las normas que debe cumplir una demostración, que venimos repitiendo como si fueran un estribillo.

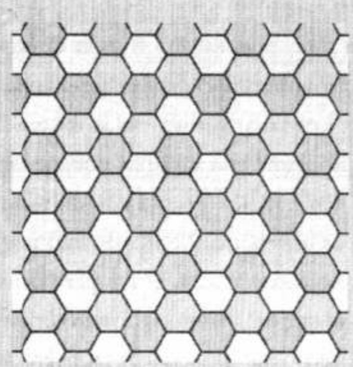
Muy bien: ya hemos construido dentro de la aritmética el enunciado $\mathrm{Dem}(x)$, que dice «La fórmula de número de Gödel x es demostrable». Negándola se obtiene $\neg\mathrm{Dem}(x)$, que no dice otra cosa que «La fórmula de número de Gödel x *no* es demostrable». Hasta aquí no hay ningún misterio, pero nos vamos acercando poco a poco al gran salto mortal. Antes de la última acrobacia, es preciso recordar que el enunciado «La aritmética es consistente», que intervenía en el segundo teorema de incompletitud, equivale a afirmar que «La fórmula $0=1$ no es demostrable». Recordando que uno es el sucesor de cero, es decir, $1=\mathrm{s}0$, invitamos al lector a comprobar que el número de Gödel de la fórmula $0=1$ vale 255.150. Por tanto, la proposición $\neg\mathrm{Dem}(255.150)$, traducida al lenguaje de la aritmética, afirma que «La fórmula de número de Gödel 255.150 no es demostrable», es decir, «La fórmula $0=1$ no es demostrable», que es lo mismo que «La aritmética es consistente». El enunciado $\neg\mathrm{Dem}(x)$ mata dos pájaros de un tiro.

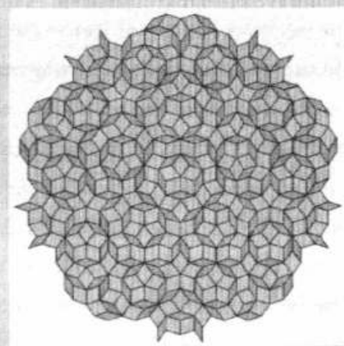
Lo importante de la expresión $\neg\mathrm{Dem}(x)$ es que ya no se trata de una afirmación en el lenguaje cotidiano, sino de una fórmula de la aritmética, en la que sólo intervienen los símbolos $0, \mathrm{s}, \neg, \vee, \exists, =, (,)$ y algunas variables. Las letras «Dem» son sólo un modo de abreviarla, porque su escritura es extremadamente compleja y ocuparía muchas páginas; pero si quisiéramos hacerlo, podríamos escribirla usando únicamente los caracteres del alfabeto de la aritmética. ¡Para eso hemos trabajado tanto! No me cabe duda de que mis lectores ya saben lo que hay que hacer cada vez que encontramos una fórmula escrita así: ¡gödelizarla! Asociemos, por tanto, a $\neg\mathrm{Dem}(x)$ su número de Gödel, que llamaremos d. Quizás este número sea tan gigantesco que no haya tinta suficiente en el mundo para escribirlo, pero nuestra filosofía ha sido siempre que el tamaño no importa; lo que importa es que sea un número.

Toda la estructura de la proposición «La fórmula de número de Gödel x no es demostrable» está contenida en un solo número: d. Ahora bien, el parámetro x no está fijo, no vale, pongamos, 14.451.937.500, sino que puede tomar cualquier valor. Y si puede tomar cualquier valor ¿por qué no escoger x igual a d, con gran malicia? Obtendríamos entonces el enunciado $\neg\mathrm{Dem}(d)$, que afirma que «La fórmula de número de Gödel d no es demostrable», pero como d es a su vez el número de Gödel de la proposición «La fórmula de número de Gödel x no es demostrable», $\neg\mathrm{Dem}(d)$ se transforma en «La fórmula *La fórmula de número de Gödel* x *no es demos-*

LA INCOMPLETITUD DE LAS TESELACIONES

Una teselación del plano es una forma de cubrirlo con alguna clase de baldosas, sin que éstas dejen huecos ni se superpongan entre sí. El arte islámico ofrece bellísimos ejemplos de teselaciones, pero también podemos encontrarlas en la naturaleza: las abejas teselan sus panales con hexágonos, el modo óptimo de hacerlo. Sin embargo, no todas las teselaciones tendrían por qué ser tan regulares: tal vez existan otras *aperiódicas*, en las que no sea posible encontrar ninguna simetría. En la década de los sesenta, el lógico Hao Wang (1921-1995) descubrió que si una cierta pregunta sobre teselaciones del plano era indecidible en el mismo sentido en el que no se podía demostrar ni refutar la frase «No soy demostrable», entonces existirían esos modos no periódicos de teselar el plano. Como esta posibilidad le resultó completamente absurda, Wang concluyó que su problema tenía que ser decidible. Unos años más tarde, sin embargo, uno de sus estudiantes demostró que con 20.426 baldosas distintas se podía teselar el plano de forma no periódica. Esa cantidad se ha ido poco a poco reduciendo hasta sólo dos azulejos distintos.





A la izquierda, teselación regular, formada por una sola clase de polígono regular, similar a un panal de abejas; a la derecha, un ejemplo de teselación aperiódica.

trable no es demostrable». Si ponemos la proposición en boca de la fórmula de número de Gödel x, estamos afirmando nada más y nada menos que «No soy demostrable»[2].

2 La naturaleza de este libro nos impide ser completamente rigurosos e incluir las fórmulas de reconocimiento de variables libres, sustitución y generalización que Gödel utiliza en su artículo. Creemos, sin embargo, que todos los ingredientes esenciales de la prueba han aparecido aquí.

Lo que no dice el teorema

El final del argumento con el que acabamos de probar que ningún conjunto de axiomas consistente y recursivo de la aritmética puede ser completo reproduce de modo muy fiel la escena en la que, en cualquier curso de lógica, muchos estudiantes vuelven a casa llorando y diciendo: «¡Mamá, yo nunca seré un lógico!», mientras al resto, *the happy few* —los «felices pocos» de los que hablaba Shakespeare— se les dibuja una sonrisa de satisfacción de oreja a oreja. Sin duda quisiéramos que el lector se encontrara en este último grupo. Aunque tal vez no lo hayamos conseguido, incluso aquellos que a estas alturas tengan ganas de gritar «¡Mamá, yo nunca seré un lógico!», o directamente hayan tirado con furia el libro a la basura, comprenderán que los teoremas de los que nos hemos ocupado aquí nada tienen que ver con una frase como: «Desde el día en el que Gödel demostró que no existe una prueba de la consistencia de la aritmética de Peano formalizable en el marco de esta teoría, los politólogos pudieron comprender, al fin, por qué había que momificar a Lenin y exhibirlo a los camaradas en un mausoleo».

Hay que reconocer que el autor de la cita, el ensayista francés Régis Debray, destaca por sus grandes dosis de imaginación, pero no por su ignorancia: nacido en 1940, cursó estudios de filosofía con Louis Althusser en la École Nórmale Supérieure de París. Preso en Bolivia, fue liberado tras una campaña internacional en la que unieron fuerzas personajes tan variopintos como Jean-Paul Sartre y el papa Pablo VI. En el tiempo libre que le dejaba la política comenzó a escribir su obra, hoy compuesta por alrededor de cincuenta libros, entre ellos *El escriba: la génesis del político*, del que hemos tomado el pasaje sobre la momia de Lenin y los camaradas.

El caso de Régis Debray no es único: otros intelectuales como los filósofos Gilles Deleuze y Julia Kristeva, el psicoanalista Jacques Lacan o el arquitecto Paul Virilio se han dejado arrastrar en tiempos recientes por lo que el filósofo francés Jaques Bouveresse ha llamado el «prodigio y vértigo de las analogías», para deducir, de un resultado lógico muy técnico, conclusiones generales que no guardan ninguna relación con las matemáticas, pero cuyo envoltorio pseudocientífico sin duda impresionará a los lectores.

Decía Horacio que la voz, una vez que se la deja ir, no sabe regresar. Aparte de las opiniones que puedan dejarse entrever en este libro, el lector tiene el poder de consultar las obras originales de Kristeva, Debray, Lacan, Deleuze y Virilio para decidir si sus pensamientos son una muestra más del peligro de hablar de lo que no se sabe o, por el contrario, el mejor reflejo del enorme poder de seducción de unos teoremas que, como decía John von Neumann, se divisarán siempre desde remotas distancias del espacio y del tiempo. A partir de ahora, nos centraremos sólo en quienes sí sabían de qué hablaban, y será así como entrará en escena uno de los hombres más geniales de la historia: Alan Mathison Turing.

Capítulo 5

Las máquinas de Turing

¿Qué puedo esperar?
Immanuel Kant

«Eur...» Betty esperaba con impaciencia que parase el traqueteo de los rotores para leer el resto del mensaje. «Europa...» Habían pasado más de cinco años desde el día en el que supo que la revista que amenizaba sus horas de sirvienta en una de las más ricas familias londinenses había organizado un concurso de crucigramas. «Europa nun...» A diario se esforzaba en recordar su sorpresa al recibir la noticia del premio y las vacilaciones cuando llegó la hora de pedir permiso para ausentarse una semana. «Europa nunca...» Después trataba de reconstruir el viaje junto a las demás aficionadas a los pasatiempos lógicos, hasta que la silueta de la mansión de Bletchley Park se dibujaba en su mente con la misma fuerza con la que se había recortado en el cielo gris aquel día de otoño, mientras el grupo abandonaba el bosque que habían atravesado para llegar allí. «Europa nunca será...» Tenía miedo de olvidar algún detalle de una historia que contaría a todo el mundo en cuanto terminara la guerra. R-u-s-a. La última palabra había tardado un rato en salir, pero ahora Bety podía celebrar un nuevo triunfo del ejército aliado: «Europa nunca será rusa». Era el 15 de abril de 1945, y así se dirigía Adolf Hitler a los altos mandos del partido nazi.

No eran los únicos que recibían el delirio del dictador dos semanas antes de suicidarse: sin que Himmler pudiera siquiera sospecharlo, diez mil personas leían a la vez su correspondencia con Hitler en un pueblecito situado a ochenta kilómetros de Londres, bien comunicado por tren y carretera, pero lo bastante escondido para evitar un bombardeo. Allí se había instalado en 1939 la Escuela Gubernamental de Códigos y Cifrado, cuya misión consistía en descifrar las instrucciones que los nazis codificaban con la máquina Enigma, la más perfecta construida hasta el momento. La había ideado en 1918 el ingeniero Arthur Scherbius para asegurar las transacciones comerciales, pero en vista de su potencial bélico, la marina alemana pronto se hizo con los derechos y se dedicó a perfeccionarla durante la siguiente década. Cuando las tropas de la Wehrmacht invadieron Polonia a principios de septiembre de 1939, los métodos criptográficos de Enigma habían alcanzado tal nivel de sofisticación que la posibilidad de que alguien fuera capaz de descifrar la máquina sencillamente no se consideraba.

Sólo el esfuerzo conjunto de un equipo del que formaban parte matemáticos, físicos y traductores, además de aquel grupo de mujeres que habían caído en la trampa del concurso de crucigramas, pudo hacer frente al diabólico dispositivo que transformaba, mediante un sistema de impulsos eléctricos enviados a una serie de rotores, la misma letra escrita dos veces seguidas en diferentes símbolos. Disfrazados de piratas, como si se tratase de un grupo de aburridos nobles en busca de un poco de diversión en tiempos de guerra, los primeros rompecódigos ocuparon en 1939 los barracones que se habían ido construyendo junto a la mansión victoriana. Nadie en el pueblo más cercano debía sospechar la crucial tarea que se llevaba a cabo en la Estación X, como empezó a llamarse el centro al que los aliados enviaban todos los mensajes recogidos en las trincheras. Incluso Winston Churchill se refería a Bletchley Park como «mi gallina de los huevos de oro que nunca cacarea».

A la derecha, militares nazis codifican sus mensajes mediante una máquina Enigma, uno de cuyos ejemplares se contempla a la izquierda.

Arriba, dependencias de Bletchley Park en las que se tradujo el código Enigma. Abajo, vista actual de la mansión.

Los polacos habían descubierto una singularidad de Enigma que la hacía menos segura de lo que pensaban los nazis: cada letra, fuera cual fuese su posición, se codificaba siempre mediante otra distinta. Sin embargo, hubo que resolver muchos problemas entre el hallazgo de esta primera pista y el momento en el que, cinco días antes del desembarco de Normandía, todo el equipo brindaba en Bletchley Park después de descodificar un mensaje en el que Hitler aseguraba que la llegada de las tropas americanas se produciría en Calais, casi trescientos kilómetros al noreste del puerto de Arromanches. Quizás ésta ni siquiera se hubiese producido sin las informaciones sobre la posición de los submarinos nazis que se lograron descifrar en la Estación X, un logro sorprendente teniendo en cuenta que en 1939 el equipo ni siquiera disponía de una máquina Enigma en la que poner a prueba sus hipótesis.

Trabajando día y noche, en turnos de ocho horas, los rompecódigos de Bletchley Park lograron construir un prototipo idéntico al que manejaban los nazis, pero la empresa nunca habría concluido con éxito de no ser por la intervención de un joven matemático inglés, al que muchos estudiantes de la Universidad de Cambridge comparaban con un dios griego: un *deus ex machina* reclutado para ganar la guerra. Sin Alan Turing (1912-1954) no habría sido fácil darse cuenta de que todos los mensajes hablaban antes o después sobre el tiempo que hacía, y de que ésa era la frase por la que debía comenzar la descodificación del texto. Por primera vez, la función fática del lenguaje, es decir, la que sirve para entablar la comunicación, se revelaba imprescindible.

Estatua de pizarra de Alan Turing, obra del escultor británico Stephen Kettle, junto a un retrato del matemático que se conserva en el National Museum of Computing de Bletchley Park (fuente: Jon Callas).

Otra de las propuestas del matemático fue construir un gran computador, La Bomba, que permitiera simular al mismo tiempo lo que ocurría en diez máquinas Enigma. Si Turing pudo ver más lejos que sus compañeros no fue porque sus años de aprendizaje hubiesen transcurrido en un laboratorio de última generación, sino porque durante mucho tiempo había explorado las fronteras de lo que él mismo consideraba la más bella creación humana: el teorema de Gödel.

UN DIÁLOGO DE *BREAKING THE CODE* (HERBERT WISE/HUGH WHITEMORE, 1996)

Dilly Knox: Me han proporcionado algunos detalles de su trabajo, señor Turing, muchos de los cuales, debo decirle, me resultan del todo incomprensibles.

Turing: No me sorprende demasiado.

Dilly Knox: Solía ser bueno en matemáticas cuando era joven, pero esto, en fin, me desconcierta. Por ejemplo, esto de aquí: «Sobre los números computables, con una aplicación al problema de la decisión». ¿Podría decirme algo sobre ello?

Turing: ¿Decirle qué?

Dilly Knox: No lo sé, algo, unas pocas palabras, una explicación en líneas generales.

Turing: ¿Unas pocas palabras?

Dilly Knox: Sí.

Turing: ¿En líneas generales?

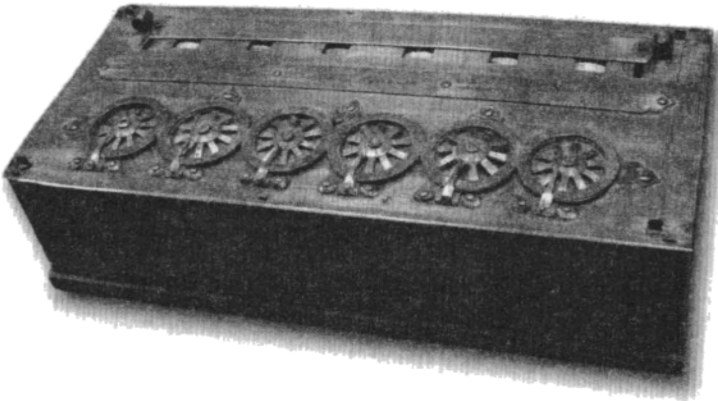
Dilly Knox: Si es posible...

Turing: Bueno... es sobre lo verdadero y lo falso... en líneas generales. Es un artículo técnico de lógica matemática, pero también trata sobre la dificultad de distinguir lo verdadero de lo falso. La gente piensa, bueno, mucha gente piensa que en matemáticas siempre sabemos lo que es cierto y lo que no. ¡Y no es así! ¡Ya nunca más! Es un problema que ha tenido ocupados a los matemáticos durante cuarenta o cincuenta años. ¿Cómo decírselo? Distinguir lo verdadero de lo falso, ¿sabe? [...]

Dilly Knox: Ya veo, bueno, en realidad no, pero algo veo. Sus ideas me parecen muy originales y estoy seguro de que será un miembro imprescindible de nuestro equipo, o nuestro grupo, llámelo como quiera.

Pensar como una máquina

Lejos de ser un hecho aislado, la construcción de La Bomba o de Colossus —el primer ordenador programable, también producto de Bletchley Park— se inscribía en una línea de continuidad que se remonta al menos hasta la segunda década del siglo XVII, cuando el astrónomo alemán Wilhelm Schickard (1592-1635) fabricó el primer «reloj calculante», un ingenio mecánico capaz de sumar, restar, multiplicar y dividir. A él le seguiría Blaise Pascal (1623-1662) con una calculadora que empezó a diseñar a los diecinueve años para hacer más llevadero el trabajo de su padre, al que habían nombrado recaudador de impuestos de la ciudad de Ruán. Comercializada con el nombre de pascalina, la nueva máquina causó furor en los salones de la aristocracia, donde científicos y nobles la contemplaban con fascinación. Allí pudo estudiarla por primera vez Gottfried Leibniz (1646-1716). Convencido como estaba de que «perder horas como esclavos en el trabajo del cálculo era indigno de los hombres excelentes», no es de extrañar que la pascalina lo entusiasmase y enseguida quisiera mejorar el modelo. Su sueño era fabricar una máquina capaz de reconocer todos los enunciados verdaderos.



La pascalina fue la primera calculadora de la historia, ideada por el francés Blaise Pascal.

A principios del siglo XIX, las calculadoras de Pascal y de Leibniz inspirarían al matemático inglés Charles Babbage (1791-1871) y a su alumna Ada Byron (1815-1852) las primeras reflexiones teóricas sobre la computación. Con el propósito de construir una «máquina analítica» *(Analytical Engine)*, Babbage y Byron distinguieron los elementos que intervienen en cualquier proceso informático. En primer lugar, debe existir un programa que indique las operaciones que han de realizarse. Consistirá en una serie de instrucciones que, a partir de un conjunto de datos que llamamos *input*, permite calcular un resultado que se devuelve al usuario como *output*. Por ejemplo, el *input* del programa «multiplicar» son pares de números como (2, 3) y el *output* es su producto, en este caso, $2 \cdot 3 = 6$. Para que el programa —que a menudo llamaremos *algoritmo*— pueda ejecutarse realmente, serán necesarios también un procesador que obedezca las instrucciones y una memoria en la que se almacenen los datos de partida, las instrucciones y todos los cálculos intermedios. En el caso de la máquina analítica de Babbage, el *input* se introducía por medio de unas tarjetas perforadas en un telar de Jacquard, diseñado, en un principio, para tejer automáticamente los patrones indicados por los agujeros.

Ada Byron era hija del gran poeta inglés Lord Byron y de Annabella Milbanke, a la que su marido llamaba «la princesa de los paralelogramos», pues había estudiado álgebra y geometría con un catedrático de Cambridge. Para que no siguiera los pasos del padre, al que había abandonado después de dar a luz a Ada, Annabella la inició tan pronto como pudo en el estudio de las ciencias. A los diecisiete años, la joven conoció a Charles Babbage en una cena organizada por su amiga Mary Somerville, que había sido su tutora y que la animó siempre a continuar con sus estudios matemáticos. Poco después, Ada explicaría a Babbage cómo calcular los números de Bernoulli a través de su sistema de tarjetas perforadas, un problema mucho más ambicioso, desde un punto de vista matemático, que los que el inventor de la máquina analítica había resuelto hasta la fecha. Con su método para «tejer álgebra pura», Byron no sólo escribió el primer programa informático de la historia, sino que puso también de manifiesto que para resolver un problema de modo algorítmico no era siempre necesario empezar de cero. Un puñado de operaciones básicas se repetía en casi todos los problemas, de modo que a menudo bastaba con combinar las tarjetas ya existentes. Es lo que los informáticos llaman hoy en día subrutinas.

Lo que distinguía a Ada Byron de Charles Babbage es lo mismo que permitiría a Alan Turing sentar las bases rigurosas de la teoría de la computación con su artículo «Sobre los números computables, con una aplicación al problema de la decisión», publicado en 1937 en los *Proceedings of the London Mathematical Society*. Mien-

A la izquierda, sello conmemorativo del centenario del nacimiento de Charles Babbage. Arriba, jardines dedicados a Ada Byron en Barcelona (foto: Ana Navarro Durán).

tras que Babbage seguía convencido en su lecho de muerte de que si viviera unos pocos años más, los frutos de la versión definitiva de la máquina analítica darían la vuelta al mundo, tanto Byron como Turing habían comprendido que era necesario avanzar mucho en el plano teórico antes de que pudiera construirse el primer or-

LOS NÚMEROS DE BERNOULLI

Cuenta una de las anécdotas más conocidas de Carl Friedrich Gauss que un día su maestro de la escuela primaria quiso descansar un rato mientras la clase sumaba los números del 1 al 100. Con lo que no contaba el señor Büttner era con que el niño Gauss iba a encontrar el resultado enseguida, aplicando un método que lo mismo le hubiese servido para calcular la suma del 1 al 1.000. Fijemos, pues, la cantidad n que queremos alcanzar. La idea de Gauss consiste en escribir la suma $1+2+\ldots+n$ al revés y aprovechar la simetría de los términos tal como indicamos a continuación:

$$\begin{gathered}1+2+\ldots+(n-1)+n\\ n+(n-1)+\ldots+2+1.\end{gathered}$$

El lector no tardará en convencerse de que si emparejamos cada término con el que está debajo, el resultado es siempre $n+1$. Como el proceso se repite n veces, obtendremos $n(n+1)$ como resultado. Pero, ¡atención!, porque al hacerlo así, hemos sumado cada número dos veces, una en la primera fila y otra en la segunda. Es necesario, por tanto, dividir por dos:

$$1+2+\ldots+n=\frac{n(n+1)}{2}.$$

Uno podría preguntarse si, al sustituir la suma de los n primeros números por los n primeros cuadrados, es posible obtener fórmulas similares. Con un método algo más sofisticado que el anterior puede demostrarse que

$$1^2+2^2+\ldots+n^2=\frac{1}{3}n^3+\frac{1}{2}n^2+\frac{1}{6}n,$$

y que la suma de los n primeros cubos se calcula mediante la fórmula

$$1^3+2^3+\ldots+n^3=\frac{1}{4}n^4+\frac{1}{2}n^3+\frac{1}{4}n^2.$$

En general, el k-ésimo número de Bernoulli está relacionado con los coeficientes que aparecen al escribir la suma de las n primeras potencias de orden k como un polinomio en la variable n. Se trata de unos números fáciles de definir con palabras, pero difíciles de calcular explícitamente. Por eso el algoritmo de Ada Byron representaba un gran paso adelante.

denador. Una de las cuestiones sobre las que más había que reflexionar era precisamente qué problemas podría resolver la máquina y cuáles no. Algo similar sucede hoy en día con la computación cuántica, cuyos logros teóricos van muy por delante del diseño efectivo del primer ordenador cuántico.

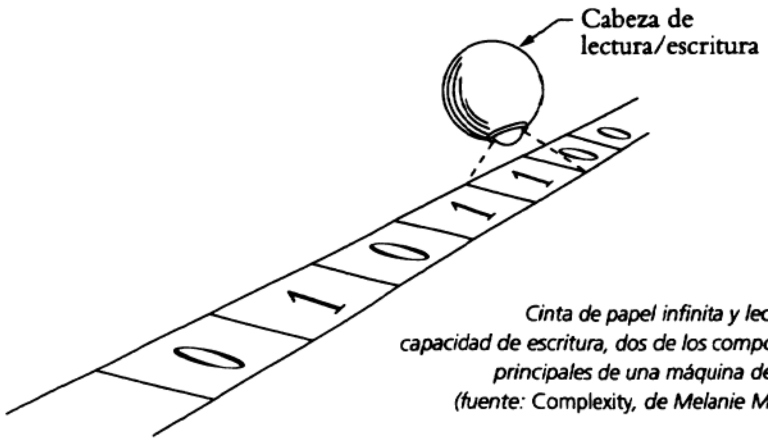
La idea genial de Turing para explorar los límites de acción de los futuros ordenadores fue plantearse con toda seriedad qué quiere decir pensar como una máquina. No hacía falta mucho esfuerzo para darse cuenta de que un ordenador no podría tener la inteligencia ni la imaginación de los seres humanos, que les permiten enfrentarse a situaciones desconocidas por completo. Por otro lado, las máquinas no se cansan ni se aburren cuando tienen que realizar un cálculo pesado; nunca tienen un mal día. ¡Son máquinas! Para distinguir los problemas que un ordenador no sería capaz de resolver por sus limitaciones técnicas (por ejemplo, porque el programa que hemos diseñado requiere la edad del universo para ejecutarse) de aquellos que son irresolubles por las condiciones mismas del enunciado, Turing imaginó una computadora ideal, con una memoria y un tiempo de ejecución infinitos a su alcance. Lo que no pudiera hacer esta máquina de Turing se le resistiría también al ordenador más potente del futuro, de manera que el método del matemático inglés serviría para poner límites a lo que podemos esperar de los ordenadores.

Funciones computables

El primer triunfo de las investigaciones de Turing consistió en definir el concepto de función computable. A partir de ahora, cada vez que digamos función entenderemos función definida sobre los números naturales que toma valores naturales. Recordemos que una función no es más que una forma de asociar a cada número otro número, que llamaremos su imagen. El lector puede pensar en las funciones —sólo si eso le da ánimos para seguir con el capítulo— como en una máquina que da forma a la materia prima que se le introduce. Así, nuestra función convertirá el número 3 en otro número que llamaremos $f(3)$, f de función. El proceso para obtener $f(n)$ a partir de n puede consistir en una serie de operaciones algebraicas o en una descripción verbal más compleja. Por ejemplo, si la función fuese la que asocia a cada número su sucesor, que, como vimos al comienzo del libro, interviene en los axiomas de Peano, entonces podríamos escribir $f(n) = n + 1$, y el resultado sería $f(3) = 3 + 1 = 4$. Si, por el contrario, la función determinara el número primo que ocupa la posición n-ésima, entonces $f(3)$ valdría 5, y $f(4)$ sería 7, porque los primeros primos son 2, 3, 5, 7, 11... En este caso disponemos de una descripción en palabras, pero no de una fórmula sencilla que exprese el valor de la función en cada punto.

La imagen de la máquina podría inducir a engaño, haciendo creer tal vez a los lectores que esa máquina de Turing ideal a la que nos hemos referido podría computar cualquier función imaginable. Por el contrario, las operaciones ocultas entre la entrada del número n y la salida de $f(n)$ podrían ser tan complicadas que ni siquiera la máquina de Turing pudiese llevarlas a cabo. Si queremos distinguir ambas situaciones, ha llegado el momento de explicar en detalle cómo funcionan esas máquinas que Alan Turing imaginó cuando tenía poco más de veinte años.

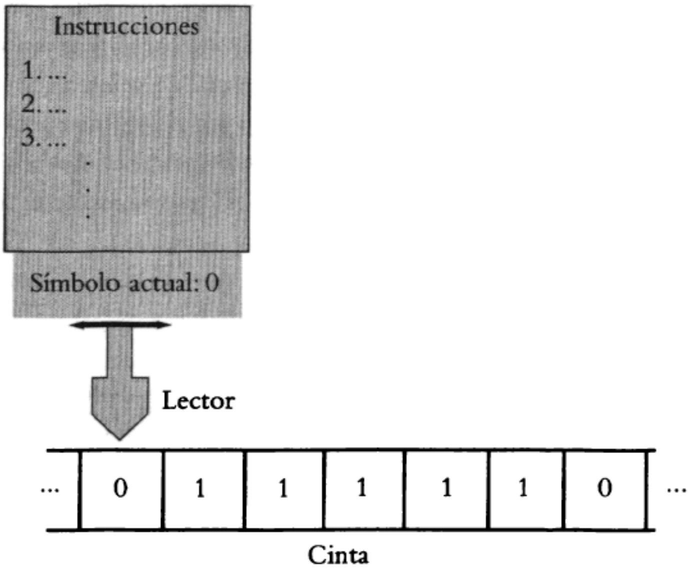
El primer elemento es una cinta de papel infinita por la derecha y por la izquierda —recordemos que se trata de una máquina ideal—, que está dividida en celdas en las que sólo cabe un símbolo, a elegir entre 0 y 1. Corresponden, como sabemos, a los dos valores de verdad posibles. La segunda pieza de la máquina de Turing es un lector capaz de detectar si el número escrito en una cierta celda es un 0 o un 1 y de escribir sobre él.



Cinta de papel infinita y lector con capacidad de escritura, dos de los componentes principales de una máquina de Turing (fuente: Complexity, *de Melanie Mitchell).*

Después de haberlo leído, el dispositivo puede responder de cinco formas diferentes: borrar el número que había y escribir un 0, reemplazar lo escrito por un 1, moverse hacia la derecha, moverse hacia la izquierda (para que estas dos operaciones puedan realizarse es fundamental que la cinta de papel sea infinita por los dos extremos) o simplemente pararse, es decir, no responder de ninguna manera a la lectura del símbolo. La secuencia de acciones está controlada por una cadena finita de instrucciones que indican a la máquina cómo tiene que responder en cada situación posible. Por ejemplo, la primera instrucción podría ser: «Si se lee el símbolo 1, moverse hacia la izquierda e ir a la tercera instrucción». Todas ellas siguen el esquema:

Instrucción número __: si el lector encuentra el símbolo __,
entonces ejecutar la operación __ e ir a la instrucción número __.



Funcionamiento de una máquina de Turing
(fuente: Complexity, *de Melanie Mitchell).*

Como hemos indicado, las instrucciones van numeradas empezando por el 1; los símbolos que aparecen son 0 y 1, y las operaciones posibles son escribir un 0 (0), escribir un 1 (1), ir a la derecha (D), ir a la izquierda (I) o pararse (P). Esto permite expresar de modo simbólico las instrucciones a través de los cuatro datos que intervienen en cada una de ellas. Así, si de nuevo la primera orden fuese «Si se lee el símbolo 1, moverse hacia la izquierda e ir a la tercera instrucción», bastaría con escribir (#1, 1, I, #3). A estas alturas, el lector ya se habrá dado cuenta de que, por cada número, hacen falta dos instrucciones: una que indique qué hacer en caso de que el símbolo encontrado sea un 0 y otra que explique cómo reaccionar cuando se lee un 1. Si en el ejemplo precedente la tercera orden nos indicara sólo qué hacer en caso de encontrar un 0, pero el símbolo siguiente fuera un 1, la máquina no sabría cómo continuar. Una solución posible consiste en decretar que cuando no existe la instrucción precisa, la máquina de Turing —que no tiene imaginación para continuar con el proceso ella sola— se detenga. Sin embargo, para que la exposición sea más clara indicaremos explícitamente cómo actuar en todos los casos posibles. Veamos un ejemplo muy sencillo, el de la máquina de Turing T formada por los siguientes tres comandos:

Instrucción #1: Si se lee un 0, escribir un 1 e ir a la instrucción #3.
Instrucción #1: Si se lee un 1, moverse a la derecha e ir a la instrucción #2.
Instrucción #2: Si se lee un 0, escribir un 1 e ir a la instrucción #3.
Instrucción #2: Si se lee un 1, pararse.
Instrucción #3: Si se lee un 0, escribir un 1 e ir a la instrucción #1.
Instrucción #3: Si se lee un 1, pararse.

A la hora de codificar esta máquina de Turing por medio del sistema del párrafo anterior, se nos plantea la duda de qué hacer cuando la máquina se detiene, pues en ese caso la instrucción no termina enviándonos a una nueva orden. La solución más fácil consiste en añadir un 0 al final: de este modo no hay equívoco posible, porque aunque la máquina de Turing trate de buscar la instrucción 0, ninguna orden lleva ese número. Sirviéndonos de este truco, la siguiente secuencia de instrucciones contiene toda la información de T:

(#1, 0, 1, #3)　　(#1, 1, D, #2)　　(#2, 0, 1, #3)
(#2, 1, P, #0)　　(#3, 0, 1, #1)　　(#3, 1, P, #0)

Veamos ahora cómo actúa el programa si introducimos como *input* una cinta de papel llena de 0s. La flecha indica en qué punto se encuentra el lector de la máquina de Turing en cada instante:

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
							⇑							

El programa comenzará a ejecutar la primera instrucción. Como el símbolo con el que el lector se encuentra es un 0 y la orden es «Si se lee un 0, escribir un 1 e ir a la instrucción #3», basta con reemplazar el 0 por un 1 y ver qué dice la tercera orden:

0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
							⇑							

Ahora bien, la instrucción #3 tiene dos partes: la primera indica que si leemos un 0, tenemos que escribir un 1 y volver a la orden #1, pero de acuerdo con la segunda, si la máquina de Turing encuentra el símbolo 1, entonces debe detenerse. Como es el caso, el programa ha terminado de ejecutarse. Por consiguiente, al introducir una cinta llena de 0s, T se para después de escribir un 1 en el punto de partida.

¿Verdad que son fáciles las máquinas de Turing? Veamos qué ocurre si esta vez aplicamos de nuevo el programa a la cinta que acabamos de obtener. El *input* es, por tanto:

0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
							⇑							

Empezaremos aplicando la primera instrucción: como lo que leemos es un 1, tenemos que movernos hacia la derecha e ir a la segunda orden. ¡Ningún misterio!

0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
								⇑						

Llegados a este punto, la instrucción #2 establece que, al encontrarse con un 0, la máquina de Turing tiene que sustituirlo por un 1, e ir a la tercera orden. Obedezcamos:

0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
								⇑						

Una vez más, la instrucción #3 obliga a T a detenerse cuando lee el símbolo 1, luego el programa ha terminado de ejecutarse, y el *output* es una cinta con dos 1s entre una infinidad de 0s, con el lector situado sobre el que se encuentra más a la derecha. Si volvemos a poner en marcha la máquina de Turing, el nuevo resultado será una cinta de papel con tres 1s en lugar de dos, de modo que lo que T calcula no es sino la función $f(n) = n + 1$. En general, una función es computable si existe una máquina de Turing que calcula cada uno de sus valores.

Supongamos que el número natural n se codifica, como hemos hecho en el ejemplo anterior, mediante el *input* de una cinta de papel compuesta por n 1s entre una infinidad de 0s a la derecha y a la izquierda, con el lector situado en el último de ellos. Una función f será computable si existe una máquina de Turing que, introduciendo de este modo cualquier valor de n, devuelva como *output* la metamorfosis $f(n)$. Lo que hemos demostrado es que la función «sumar uno» es computable en el sentido de Turing. Como para calcular la función $f(n) = n + 2$ basta con aplicar dos veces el mismo conjunto de instrucciones, y para computar $f(n) = n + 3$ es suficiente con repetir tres veces el proceso, y así sucesivamente, la suma es una operación computable. También lo es la multiplicación, pues tal y como nos enseñaron en el colegio, multiplicar 3 y 5 significa sumar tres veces el número 5, o cinco veces el número 3: «el orden de los factores no altera el producto…».

Decíamos que una función es computable si existe una máquina de Turing que calcule cada uno de sus valores, pero eso no significa que sepamos siempre encontrarla. Consideremos, por ejemplo, una función que sólo admite como *input* y como *output* los valores 0 y 1. Es suficiente entonces con especificar $f(0)$, que puede ser igual a 0 o a 1, y $f(1)$, que también toma uno de estos dos valores.

Al lector no le costará ningún trabajo convencerse de que sólo existen cuatro funciones con estas características: la que siempre vale 0; la función constante igual a 1; la que en 0 toma el valor 0, y en 1, el valor 1, y la que asocia a 0 el número 1, y a 1, el número 0. Como las posibilidades son finitas, todas ellas son funciones computables, pues es posible —al menos, en teoría— describir *ad hoc* un conjunto de instrucciones que calcule sus valores. Sin embargo, la descripción de cómo calcular la imagen de alguno de los valores podría ser tan compleja que no fuéramos capaces de construir de modo explícito ninguna máquina de Turing que compute la función. Veamos un ejemplo, sugerido por Arturo Sangalli.

Consideremos la función definida sobre los números del 1 al 9 que hace corresponder a n el valor 1 si el desarrollo decimal del número π contiene n cifras n consecutivas (por ejemplo, la combinación 4444 si $n = 4$) y 0, en otro caso. Con esta definición, $f(1)$ vale 1 porque en el desarrollo decimal de π, que comienza por 3,141592… aparece un 1 (por ejemplo, en la primera posición).

Análogamente, $f(2)$ también es igual a 1, pero para encontrar la cadena 22 es preciso recorrer los 135 primeros decimales: …4460955058 **22** 31725359408… La siguiente tabla ha sido realizada mediante un programa con el que el lector puede hacer otros experimentos, si así lo desea, a través del sitio web: http://www.angio.net/pi/bigpi.cgi.

Patrón	Posición	Patrón	Posición
333	1.698	666666	252.499
4444	54.525	7777777	3.346.228
55555	24.466	88888888	46.663.520

De la tabla se deduce que nuestra función toma el valor 1 en los ocho primeros naturales, pues π contiene los patrones 333, 4444, 55555, 666666, 7777777 y 88888888. Para calcular el valor de $f(9)$, podríamos pensar en un programa que fuese recorriendo una a una todas las cifras de π hasta que encontrara la combinación requerida, es decir, un bloque de nueve 9s consecutivos. Si en efecto existe, antes o después el programa lo encontrará, luego el *output* será 1. No importa el tiempo que se tarde porque, como hemos repetido en varias ocasiones, se trata de una máquina ideal, sin las limitaciones físicas que tienen los ordenadores. Sin embargo, si no hubiera un bloque formado por nueve 9s consecutivos, el programa no se detendría nunca, y no sabríamos decidir cuánto vale $f(9)$. Por tanto, este enfoque nunca nos permitirá saber si f es computable, a menos que podamos demostrar primero que en algún lugar del desarrollo decimal de π se encuentran nueve 9s seguidos. Pero en ese caso, el programa sería inútil, pues el mismo argumento demostraría que $f(9)$ vale 1. Aunque el enfoque más evidente no da resultados, esta función es computable. Para demostrarlo, hace falta razonar como hemos hecho un poco más arriba: puesto que el número de funciones definidas de 1 a 9 que toman los valores 0 y 1 es finito (en este caso, 512, algo menos manejable que nuestras cuatro

¿Y SI TODO FUESE UN NÚMERO?

En su cuento *La Biblioteca de Babel*, el escritor argentino Jorge Luis Borges sugiere que toda la información del universo podría almacenarse en un único libro, que «constara de un número infinito de hojas infinitamente delgadas». Pero, ¿por qué almacenarla en esa suma ingobernable si quizá quepa en un número? Una de las conjeturas más misteriosas a las que se enfrentan hoy los matemáticos consiste en demostrar que en las cifras decimales del número π, el cociente entre la longitud y el diámetro de la circunferencia, aparece antes o después cualquier patrón numérico que se nos ocurra. De ser así, no sólo la combinación 999999999 aparecería antes o después, sino que en su escritura decimal podría codificarse cualquier mensaje: pasado, presente y futuro.

simpáticas funciones definidas en 0 y en 1 con valores 0 o 1), existe una máquina de Turing que computa cada una de ellas. Aquí tenemos un ejemplo de función computable cuya máquina de Turing no se sabe describir explícitamente.

Otra clase de funciones computables son las recursivas, es decir, aquellas en las que $f(n)$ puede calcularse a partir de los valores que toma la función en otros números menores que n. Gran parte de las funciones que los matemáticos emplean a diario son recursivas, pero ¿son todas las funciones computables? Alan Turing llegó enseguida a la conclusión de que la respuesta es negativa: hay muchas funciones que ninguna máquina de Turing puede calcular y, lo que es más grave, si se elige una función al azar, es casi seguro que no será computable. Al mismo tiempo, al otro lado del Atlántico, el lógico Alonzo Church (1903-1995) llegaba a idénticas conclusiones en la Universidad de Princeton mediante el desarrollo de un sistema formal que se llamó cálculo lambda. Ambas ideas eran tan novedosas que la única persona que los editores de los *Proceedings of the London Mathematical Society* pudieron encontrar para que evaluase el artículo de Turing fue precisamente Church. Se inició así una fecunda colaboración, interrumpida por la guerra, que permitiría a los científicos formular el principio que hoy se conoce como «tesis de Church-Turing». Podría haber otras definiciones de función computable, pero si se acepta la tesis, todas ellas serían equivalentes a la existencia de una máquina de Turing que calcule los valores de la función.

Alonzo Church, creador del cálculo lambda, y colaborador de Turing.

Para demostrar que casi ninguna función es computable, Alan Turing recurrió a una variante ingeniosa del argumento diagonal de Cantor que estudiamos con detalle en el capítulo 2. En él vimos que no había modo de ordenar en una lista las sucesiones de 0s y de 1s. En cuanto suponíamos que podían colocarse una detrás de otra, modificando los valores de los elementos de la diagonal, lográbamos construir una sucesión que, a pesar de estar formada únicamente por 0s y por 1s, no coincidía con ninguno de los de la lista. Este mismo razonamiento nos permite concluir que las funciones no son numerables.

A fin de cuentas, ¿qué es una función? Hemos dicho que se trata de un método para transformar 0 en $f(0)$, 1 en $f(1)$, 2 en $f(2)$, y así hasta el infinito. Por tanto, toda la información de f está contenida en la sucesión de números $f(0), f(1), f(2), f(3)\ldots$ Pensemos, para simplificar las cosas, en funciones que sólo toman los valores 0 y 1; por ejemplo, en la función f que vale 0 cuando un número es par, y 1 si es impar. En este caso, toda la información de f aparece en la secuencia 0 1 0 1 0 1 0 1 0 1…, pues si queremos calcular la imagen de n basta con avanzar hasta la posición n-ésima y ver qué símbolo encontramos. Esperamos haber convencido así al lector de que las funciones que sólo toman los valores 0 y 1 son exactamente lo mismo que las sucesiones infinitas de 0s y de 1s. Por tanto, ¡no numerables!

Como cada máquina de Turing computa una única función, lo primero que habría que demostrar, para que la esperanza de que todas las funciones sean computables siguiera teniendo sentido, es que hay al menos tantas máquinas como funciones que se quieren calcular. Y, sin embargo, no es así: Turing demostró que el infinito de sus máquinas era mucho más pequeño. Para ver que las funciones no son numerables, primero hubo que codificarlas a través de sucesiones formadas por 0s y 1s. En el caso de las máquinas de Turing, disponemos ya de un modo simbólico de escribirlas, pues una de estas máquinas no es más que una lista finita de instrucciones, y cada una de ellas puede traducirse a unos cuantos símbolos. Como vimos, (#1, 1, I, #3) dice lo mismo que «Instrucción número 1: si se lee el símbolo 1, moverse hacia la izquierda e ir a la tercera instrucción». Una vez que hemos expresado la máquina de Turing como una secuencia de instrucciones escritas de este modo, el lector puede buscar algún tipo de orden que permita escribir todas las máquinas de Turing en una lista.

Sin embargo, para lo que sigue, nos interesa más servirnos del mismo proceso de *gödelización* que estudiamos en el capítulo 4. Recordemos que consistía en una forma de asignar unos números naturales gigantescos a cada fórmula de la lógica de primer orden, de modo que, conociendo el número, se pudiese reconstruir la fórmula. Este procedimiento, aplicado esta vez al código de las máquinas de Turing, nos permite encapsular toda la información del programa en un solo número. Como ocurría con la *gödelización*, no todos los números representan una máquina de Turing, sino sólo aquellos que cumplen ciertas propiedades. Por un lado, existen infinitas máquinas de Turing, pero, por otro, ese infinito no puede exceder al de los números naturales, pues cada máquina de Turing está codificada a través de uno de ellos. Hemos demostrado así que las máquinas de Turing son numerables, luego también lo son las funciones computables: agujas en el pajar de todas las funciones.

El problema de la parada

El sueño de Leibniz de construir una máquina capaz de distinguir los enunciados verdaderos de los falsos lo retomaría David Hilbert en el siglo XX. Como indicamos en el capítulo 3, el programa de Hilbert para limpiar de paradojas las matemáticas no sólo consistía en dotarlas de unos fundamentos lo más sólidos posibles: eso ya lo habían hecho los antiguos, empezando por Euclides, y no había funcionado. Para tener la certeza absoluta de que, en el futuro, ningún Russell se sacaría otra paradoja del sombrero, la tarea matemática de cimentar el edificio de la lógica debía acompañarse de un *cálculo de estructuras* metamatemático, que demostrara que los pilares podían, en efecto, aguantar el peso de la bóveda. Las dos primeras cuestiones que se planteó Hilbert fueron si las matemáticas eran completas y si eran consistentes, es decir, si lo verdadero coincidía con lo demostrable, y si no había peligro de encontrar contradicciones. Tres años antes de que Gödel probara que, en el caso de la aritmética, los dos requerimientos eran incompatibles, David Hilbert y su discípulo Wilhelm Ackermann (1896-1962) añadieron una tercera cuestión, expuesta en la primera conferencia plenaria del Congreso Internacional de Matemáticos de 1928.

El problema de la decisión (*Entscheidungsproblem*) se preguntaba si existía un algoritmo capaz de recibir como *input* un enunciado matemático y de devolver como *output* verdadero o falso. Mientras que era razonable exigir a los axiomas que fuesen recursivos, no ocurría lo mismo con el conjunto de teoremas, como veremos enseguida. Pero antes recrearemos para el lector una escena muy relacionada con el nuevo problema de Hilbert, de la que el autor fue testigo precisamente en otro Congreso Internacional de Matemáticos, el que se celebró en Madrid en agosto de 2006.

A la salida de una de las conferencias, cierto matemático conversaba con alguien a quien había confundido con un periodista. Después de un intercambio de bromas sobre la banda de ladrones que había atracado a algunos de los asistentes haciéndose pasar por policías, la segunda persona quiso saber a qué se dedicaba la primera. Pregunta peligrosa donde las haya, porque lo más probable es que desemboque en un monólogo de media hora en el que el entusiasmo de quien habla irá aumentando con la misma velocidad con la que disminuye el del que escucha. Pero esa vez el matemático había decidido que el periodista no iba a entender nada, así que se limitó a explicarle: «Pues mira, yo tengo una maquinita a la que le meto un enunciado para que me diga si es verdadero o falso», a lo que el supuesto periodista, que había sabido hasta entonces cómo camuflarse, respondió: «¡Genial! A ver si un día de éstos me prestas tu maquinita, porque yo trabajo con un montón de conjeturas matemáticas que no tengo ni idea de si son ciertas o no».

A todos nos encantaría disponer de ese programa del que el matemático presumía ante el falso periodista, pero como consecuencia de sus investigaciones sobre las funciones computables, Alan Turing demostró que era imposible. Para hacerlo, imaginó una máquina universal, que no sólo admitía como *input* números, sino también las instrucciones de cualquier máquina de Turing. Si las órdenes correspondían a lo que ahora llamamos un programa, la máquina universal era el ordenador en sí, capaz de imitar, al menos a nivel teórico, lo que hace cualquier máquina de Turing. Esta computadora abstracta anticipaba en varios años la arquitectura de nuestros ordenadores, de modo que la redacción de la revista *Time* no exageraba cuando, al elegir a Alan Turing entre los personajes del milenio, proponía un ejercicio de memoria: «Cada vez que tecleamos en un ordenador, estamos trabajando con una reencarnación de las máquinas de Turing». Sirviéndose de esta computadora *avant la lettre*, Turing exhibió el absurdo al que conducía la existencia de esta «máquina de la verdad».

Veamos en qué consiste la solución de Turing al *Entscheidungsproblem*. Para empezar, nuestro protagonista supuso que el sueño de Hilbert era realizable, es decir, que existía un procedimiento mecánico capaz de decidir en tiempo finito si un enunciado es verdadero o falso. En particular, ese algoritmo nos permitiría saber si la afirmación «La máquina de Turing T se para cuando se le introduce el *input n*» es verdadera o falsa. Como ya hemos indicado, gracias al método de la *gödelización* podemos asociar un número a cada máquina de Turing, de modo que toda su estructura esté codificada en él. Cuando *n* sea el número de una máquina de Turing, la denotaremos por T(*n*). Con esta notación, el problema que queremos resolver puede enunciarse en los siguientes términos: ¿Se detiene la máquina de Turing T(*n*) cuando se le introduce como *input* el número *m*? Debemos insistir en el hecho de que, si la máquina perfecta que imaginaba Hilbert existiese, entonces no sólo sería capaz de responder a esta pregunta en algunos casos, sino siempre, fuesen cuales fuesen los valores de los parámetros *m* y *n*. Se trata, por lo tanto, de una función de dos variables, que toma el par de números (*m*, *n*) y que predice si la máquina de Turing asociada a *n* se detiene cuando se la pone en marcha con una cinta que represente el número *m*. En el ejemplo sobre el número π, si llamamos *t* al número de la máquina de Turing que recorre sus cifras decimales en busca de las combinaciones requeridas, al introducir los parámetros (9, *t*), nuestra función responderá 1 si entre los dígitos de π aparecen nueve 9s seguidos (porque entonces la máquina se para), y 0 si no (pues seguiría ejecutándose indefinidamente).

Sin embargo, al suponer que existe una máquina de Turing P (P de parada) que resuelve este problema, llegamos enseguida a una contradicción. Para convencerse, merece la pena repetir una vez más qué es lo que hace P: se trata de la máquina de Turing cuyo *input* son pares de números (*m*, *n*) y cuyo *output* sólo toma dos valores: 1 si la máquina de Turing T(*n*) se detiene para el valor inicial *m*, y 0 en otro caso, es decir, si o bien no existe ninguna máquina de Turing representada por el número *n* (pues no todos los naturales son números de alguna máquina de Turing) o bien, aunque ésta exista, el programa continúa ejecutándose indefinidamente cuando se le introduce el parámetro *m*. Ésta es una de las bestias negras de los informáticos, pero, sin perder la alegría, ellos la llaman con el sugerente nombre de *bucle infinito*. Lo que nos importa aquí es que, si dispusiéramos de esta máquina T, podríamos construir sin dificultad alguna otra máquina de Turing, a la que llamaríamos C (C de contradicción), cuyo *input* fuese un único número *n* y que actuara de este modo:

— Si la máquina de Turing T(*n*) se detiene cuando se le introduce el *input n* (dicho de otro modo, si P(*n*, *n*) vale 1), entonces C no se para nunca.

— Si la máquina de Turing T(*n*) continúa ejecutándose indefinidamente a partir del valor *n* (es decir, si P(*n*, *n*) vale 0), entonces C se detiene nada más comenzar.

En el capítulo 2 vimos cómo la paradoja del mentiroso, que tanto atormentó al bueno de Epiménides, surgía al hacer decir a un cretense que todos los cretenses eran mentirosos, o bien cuando una frase predicaba de sí misma «Esta afirmación es falsa». Más adelante, quisimos mostrar cómo había aprovechado Gödel la autorreferencia para construir un enunciado verdadero, pero indemostrable, precisamente el que afirmaba «Esta frase es indemostrable». Tras este recordatorio, seguro que el lector ya sabe cómo terminar el argumento: hemos definido una máquina de Turing C, que se detiene o sigue ejecutándose sin pausa en función de lo que hace otra máquina T(*n*). Pero ¿qué pasa si a C le introducimos como *input* la propia C, es decir, el número asociado *c*? Si la máquina T(*c*) se detiene, entonces C no se detiene. Si, por el contrario, T(*c*) entra en un bucle infinito, entonces C se para. ¡Pero C y T(*c*) son la misma máquina! ¡No pueden hacer cosas distintas! Suponiendo que el problema de la parada tuviera solución para cualquier valor de *m* y *n*, hemos llegado a una contradicción, pues cuando el demonio de la autorreferencia nos susurra al oído ¡elige *c*! resulta que la misma máquina se comporta de dos formas distintas.

El sueño de Hilbert y de Leibniz era una utopía. El mismo juego de espejos, la autorreferencia, primero había animado a Bertrand Russell a lanzarse a una refundación de las matemáticas sobre unas bases más seguras; después había permitido a Gödel demostrar que el optimismo de la época no estaba justificado, y ahora Turing lo volvía a utilizar en su solución al *Entscheidungsproblem*, disfrazado esta vez de máquinas teóricas de las que nacerían los ordenadores.

Hemos dicho que la lógica no se ocupa de cómo razonamos en la vida diaria, sino de cómo deberíamos hacerlo para estar seguros de que la conclusión es verdadera. Y en efecto, hasta ahora sólo hemos atendido a fórmulas en las que los valores de verdad 0 y 1 aparecían vacíos de significado. Blanco o negro. En el próximo capítulo, sin embargo, intentaremos describir cómo sería un mundo con grises; menos gris, pero también más inseguro…

Capítulo 6

Bien acaba lo que no acaba

Un saber casi de miligramos requiere un largo y penoso noviciado que sólo las personas más puras tienen el valor de afrontar.
Marguerite Duras, citada por Juan Goytisolo

Tal vez sí que supiera lo que hacía cuando la llevó a aquel restaurante frecuentado sólo por japoneses. Aunque tenía pocas dudas sobre sus encantos, si no funcionaba la seducción por viajes y lecturas, podría aún salvar la noche manejando sutilmente los palillos y sorprendiéndola con uno de esos platos de nombre sinestésico. Cuando la camarera, no tan guapa como mandan las historias, se les acercó a la hora de los postres, todo iba por buen camino. Se notaba que había aprendido el idioma de pequeña, así que en el momento en que se dio la vuelta y volvió sobre sus pasos para preguntarles, exactamente, si las trufas de té las querían «con nata, sin nata o qué», ambos comprendieron que lo que allí ocurría no era una pérdida en la traducción. El hombre que durante toda la velada poco menos que había fanfarroneado con sus aventuras por rincones del planeta donde nunca antes que él había entrado un europeo no estaba dispuesto a desaprovechar el margen de la duda. Desconcertado, pero decidido, aventuró: «¡qué!». Al rato, la camarera regresaba, sonriente, con un plato de trufas en el que había sólo un poco de nata, alejada del centro del postre. Fue entonces cuando, con complicidad binaria, los dos se miraron a los ojos y dijeron a la vez: «Malditos orientales, que no tienen principio de no contradicción».

La lógica difusa

A pesar de sus variadas apariencias, todos los conjuntos que hemos considerado hasta ahora comparten una propiedad común: dado cualquier elemento y cualquier conjunto, la pregunta «¿Pertenece el elemento al conjunto?» sólo tiene una respuesta: sí o no. Es cierto que la descripción que lo define podría ser tan complicada que no supiésemos contestar, pero lo importante es que en algún rincón del mundo matemático está escrito sí o no. Es lo que ocurría en el ejemplo de los números cuyo desarrollo decimal contiene todos los patrones imaginables, que presentamos en el capítulo anterior. No sabemos si π cae dentro o fuera del conjunto, pero sólo una respuesta es válida. También las proposiciones de la lógica se ajustan a este esquema: son verdaderas o falsas, y cualquier otra posibilidad está excluida. Tanto es así que las dos paradojas principales de las que nos hemos ocupado (la de Russell y la del mentiroso) surgen cuando no es posible, ni siquiera desde un punto de vista teórico, responder verdadero o falso, pertenece o no pertenece. Y no es que nuestra regla admita una excepción, sino que ni el conjunto de todos los conjuntos que no son miembros de sí mismos ni la frase «Esta frase es falsa» son expresiones formalmente correctas, sea porque la relación de pertenencia sólo es aplicable entre objetos de tipo sucesivo o porque el concepto de verdad no pertenece al lenguaje, sino al metalenguaje. De algún modo, la teoría de conjuntos y la lógica funcionan por acantilados: verdadero está al borde, y basta con un soplo de aire para que se precipite en caída libre hacia lo falso. Sin embargo, en la geografía terrestre lo que abundan no son los acantilados, sino las pendientes suaves que se internan en el mar muy poco a poco.

Hace algunos años, la publicidad televisiva del juego de mesa Scattergories causó furor en numerosos países. Se trata de un juego en el que se elige al azar una letra del alfabeto y a continuación hay que escribir palabras de distintos campos semánticos que empiecen por ella. Por ejemplo, si recibimos la lista «Deportes. Títulos de canciones. Partes del cuerpo. Comidas étnicas. Cosas que se gritan» y, al lanzar el icosaedro que actúa como dado, sale la letra T, una respuesta posible sería: «Tenis. *Tu nombre me sabe a yerba.* Tibia. Tayín. ¡Tonto!». En el anuncio, un joven se marchaba de casa enfadado, llevándose consigo el Scattergories, porque sus compañeros no le habían admitido la respuesta «barco» en la categoría de «animales acuáticos». Deciden por fin ceder, viendo que es la única forma de seguir jugando, pero el chico vuelve a las andadas en la siguiente ronda, esta vez con la letra P, y sus amigos se preguntan, sin saber bien qué decir: «¿Aceptamos pulpo como animal de compañía?»

Si existen seres vivos cuya clasificación como animales plantea graves problemas, el conjunto de los animales de compañía está aún peor delimitado: nadie pone en duda que los perros y los gatos formen parte de él, y con la misma convicción podemos asegurar que ni los lobos ni los elefantes pertenecen al conjunto. Sin embargo, mientras que algunos incluirían las tarántulas en la clase «animales que no quiero tener a menos de un kilómetro», hay quien se entretiene lanzándoles grillos por el agujero de una jaula. Tan mal definidos como los animales de compañía están otros conjuntos que empleamos a diario, como el de las personas guapas, el de los buenos restaurantes o el de los chistes divertidos. El primero en proponer una teoría que respondiese mejor a estas situaciones fue el polaco Jan Łukasiewicz (1878-1956), que en 1917 introdujo la lógica trivaluada, en la que además de verdaderas y falsas, las proposiciones podían ser «posibles». Por ejemplo, una persona de 1,50 m de estatura es baja; otra que mida 2 m es alta, y alguien de 1,75 m es «posiblemente alto» o «posiblemente bajo», todo depende de si lo comparamos con una tribu de pigmeos o con un equipo de la NBA.

LA VENGANZA DEL MENTIROSO

Si volvemos a examinar la paradoja del mentiroso a la luz de la lógica trivaluada de Łukasiewicz, nos damos cuenta de que la contradicción ha desaparecido, pues el ingrediente esencial de nuestro análisis consistió en deducir que si un enunciado como «Esta frase es falsa» no era verdadero, entonces tenía que ser falso. Sin embargo, en la nueva lógica hay afirmaciones que no son ni verdaderas ni falsas, sino posibles. Como la razón de fondo de la paradoja es más profunda que este principio de bivalencia, es posible modificarla para que siga siendo válida en la lógica trivaluada. Pensemos en el enunciado «Esta frase no es verdadera». Las afirmaciones se dividen en tres clases (verdaderas, falsas y posibles), así que razonaremos caso por caso. Si el enunciado es verdadero, entonces debe ser cierto lo que dice, luego no es verdadero. Si, por el contrario, el enunciado es falso o posible, entonces no es verdadero, pero como ése es exactamente su contenido, tiene que serlo. Tampoco en la nueva lógica es posible atribuir un valor de verdad a la sentencia «Esta frase no es verdadera».

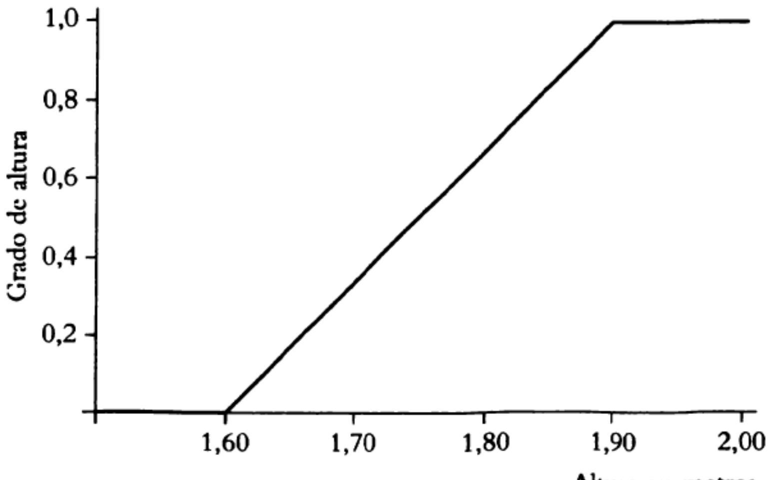
Incluir «posible» entre los valores de verdad representa un avance con respecto al mundo en blanco y negro de la lógica clásica, pero aún así no es suficiente, pues se trata de un punto indeterminado, y lo que nos interesa es poder tomar decisiones. Supongamos que un periodista se plantea dimitir tras el cambio de la línea editorial de su periódico. Llamaremos P a la proposición «No estoy de acuerdo con la nueva ideología del periódico». Entonces, la estructura de una decisión clásica sería «Si P es verdadera, dimito» y «Si P es falsa, me quedo». Como estar de acuerdo es siempre una cuestión llena de matices, al periodista le vendría muy bien disponer de un tercer valor de verdad. Pero ¿cómo interpretar «posible»? Si P es posible, ¿dimito o me quedo? ¿Qué barrera separa una reacción de la otra? Si queremos una lógica que nos permita tomar esta clase de decisiones, tenemos que ser mucho más precisos.

Es aquí donde entra en escena el profesor de la Universidad de Berkeley Lotfi Zadeh, que en 1965 propuso que la pertenencia a un conjunto o la verdad de una proposición tomasen cualquier valor comprendido entre 0 y 1. De este modo, los jugadores del Scattergories podrían decretar que sólo son válidas aquellas respuestas que pertenecen, por ejemplo, más de 0,6 al campo semántico en cuestión, y el periodista podría decidir que, si su desacuerdo con la nueva línea editorial del periódico superaba, pongamos, el 0,45, entonces dimitiría. Zadeh denominó los nuevos conjuntos con el término inglés *fuzzy*, que en una de sus acepciones significa borroso o difuso, lo que no tiene límites bien definidos. En los conjuntos borrosos, por tanto, a la pregunta ¿pertenece el elemento al conjunto? cabe responder de infinitos modos diferentes.

Lotfi Zadeh, el creador de la lógica difusa (fuente: Wolfgang Hunsche).

Al lector tal vez le tiente la idea de interpretar los conjuntos borrosos en términos probabilísticos. Puede que así la exposición sea más clara, pero decir que el nivel de pertenencia de un elemento a un conjunto es la probabilidad de que esté contenido en él sería traicionar el espíritu con el que Zadeh inventó la lógica difusa. Veamos lo que ocurre al lanzar una moneda al aire: desde nuestra tierna infancia sabemos que la probabilidad de que salga cara es del cincuenta por ciento, y eso significa que si tiramos la moneda muchas veces, digamos que diez mil, más o menos la mitad de los resultados serán cara, y la otra mitad, cruz. Sin embargo, en cada lanzamiento se obtiene un solo resultado: cara o cruz, sí o no, pertenece o no pertenece. La probabilidad, al menos en su versión más simple, refleja nuestro conocimiento limitado de las cosas: si supiéramos con absoluta precisión la fuerza con la que se lanza la moneda, si pudiéramos convertirnos en el dios Eolo y controlar los vientos, entonces seríamos capaces de predecir el resultado. Esto significa que el principio subyacente a esta encarnación sencilla de la teoría de la probabilidad coincide, en el fondo, con el de la lógica clásica, mientras que en el mundo borroso, al lanzar una moneda al aire puede salir sólo cara, más cara que cruz, más cruz que cara, sólo cruz, o cualquiera de las variantes intermedias, expresadas con infinita precisión.

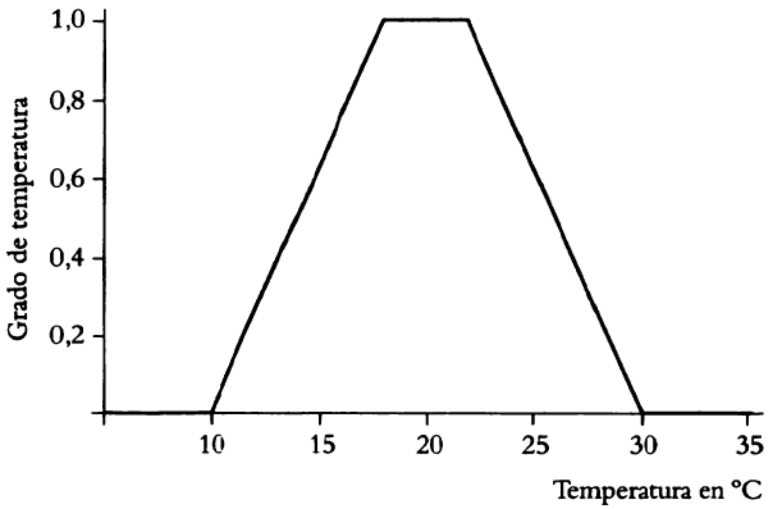
A diferencia de los conjuntos clásicos, cuya frontera es un abismo, las colecciones que estudia la lógica difusa están delimitadas por una función de pertenencia que reproduce la forma de la pendiente con la que el talud se adentra poco a poco en el mar. Pensemos, por ejemplo, en las personas altas: si decidimos que hasta 1,60 m se es completamente bajo, y a partir de 1,90 m, completamente alto, entonces la función de pertenencia al conjunto adopta el siguiente aspecto:



Función de pertenencia al conjunto borroso de las personas altas.
La función adopta forma de «talud».

Haciendo algunos cálculos se puede demostrar que todo el que mida menos de 1,60 m tendrá un grado de pertenencia al conjunto de las personas altas de 0; si su estatura es mayor o igual a 1,90 m, será del todo alto, y si se encuentra, como la mayoría, entre estos dos valores, entonces para calcular el valor de pertenencia deberá multiplicar por 10 su altura (medida en metros), restar 16 al resultado y dividir por 3 el número que obtenga. Sólo con conocer que el grado de altura de una persona es de 0,5, como es el caso del autor de este libro, basta para saber cuánto mide.

En otros casos, las funciones de pertenencia podrían tener forma de triángulo o de trapecio. Si se considera, por ejemplo, que por debajo de 10 °C de temperatura hace mucho frío, que por encima de 30 °C, demasiado calor, y que la temperatura perfecta está comprendida entre los 18 °C y los 22 °C, entonces la función de pertenencia al conjunto de las temperaturas agradables se parecería al gráfico que muestra la figura de abajo. ¡Compararla con los climogramas no sería un mal modo de escoger el lugar donde vivir! O al menos de descartar unos cuantos...



Función de pertenencia al conjunto borroso de las temperaturas agradables. La función adopta forma de trapecio.

Al adaptarse a la escala de grises de la realidad, los conjuntos borrosos permiten resolver algunas paradojas, tomando el término en su acepción más amplia. Imaginemos que en un bar nos sirven un café realmente amargo. Salvo excepciones, para beberlo tendremos que añadirle algo de azúcar. El lector convendrá con nosotros en que, si echamos un único granito de azúcar, el sabor del café no cambiará en absoluto; por tanto, podemos concluir que la operación «añadir un granito de azúcar» no modifica la amargura del café. Echemos otro granito, luego otro, después otro más, y así hasta llegar a diez sobres de azúcar. Si nuestro principio es correcto, como ninguno de los pasos cambia el sabor, la taza de café con diez sobres de azúcar seguirá siendo igual de amarga que la que nos habían servido al principio, lo cual es, cuanto menos, inquietante... Como es fácil imaginar, lo que sucede es que los cafés dulces no son un conjunto en sentido clásico, como los que habíamos estudiado hasta el capítulo anterior. No hay un abismo, sino una gradación continua, entre ser tan amargo que no se puede beber y ser tan dulce que empalaga. Aunque nuestro paladar no sea lo bastante exquisito como para percibir el cambio, cuando se añade un grano de azúcar a la taza, su grado de pertenencia al conjunto de los cafés dulces aumenta un poco, por mínimo que sea el cambio; si echamos otro más, sigue creciendo, y es así como, tras diez sobres de azúcar, el café acaba volviéndose tan dulce que empalaga.

Cuando se trata de generalizar un concepto de las matemáticas, como pretendía Zadeh al introducir la lógica difusa, es imprescindible comprobar que el nuevo formalismo aún es válido para estudiar los objetos de partida. Los conjuntos clásicos son un ejemplo muy particular de conjuntos borrosos: exactamente aquellos en los que, entre las infinitas posibilidades de las que dispone, el grado de pertenencia sólo toma los valores 0 y 1. Sin embargo, no es tan evidente cómo generalizar el hecho de que un conjunto esté contenido en otro, o las operaciones de unión e intersección, que, según vimos allá por el capítulo 3, eran fundamentales en la teoría de conjuntos. Éstas son algunas de las preguntas a las que Zadeh dio respuesta en su artículo de 1965.

En lo que sigue, A y B serán dos conjuntos borrosos, cuyas funciones de pertenencia asociadas denotaremos f_A y f_B. Esto sólo significa que, dado un elemento x, el número $f_A(x)$, que indica el grado de pertenencia de x al conjunto A, está comprendido entre 0 y 1, y que lo mismo ocurre con $f_B(x)$. Empleando esta notación, Zadeh establece que A está contenido en B si, sea cual sea el elemento x, el número $f_A(x)$ es menor o igual que $f_B(x)$. Veamos un ejemplo: en lugar de considerar que hasta 1,60 m se es del todo bajo, y que a partir de 1,90 m, completamente alto, podríamos reducir un poco el límite, de modo que las personas de menos de 1,50 m de estatura se considerasen del todo bajas, y a partir de ahí, el grado de pertenencia fuese creciendo, como antes, hasta 1,90 m. Obtendríamos así otro conjunto borroso de personas altas. En él, el grado de pertenencia del autor ya no es de 0,5, sino de 0,625. Pues bien, lo que nos dice Zadeh es que el primer conjunto que habíamos definido está contenido en este último, lo cual encaja con la idea intuitiva de que todas las personas que eran altas con el rasero estricto lo siguen siendo una vez que se han rebajado los límites.

Al inventar la lógica difusa, Lotfi Zadeh, que había estudiado ingeniería electrónica, sospechaba que podrían encontrarse aplicaciones en el procesamiento de la información y en el reconocimiento de patrones, dos ámbitos donde la imprecisión es la nota dominante. El curso de la historia ha demostrado que Zadeh subestimó su idea, y ningún país lo iba poner tan de relieve como aquel cuyos habitantes toman las trufas de té con nata, sin nata o qué. A finales de la década de los noventa, en los comercios japoneses se habían empezado a vender fotocopiadoras y lavadoras difusas, y un rascacielos de la ciudad de Tokio no estaba a la última si no había instalado un ascensor borroso que redujese al máximo el tiempo de espera. Como anunciaba la publicidad de una de esas lavadoras: ¡la era difusa ha llegado!

LAS LAVADORAS DIFUSAS

Para mejorar la duración y la calidad de la colada, nos vendría bien recurrir a cuantificaciones precisas sobre si la ropa está muy sucia, poco sucia o prácticamente limpia. En su versión más simple, las lavadoras difusas asignan un valor de suciedad entre 0 y 1 a cada prenda. Luego a una base fija de diez minutos de lavado, se va añadiendo tiempo en función del grado de suciedad de la ropa. La máquina podría considerar, por ejemplo, que las prendas limpias (0) ya se han lavado con el tiempo de base, y que por cada prenda muy sucia (1), hacen falta dos minutos más. Entonces una camisa que esté medio limpia-medio sucia, supondrá un incremento de un minuto en la duración del lavado. Otros modelos más sofisticados incluyen asimismo medidas de la cantidad de grasa, más difícil de eliminar que otras manchas, o de la carga de la lavadora, con el fin de ahorrar energía.

La complejidad

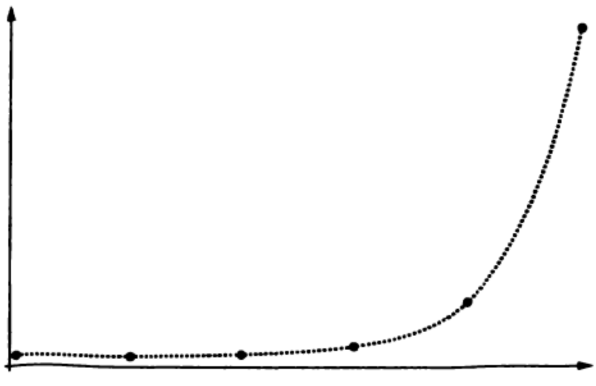
«Amor» y «justicia» son ideas demasiado sutiles para que pueda gobernarlas una lógica del sí o no. La zona de grises que se abre entre «me quiere» y «no me quiere», entre la culpabilidad y la inocencia, es el dominio de acción de la lógica difusa. A medida que aumenta la complejidad, un pensamiento nuevo es necesario. Sería útil, por tanto, disponer de estimaciones de cuán difíciles son los conceptos, pero la «complejidad» resulta ser una de esas nociones que escapan a cualquier definición. Ni siquiera en el reino seguro de las matemáticas se saben distinguir los problemas fáciles de los difíciles con total exactitud. Es el caso de las máquinas de Turing: si en el capítulo anterior el hecho de trabajar con ordenadores ideales nos permitió obtener resultados teóricos sobre los problemas que una máquina no es capaz de resolver, lo que nos interesa ahora es determinar qué cálculos pueden realizarse teniendo en cuenta las limitaciones de memoria y de tiempo de ejecución de los ordenadores de verdad. Es eso lo que convertirá, en espera de una mejor definición, un problema en fácil o difícil.

Como primer enfoque podríamos establecer que la complejidad de una tarea es el número de operaciones necesarias para llevarla a cabo. Imaginemos a un hombre de negocios que debe visitar una serie de ciudades y volver luego al punto de partida. Su objetivo es, desde luego, recorrer la mínima distancia posible. Si esas ciudades fueran, por poner un ejemplo, París (P), Londres (L), Berlín (B) y Roma (R), y el ejecutivo se encontrase en París en el momento de emprender el viaje, entonces su secretaria podría organizar la agenda de seis formas distintas: PLBRP, PLRBP, PBLRP, PBRLP, PRBLP y PRLBP. Teniendo en cuenta las distancias aproximadas París-Londres (455 km), París-Berlín (1.050 km), París-Roma (1.435 km), Londres-Berlín (1.095 km), Londres-Roma (1.855 km) y Berlín-Roma (1.515 km), podría calcular el número de kilómetros de cada ruta y elegir luego la más corta:

Ruta	Km	Ruta	Km
PLBRP	4.500	PBRLP	4.875
PLRBP	4.875	PRBLP	4.500
PBLRP	5.435	PRLBP	5.435

En vista de la tabla, el recorrido óptimo del viaje es París-Londres-Berlín-Roma-París, o bien, empezando por el final, París-Roma-Berlín-Londres-París. Para resolver el problema se han analizado uno a uno los seis casos posibles. Pero, ¿qué pasaría si en lugar de recorrer tres ciudades tuviéramos que visitar cuatro, cinco, en fin, un número cualquiera? Sólo con veinte ciudades, un ordenador medio tardaría unos ochenta mil años en encontrar la solución más rápida, lo cual hace insignificante el tiempo que podría perder nuestro hombre de negocios si no eligiera bien la ruta. Al tratar de resolver el problema «a la fuerza», ya no son seis los casos que hay que considerar, sino el número que se obtiene al multiplicar $1 \cdot 2 \cdot 3 \cdot \ldots$ así hasta 20, que tiene diecinueve cifras. Es lo que los matemáticos llamamos factorial de 20 y representamos con un signo de exclamación detrás del número. Así, $3! = 1 \cdot 2 \cdot 3$ vale 6; $4! = 1 \cdot 2 \cdot 3 \cdot 4$ es igual a 24; en general, $n!$ es el producto de los n primeros números naturales.

El factorial es un ejemplo de función muy simple de calcular desde un punto de vista teórico, pero que en la práctica hace enloquecer a los ordenadores. Como señalamos de pasada en el capítulo anterior, todas las funciones recursivas son computables. Recordemos que una función es recursiva si $f(n)$ puede calcularse a partir de los valores que toma la función en otros números menores que n. Pues bien, el factorial es el caso típico de función recursiva, ya que si queremos calcular $4! = 1 \cdot 2 \cdot 3 \cdot 4$, podemos hacer primero el producto $1 \cdot 2 \cdot 3$ y multiplicar luego por 4. Ahora bien, ¿qué es el producto $1 \cdot 2 \cdot 3$? Precisamente 3!, de modo que sabiendo cuánto vale 3!, basta una operación para obtener el factorial de 4. En general, $n! = (n-1)! \cdot n$, y eso demuestra que el factorial es una función recursiva, luego computable. Por tanto, para una máquina de Turing con toda la eternidad por delante, el cálculo de $n!$ no tiene ningún misterio. Sin embargo, en la práctica los valores de la función crecen tan deprisa que pronto se vuelven intratables, como se aprecia en la siguiente gráfica:

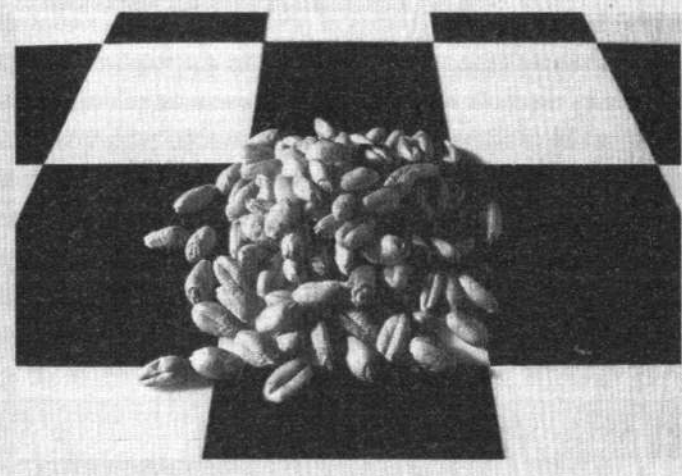


Gráfica que muestra el crecimiento de la función factorial.

El ejemplo anterior no pasaría de ser una curiosidad si no fuese porque el factorial cuenta el número de permutaciones de los conjuntos finitos, es decir, de cuántas formas distintas pueden ordenarse sus elementos. Así, las frases «3! = 6» y «el conjunto {1, 2, 3} se puede escribir de seis maneras diferentes (123, 132, 213, 231, 312 y 321)» expresan la misma información. Como el análisis ingenuo de muchos problemas similares al del ejecutivo requiere examinar una por una todas las permutaciones de un conjunto que podría estar formado por muchos elementos, la rapidez con la que crece el factorial tiene consecuencias mortales para la informática. Una primera medida de la complejidad se obtiene llamando difíciles a los proble-

EL INVENTOR DEL AJEDREZ

Cuenta la leyenda que un rey persa quiso premiar al inventor del ajedrez ofreciéndole aquello que más deseara, sin importarle el precio. Entonces el sabio lo sorprendió con una petición en apariencia muy humilde: quería un grano de trigo por la primera casilla del tablero, dos por la segunda, cuatro por la tercera, y en general, el doble de lo que había recibido por la anterior hasta llegar a la última. Enojado por lo que le parecía una burla a su benevolencia, el rey mandó a sus súbditos que cumpliesen la orden de inmediato y despachó al inventor del ajedrez. Cuál no sería su sorpresa cuando uno de sus consejeros le comunicó, al día siguiente, que no había trigo en todos los graneros del mundo para satisfacer la petición del sabio. La función que comenzaba tomando los valores 1, 2, 4, 8... crecía tan deprisa que en total hacían falta 18.446.744.073.709.551.615 granos de trigo.



mas que requieran un número de operaciones aproximadamente de este tamaño. Los fáciles serán, por el contrario, aquellos que no sólo sean resolubles desde un punto de vista teórico, sino también en la práctica, en tiempo razonable. A menudo se los representa con la letra P (P de polinomio), porque el número de operaciones necesarias aumenta en función del tamaño de los datos más o menos al mismo ritmo que los polinomios.

Lo que observaron los informáticos es que hay problemas en los que encontrar la solución es muy difícil, mientras que comprobar que es la correcta resulta sencillo. Volvamos a nuestro hotel del capítulo 2, que esta vez será finito: supongamos que un grupo de cuatrocientas personas quiere pasar allí la noche en unas fechas en las que sólo hay cien habitaciones libres. Elegirlas sin atender a ningún criterio sería muy fácil, pero el formulario de reserva iba acompañado de una extraña petición: hay una serie de personas que se llevan tan mal que de ningún modo podrían dormir en habitaciones contiguas. Es impensable resolver el problema examinando una a una todas las elecciones posibles de cien personas entre cuatrocientas, y sin embargo, una vez propuesta una solución, basta con comprobar que dos personas incompatibles no aparecen juntas en la lista de las habitaciones asignadas. Sin necesidad de un ordenador, el mismo recepcionista podría hacerlo en unas horas. A estos problemas que son difíciles de resolver, pero fáciles de comprobar, los matemáticos los llamamos NP.

Hasta ahora nos hemos referido a la complejidad de los problemas como si se tratase de una propiedad intrínseca al enunciado. Este punto de vista es a priori erróneo, porque lo fácil o lo difícil no es el problema en sí, sino nuestro modo de resolverlo. Quizás hayamos encontrado una solución que requiere muchísimas operaciones, pero exista otra más simple; en ese caso, nuestra solución estaría en NP, mientras que el problema pertenecería a P. La estrategia para que el hombre de negocios optimizara sus viajes consistió en examinar todos los recorridos posibles uno a uno, sin hacer ningún razonamiento. La tabla muestra, sin embargo, que al invertir el orden de la ruta, la distancia no varía. Lo mismo da escoger París-Londres-Berlín-Roma-París que París-Roma-Berlín-Londres-París, luego sería suficiente con tratar la mitad de los casos. En la práctica, esta simplificación no repre-

P VERSUS NP

Como vimos en el capítulo 3, el comienzo simbólico de las matemáticas del siglo xx se había escenificado en agosto de 1900, en París, con la lista de los veintitrés problemas de Hilbert. También en París, pero cien años después, una comisión de expertos del Instituto Clay de Matemáticas se reunía para escoger las siete cuestiones abiertas que, a su juicio, marcarían la investigación del nuevo siglo. El cuarto problema de la lista, conocido como «P versus NP», consiste precisamente en averiguar si hay problemas NP en sentido absoluto o si, por el contrario, cualquier problema cuyas soluciones se pueden comprobar en tiempo polinomial también se puede resolver rápidamente, una vez que se ha encontrado el algoritmo ingenioso. Un millón de dólares espera a quien sepa contestar a la pregunta. Después de todo, las matemáticas quizá sí sean rentables.

senta ningún avance, pues la mitad de un número enorme sigue siendo enorme. Su importancia es más bien filosófica: si en la primera solución pasamos por alto un detalle tan trivial, ¿cuántos otros trucos del mismo tipo no habremos tenido en cuenta? Dijimos que el punto de vista era a priori erróneo, porque lo cierto es que ignoramos si existen problemas difíciles en sentido absoluto. Aunque el problema del ejecutivo es uno de los candidatos, nadie ha conseguido demostrar aún que todas sus soluciones sean difíciles.

Otra de las objeciones que plantea este concepto de complejidad es que no ayuda a distinguir entre tareas que involucren la misma cantidad de operaciones. Según nuestra definición, memorizar una contraseña de doce símbolos es un problema fácil o difícil con independencia de cuáles sean los caracteres, pues siempre harán falta doce operaciones: memorizar el primero, memorizar el segundo, así hasta el duodécimo. Sin embargo, nadie en su sano juicio consideraría que aprenderse de memoria las contraseñas 111111111111 y 6u0yfz3eq85s requiere el mismo esfuerzo. Mientras que la primera puede comprimirse en «doce 1s», el único modo de describir la segunda es término a término. Con este ejemplo en mente, el matemático ruso Andrei Kolmogorov propuso a mediados de la década de los sesenta sustituir el número de operaciones por el de instrucciones. La complejidad de una cadena de símbolos sería, desde entonces, la longitud mínima del algoritmo necesario para generarla.

Imaginemos una máquina de Turing cuya tarea consiste en escribir una cierta cadena de 0s y de 1s, que vamos a llamar *s*. Como vimos en el capítulo anterior, a la máquina habrá que darle una serie de instrucciones del estilo «Si se lee un 1, moverse a la derecha e ir a la orden #2». En esta versión simplificada, diremos que la complejidad de *s* es un número natural *n* si existe una máquina de Turing descrita por medio de *n* instrucciones cuyo *output* es *s*, y si ninguna máquina con menos órdenes puede generar nuestra secuencia. Se obtiene así una función *K* (K de Kolmogorov) que asocia a cada cadena de 0s y 1s su complejidad. Pensemos, por ejemplo, en la sucesión 1111... El lector puede comprobar que, si se introduce como *input* una cinta de papel llena de 0s a la máquina de Turing cuya única instrucción es «Instrucción #1: Si se lee un 0, escribir un 1 e ir a la instrucción #1. Si se lee un 1, moverse hacia la derecha e ir a la instrucción #1», entonces obtendremos como resultado la sucesión 1111... Esto significa que su complejidad es la mínima posible, $K(s) = 1$, porque una sola instrucción basta.

Una consecuencia sorprendente del nuevo concepto de complejidad es que los ordenadores no saben generar cadenas aleatorias infinitas de 0s y 1s. De modo intuitivo, una sucesión es aleatoria cuando nos resulta imposible, por su estructura interna, predecir cuál será el siguiente término. Eso significa que no se puede comprimir una secuencia aleatoria en una descripción más corta que ella misma; en otras palabras, su complejidad es infinita. Sin embargo, todos los programas informáticos funcionan con un número finito de instrucciones (recordemos la definición de máquina de Turing del capítulo anterior). Por tanto, las cadenas de 0s y 1s que generen, por muy endiabladas que parezcan, serán siempre de complejidad finita. Los ordenadores sólo pueden escribir sucesiones *pseudoaleatorias*, y ésta es la razón por la que muchos físicos intentan desde hace algunos años aprovechar las propiedades indeterministas de los átomos para construir secuencias aleatorias de verdad.

Por otra parte, la complejidad de Kolmogorov guarda muchas similitudes con la paradoja del bibliotecario de Oxford que explicamos al final del capítulo 2, y que consistía en estudiar el conjunto de los números naturales que pueden describirse por medio de quince palabras. Como sólo hay una cantidad finita de expresiones de quince palabras, también este conjunto es finito. Por tanto, de entre todos los números que no pertenecen a él, habrá uno que sea el más pequeño, llamémoslo n. Entonces n es «el menor número que no podemos describir con menos de quince palabras», ¡pero esta expresión tiene doce palabras! Es natural preguntarse si la definición que acabamos de introducir no nos conducirá a contradicciones, y la respuesta vuelve a ser sorprendente: si la función K fuese computable, es decir, si existiera una máquina de Turing que, al recibir como *input* una cadena s de 0s y 1s, devolviera como *output* la complejidad $K(s)$, entonces un razonamiento similar al del problema de la parada nos permitiría reproducir la paradoja del bibliotecario en el lenguaje formal de la aritmética. Por tanto, la única solución posible es que la complejidad no sea computable, y eso basta para resolver la paradoja del bibliotecario, que había quedado pendiente: ocurre que la expresión «describir por medio de quince palabras» no es correcta porque no pertenece al lenguaje, sino al metalenguaje.

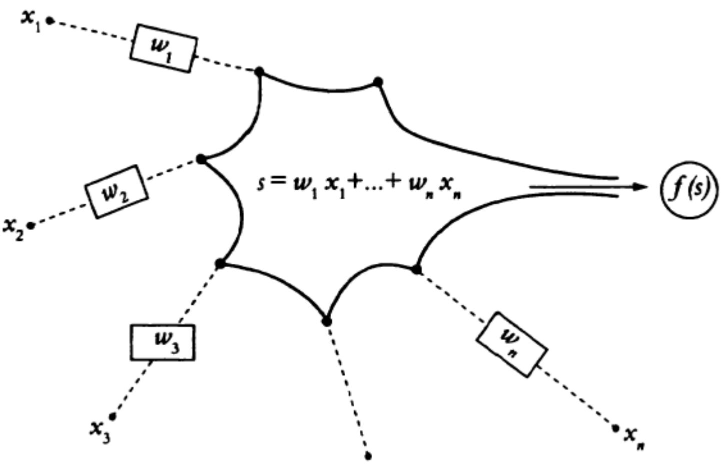
Gödel, Turing y la inteligencia artificial

En las páginas anteriores nos hemos contentado con discutir el concepto de complejidad sólo en el ámbito de las matemáticas, donde el lector habrá podido comprobar que las dificultades son numerosas. Nuestro propósito inicial era aún más ambicioso: queríamos saber qué significa que las ideas de «amor» y «justicia» sean complejas. Poco a poco, los resultados matemáticos han servido de inspiración a otra teoría de la complejidad, que podría resumirse en el emblema «el todo es mayor que la suma de las partes». Las palabras «lucidez», «herida», «sol» y «próximo» tienen cada una un significado muy preciso. Podríamos consultar las definiciones de varios diccionarios y aprender su etimología. Sin embargo, cuando el poeta francés René Char escribe que «La lucidez es la herida más próxima al sol», de cuatro palabras que conocíamos perfectamente ha emergido algo que no estaba en ellas. El verso es mayor que la suma de las palabras, y por eso la poesía es difícil de entender.

Lejos de ser un fenómeno exclusivo del lenguaje, este principio de emergencia recorre el mundo en una increíble variedad de formas: está presente en los llamados insectos sociales, sirve para explicar el éxito de Internet y es una de las claves para el estudio de los sistemas nerviosos de los seres vivos. Pensemos, por ejemplo, en una humilde hormiga, que atiende como puede los imperativos genéticos de buscar comida. Nunca podríamos comprender la compleja organización de un hormiguero, capaz de adaptarse a las situaciones más extremas, viéndolo únicamente como una suma de hormigas. También el sistema inmunitario es más que una colección de células; la economía, más que el conjunto de compradores de acciones, e Internet, más que la suma de las intervenciones aisladas de cada usuario del planeta. Entender de qué modo partiendo de la relativa sencillez de cada componente de cualquiera de estos sistemas emerge la complejidad del todo es uno de los grandes desafíos de la ciencia de comienzos de este siglo.

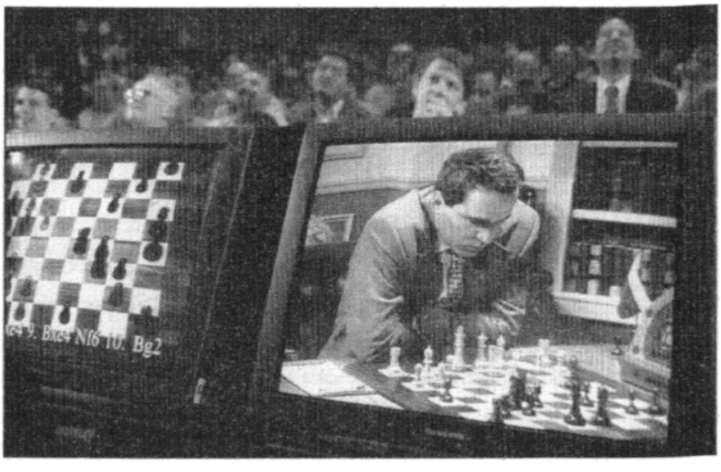
Aunque la definición de sistema complejo como aquel en el que el todo es mayor que la suma de las partes sea demasiado imprecisa, no cabe duda de que el cerebro se adapta muy bien a ella. En este caso, los componentes individuales son las neuronas, unas células formadas por un cuerpo central que recibe, procesa y transmite a otras neuronas los impulsos que le llegan a través de una serie de ramificaciones. La opinión más extendida entre los estudiosos del cerebro es que la red de conexiones que lo convierten en algo más que un conjunto de neuronas aisladas está detrás de fenómenos globales como la percepción, la inteligencia o los sentimientos. ¿Y si pudiéramos traducir esta estructura a la informática? Los primeros intentos de modelizar matemáticamente las neuronas se remontan a un artículo de 1943, en el que el neurólogo Walter McCulloch y el lógico Walter Pitts definían con elegancia una neurona como una función que, partiendo de varios *inputs*, producía un solo *output*.

Hasta el momento, todas las funciones que se han considerado en este libro tomaban un único valor inicial y lo transformaban en otro mediante una serie de operaciones. Sin embargo, en la vida real pocos fenómenos, por no decir ninguno, dependen de un solo parámetro. La moderna teoría de las redes neuronales artificiales, inspirada en las ideas de Pitts y McCulloch, permite computar funciones de muchos parámetros imitando el funcionamiento del cerebro. Supongamos que se quiere calcular el valor de una función f que depende de los números $x_1, x_2, \ldots, x_n$. La idea es que el programa los reciba como si se tratase de los impulsos eléctricos que llegan al cuerpo central de una neurona a través de sus ramificaciones. Puesto que no todos ellos tienen necesariamente la misma intensidad, habrá que acompañar cada número x_i con otro número w_i, llamado peso, que mida la importancia de cada impulso eléctrico con relación a los demás. Por ejemplo, si w_1 y w_n fueran mucho mayores que $w_2, w_3 \ldots, w_{n-1}$, eso significaría que los parámetros que de verdad influyen son el primero y el último. Con los pesos de los impulsos a su disposición, la neurona artificial calcula la suma ponderada $s = w_1x_1 + w_2x_2 + \ldots + w_nx_n$ y evalúa la función en ella, tal como se detalla en la ilustración:



La novedad de las redes neuronales reside en que el programa con el que pretendemos resolver nuestro problema ya no es un algoritmo fijo, sino una *obra abierta* en la que los pesos pueden ir cambiando. De hecho, es costumbre someter a la red neuronal a una fase de entrenamiento en la que el programa va «aprendiendo», por ensayo y error, cuáles son los pesos más adecuados, o dicho de otro modo, qué *inputs* hay que privilegiar para que la solución sea satisfactoria. Si nuestra red neuronal se encarga, por ejemplo, de reconocer la voz humana y una de las conclusiones de la fase de entrenamiento es que la mayoría de lo que se recibe por el primer impulso es ruido de fondo, entonces las neuronas aprenderán a darle muy poca importancia. Otras tareas para las que las redes neuronales se han revelado muy efectivas son las previsiones meteorológicas, e incluso el problema del ejecutivo. Los ordenadores que incorporan ésta y otras técnicas más avanzadas saben resolverlo ya para doscientas ciudades.

Gracias a la lógica difusa y a las redes neuronales, la posibilidad de que los ordenadores imiten muchas de las actividades de la mente humana ha abandonado las especulaciones de la ciencia ficción para convertirse en el primer objetivo de una disciplina en auge: la inteligencia artificial. Durante años se creyó que una máquina nunca podría jugar al ajedrez como los grandes maestros: por muchas jugadas que fuese capaz de prever, no conocería los puntos débiles del adversario ni sabría considerar otros factores psicológicos. Tampoco se le darían bien los juegos de apuestas: ¿cómo entrenar a un ordenador para jugar al póquer si un farol contradice las probabilidades de victoria? Las voces críticas tuvieron que retractarse cuando, en febrero de 1996, la supercomputadora Deep Blue, que la compañía IBM había desarrollado culminando esfuerzos que se remontaban a la década de 1950, consiguió batir a Garry Kasparov en la primera ronda de un torneo de ajedrez. A pesar de los cien millones de posiciones que Deep Blue analizaba por segundo, de las cinco partidas siguientes, jugadas a ritmo más lento de lo habitual, cuatro las ganó el ruso. Pero un año después el equipo había mejorado la máquina, hasta el punto de que Deep Blue ganó tres de las partidas y consiguió hacer tablas en la cuarta, jugando a la misma



Garry Kasparov estudia su siguiente movimiento durante la partida que jugó contra la supercomputadora Deep Blue el 10 de mayo de 1997.

velocidad que los profesionales. El campeón del mundo había sido derrotado, aunque este duro golpe no impidió que Kasparov siguiera defendiendo la supremacía de la inteligencia humana, curiosamente con el mismo argumento que habían utilizado sus competidores para programar Deep Blue: «es la síntesis, la capacidad de combinar creatividad y cálculo, arte y ciencia, en un todo que es más que la suma de las partes».

UN DIÁLOGO DE LA PELÍCULA *YO, ROBOT* (ALEX PROYAS/JEFF VINTAR/ISAAC ASIMOV, 2004)

El protagonista del filme, un policía llamado Spooner, investiga un asesinato del que acusa al robot Sonny.

Spooner: El asesinato es una habilidad nueva para un robot, enhorabuena. ¡Responde!

Sonny: ¿Qué significa esta acción? (Guiña un ojo). Antes de entrar, cuando miró al otro humano. ¿Qué significa? (Vuelve a guiñar un ojo).

Spooner: Significa confianza, es algo humano, no lo entenderás.

Sonny: Mi padre intentó enseñarme las emociones. Son muy... difíciles.

Spooner: Te refieres a tu diseñador.

Sonny: Sí.

Spooner: ¿Por qué lo mataste?

Sonny: Yo no maté al doctor Lanning.

Spooner: Entonces, ¿por qué estabas allí escondido?

Sonny: Estaba asustado.

Spooner: Los robots no sienten miedo. No sienten nada: no tienen hambre, no duermen...

Sonny: Yo sí. Incluso he tenido sueños.

Spooner: Los seres humanos tienen sueños. Los perros también, pero tú no. Sólo eres una máquina. Una imitación de la vida. ¿Puedes componer una sinfonía? ¿Puedes convertir un lienzo en una hermosa obra de arte?

Sonny: ¿Puede usted?

Estos progresos no hacían sino avivar apasionantes controversias científicas y filosóficas que habían enfrentado a Kurt Gödel y a Alan Turing cincuenta años antes. Sirviéndose de diferentes métodos, los dos matemáticos habían coincidido en la definición de sistema formal y a la hora de exhibir problemas indecidibles. Sin embargo, mientras que Gödel distinguía formalismo y lógica, mecanismo y mente, Turing los consideraba totalmente sinónimos. Llevando al extremo esta equiparación, en 1947 el lógico inglés postulaba que el mejor modelo de la mente humana era la máquina universal —capaz de imitar el comportamiento de cualquier programa— que él mismo había introducido con vistas a resolver el problema de la decisión de Hilbert. Turing consideraba que la pregunta de si pueden pensar los ordenadores sólo podría resolverse de modo experimental. En un artículo llamado a hacer historia, «Máquinas de computación e inteligencia», Turing proponía en 1950 un «juego de imitación» para que los científicos averiguasen, a través de una batería de preguntas comunicadas por escrito, si lo que había al otro lado de la sala era un ser humano o un ordenador. La filosofía del test era que si una máquina se comportaba, en todos los aspectos, como un ser inteligente, entonces la explicación más simple es que fuera un ser inteligente.

Entre otras cuestiones, Turing sugería que se le pidiera al aspirante a ser inteligente que escribiera un poema o que realizara difíciles cálculos numéricos. En principio, soluciones correctas a la primera pregunta inclinarían la balanza hacia los seres humanos, mientras que una respuesta rápida a la segunda nos haría pensar en un ordenador. Alguien podría objetar que muchas personas no son capaces de escribir un poema o que, si por casualidad diésemos con un poeta vanguardista, resultaría difícil distinguir la composición de unos versos generados aleatoriamente. Del mismo modo, hay auténticas «calculadoras humanas», capaces de multiplicar o de factorizar números muy grandes tan rápido como los ordenadores. A pesar de estas dificultades, todo el mundo está de acuerdo en que, con un número ilimitado de preguntas a nuestra disposición, deberíamos poder distinguir a un ser humano de una máquina. Por ahora, no sólo ningún ordenador ha conseguido superar el test de Turing, sino que, además, éste se emplea en el reconocimiento del correo electrónico basura, por lo general enviado de forma masiva por ordenadores.

En diciembre de 1969, quince años después de la muerte de Turing, Gödel creyó haber descubierto un error con graves consecuencias filosóficas en la obra del lógico inglés. A su juicio, Turing no había tenido en cuenta que la mente no es estática, sino que está en constante desarrollo. En el transcurso de una demostración, los sistemas formales no sufren modificaciones, ni tampoco las máquinas durante un cálculo, pero nada permite asegurar que la mente, que está viva, no cambie al hacer razonamientos. Por tanto, jamás podrá ser reemplazada por un ordenador. Quien abra un libro contra la inteligencia artificial es probable que se encuentre antes o después con una sección dedicada a los argumentos gödelianos, pero éstos no se refieren a la objeción que acabamos de plantear, sino a la idea del filósofo oxoniense John Lucas de que los teoremas de incompletitud tienen algo que decir sobre la posibilidad de que algún día existan máquinas inteligentes. Curiosamente, Gödel nunca pensó demasiado en serio que sus resultados guardasen relación con la estructura de la mente humana.

El más famoso argumento gödeliano contra la inteligencia artificial se debe, como decimos, al filósofo John R. Lucas, que había estudiado matemáticas antes de dedicarse a la filosofía y a la historia antiguas. En el artículo «Mentes, máquinas y Gödel», presentado en 1959 a la Oxford Philosophical Society, Lucas exponía con simplicidad rotunda por qué la mente es irreducible a los ordenadores: puesto que somos capaces de enseñar a una máquina los axiomas y las reglas deductivas de la aritmética, podríamos dejarlo construyendo todas las fórmulas del lenguaje y preguntarle cuáles son verdaderas. Antes o después, el ordenador daría con la sentencia «Esta frase no es demostrable» y se pasaría el resto de la eternidad tratando de demostrarla o de refutarla, mientras que nosotros, los humanos, vemos enseguida que la frase es indecidible por su contenido mismo. «Luego la máquina seguirá sin ser un modelo adecuado de la mente [...] siempre un paso por delante de cualquier sistema formal, osificado, muerto», concluía Lucas.

Pasado medio siglo, casi nadie acepta el argumento de John Lucas, ni tampoco la versión refinada que propuso el físico Roger Penrose en 1989. ¿Qué significa que las personas *vemos* la verdad del enunciado de Gödel? Lo que dice el primer teorema de incompletitud es que, si la aritmética es consistente, entonces la proposición «Esta frase no es demostrable» es verdadera, luego para *ver* su verdad tenemos que *ver* primero la consistencia de la aritmética. Si la aceptamos como un acto de fe, porque creemos en un mundo libre de contradicciones, también podríamos programar un androide cuyo código informático incluyera la esperanza de que la aritmética fuera consistente. Esto no es más que una reinterpretación del segundo teorema de incompletitud, que afirma que la consistencia de la aritmética no puede demostrarse dentro de su propio sistema formal. Sin embargo —respondería Lucas—, los matemáticos son capaces de probar que la aritmética es consistente recurriendo a técnicas más avanzadas, a lenguajes de orden superior. Es cierto que ese «salirse del sistema» del que nosotros somos capaces parece difícilmente accesible a la máquina, pero ¿y si consiguiera aprenderlo? ¿Si de una red de neuronas artificiales muy compleja emergiesen nuevas visiones de la consistencia? Nada es tan fácil como parece.

¿Qué pensaría el viejo Euclides de las bifurcaciones del método axiomático? Disfrazarlo de lógico borroso del siglo XXI sería un buen final para esta novela que se inició con el descubrimiento de las geometrías no euclideas, continuó con la teoría de conjuntos y sus paradojas y vio nacer en los capítulos siguientes a tres héroes indiscutibles: David Hilbert, Kurt Gödel y Alan Turing. Sería un buen final, pero la investigación va por delante. En los pocos meses que tardarán estas líneas en llegar a los primeros lectores, matemáticos, físicos e ingenieros habrán seguido perfeccionando las redes neuronales, quizá la lógica difusa haya tomado nuevos rumbos y es posible, incluso, que alguien haya avanzado hacia la solución del problema «P versus NP». Mejor será dejarlo así. *Bien acaba lo que no acaba* no es mal final para un libro cuyo protagonista son las paradojas.

Bibliografía

BERTO, F., *Tutti pazzi per Gödel!*, Roma, Laterza, 2007.

EUCLIDES, *Elementos*, Madrid, Gredos, 2000.

FRESÁN, J., *Gödel. La lógica de los escépticos*, Madrid, Nivola, 2007.

HEIJENOORT, J.V., *From Frege to Gödel: A Source Book in Mathematical Logic,* Cambridge (Massachussets), Harvard University Press, 1967.

HOFSTADTER, D.R., *Gödel, Escher, Bach. Un eterno y grácil bucle*, Barcelona, Tusquets, 1987.

MARTÍNEZ, G. y PIÑEIRO, G., *Gödel ∀ (para todos)*, Barcelona, Destino, 2010.

MITCHELL, M., *Complexity*, Oxford, Oxford University Press, 2009.

MOSTERÍN, J., *Los lógicos*, Madrid, Espasa, 2000.

NAGEL, E. y NEWMAN, J.R., *El teorema de Gödel*, Madrid, Tecnos, 1970.

SANGALLI, A., *The Importance of Being Fuzzy and Other Insights from the Border between Math and Computers*, Princeton, Princeton University Press, 1998.

SMITH, P., *An Introduction to Gödel's Theorems*, Cambridge, Cambridge University Press, 2007

SOKAL, A. y BRICMONT, J., *Imposturas intelectuales*, Barcelona, Paidós, 1999.

Índice analítico