más sincero y preciso de la falacia ha procedido de un autor soviético: «No es la máquina creada por el hombre sino el propio hombre quien es la máxima manifestación de cultura, pues los pensamientos y los sueños, los amores y las aspiraciones del hombre, *el creador*, son a la vez complejos y grandes»[151].

Antropólogos e historiadores hace tiempo que han pensado que la introducción de cualquier innovación económica en una colectividad tiene éxito sólo si la colectividad puede adaptarse culturalmente a ella, esto es, sólo si la innovación llega a ser socialmente aprobada y entendida[152]. Entre los economistas angloamericanos, hubo una época en la que sólo un rebelde como Veblen afirmó que es peligroso colocar las máquinas modernas en manos de gentes que tienen todavía una *Anschauung* económica feudal[153]. Sin duda alguna, «peligroso» difícilmente es el término adecuado aquí, pero probablemente Veblen quería subrayar tanto la inmensa pérdida económica como los grandes males sociales derivados de una forzada introducción de industrias modernas en una colectividad carente de las correspondientes propensiones[154]. Pero seamos sinceros al respecto: ¿quién puede negar que el peligro creado por el descubrimiento de la energía atómica se deriva del atraso cultural de la humanidad con respecto a la nueva tecnología? Todas las culturas han ido siempre a la zaga del progreso tecnológico de su época, unas más, otras menos; pero el atraso, ya sea de la humanidad como un todo o de partes de ella, nunca ha sido tan grande como en la actualidad.

La cuestión tiene evidentes implicaciones para toda política dirigida a acelerar la tasa de crecimiento de una economía. Esas implicaciones han sido esporádicamente reconocidas, principalmente por economistas «no ortodoxos». Leonard Doob, por ejemplo, insistió en que ninguna planificación puede tener éxito a no ser que se base en un conocimiento del entorno social, es decir, de la tradición de las gentes que se verán afectadas por ella. Una tesis todavía más potente se ha postulado por J. J. Spengler, quien afirma que la tasa de crecimiento económico depende del grado de compatibilidad existente entre los componentes económicos y los no económicos de la cultura respectiva[155]. No habría que desechar sin más esas observaciones, pues todos los análisis de por qué los resultados de nuestra ayuda económica exterior no han sido con frecuencia proporcionales a su esencia convergen en una explicación: las costumbres locales.

[151] S. T. Konenkov, «Communism and Culture», *Kommunist*, núm. 7, 1959. Versión inglesa en *Soviet Highlights*, núm. 3, I (1959), pp. 3-5. Las cursivas son mías.

[152] G. Sorel, en la «Introduction» a G. Gatti, *Le socialisme et l'agriculture* (París, 1902), p. 8; Richard Thurnwald, *Economics in Primitive Communities* (Londres, 1932), p. 34; V. Gordon Childe, *Social Evolution* (Nueva York, 1951), p. 33.

[153] Thorstein Veblen, *Imperial Germany and the Industrial Revolution* (Nueva York, 1964), pp. 64-66, y *Essays in Our Changing Order*, ed. L. Ardzrooni (Nueva York, 1934), pp. 251 y s.

[154] O, como llegó a expresarlo P. N. Rosenstein-Rodan en «Problems of Industrialization of Eastern and South-Eastern Europe», *Economic Journal*, LIII (1943), p. 204, «Un marco institucional diferente del actual es evidentemente necesario para llevar a cabo con éxito la industrialización en las áreas internacionales deprimidas».

En realidad, hay unos cuantos hechos que sugieren que la influencia de la *Anschauung* económica sobre el proceso económico es mucho más profunda de lo que sospecharon los autores antes citados. Mencionaré únicamente los más convincentes. La Rusia soviética, en un momento en el que apenas había introducido ninguna innovación salvo la planificación central, sintió la necesidad de actuar sobre la *Anschauung* económica de las masas: «El propósito del trabajo políticamente educativo [en los campos de trabajos forzados] es el de erradicar de los obreros convictos los viejos hábitos y tradiciones nacidos de las condiciones imperantes en los modos de vida de épocas anteriores»[156]. Por muy intensa que fuese la presión ejercida sobre el pueblo de la URSS a través de numerosas obras educativas semejantes, el resultado fue tal que, en el Vigésimoprimer Congreso del PCUS, Nikita Jruschev tuvo que seguir anunciando: «Para alcanzar el comunismo... hemos de criar el hombre del futuro precisamente ahora»[157].

Un caso mucho más familiar es el gran milagro económico de Japón. No tengo duda alguna de que sólo la peculiar *Anschauung* económica del japonés medio puede explicar ese milagro, porque, estoy seguro, ningún experto en planificación podría trazar un plan económico para llevar a una economía desde las condiciones dominantes en el Japón de 1880 a las actualmente existentes; y, en caso de poder, debe haber sabido de antemano que el pueblo eran los japoneses y haberse dado cuenta igualmente de que los datos completos de cualquier problema económico deben incluir también las propensiones culturales.

Nada está más lejos de mi mente que negar las dificultades que plantea la manera de estudiar la *Anschauung* económica de una sociedad en la que uno no se ha criado culturalmente, ni estoy preparado para anotar toda una serie de instrucciones acerca de cómo hacerlo mecánicamente. Ahora bien, si negamos la facultad empática del hombre, entonces no existe en verdad juego alguno al que podamos jugar, ya sea en filosofía, literatura, ciencia o familia. En realidad, hemos de reconocer que el juego no es el mismo en las ciencias físicas que en las ciencias del hombre; que, contra-

[155] Leonard Doob, *The Plans of Men* (New Haven, 1940), pp. 6 y s.; J. J. Spengler, «Theories of Socio-Economic Growth», en *Problems in the Study of Economic Growth*, National Bureau of Economic Research (Nueva York, 1949), p. 93. Véase también K. Mannheim, «Present Trends in the Building of Society», en *Human Affairs*, ed. R. B. Cattell *et al.* (Londres, 1937), pp. 278-300.

[156] Resolución del Congreso Panruso de Obreros de la Judicatura, de 1931, en *Report of the Ad Hoc Committee on Forced Labor*, Naciones Unidas, OIT, Ginebra, 1953, pp. 475 y s.

[157] Citado en Konenkov, «Communism and Culture».

riamente a lo que predicaron Pareto y otros muchos, no existe un único método para conocer la verdad[158].

En la física únicamente podemos confiar en el instrumento de lectura puntual, porque no estamos dentro de la materia; y, sin embargo, tiene que haber un hombre en el otro extremo del instrumento para leerlo, comparar las lecturas y analizarlas. La idea de que no se puede confiar en el hombre como instrumento en el proceso cognoscitivo es, por eso, tanto más incomprensible. Curiosamente, los físicos son conscientes de su hándicap, es decir, del hecho de que no pueden interrogar a la Naturaleza: lo único que pueden hacer es observar el *comportamiento* de la materia. Como lo ha señalado un gran físico tras otro, el estudioso del hombre tiene medios adicionales a su disposición; puede tener sensaciones dentro de los hechos, o recurrir a la introspección, o, sobre todo, descubrir los motivos de su objeto de estudio interrogándole[159]. Si *per absurdum* un físico pudiese conversar con los electrones, ¿rehusaría preguntarles: por qué saltáis? Ciertamente, no. Sin embargo, el paralelismo físico ha sido exagerado por algunos científicos sociales hasta el punto de que, puesto que no podemos conversar con la materia inerte, tampoco deberíamos conversar con la gente. Existe una razón fundamental para que los físicos abracen el conductismo puro, pero el conductismo puro no tiene sitio en las ciencias del hombre. Como observó F. A. Hayek en su espléndida denuncia del dogma conductista en las ciencias sociales, «cuando hablamos del hombre, implicamos necesariamente la presencia de ciertas categorías mentales familiares»[160], es decir, las mismas categorías mentales que las poseídas por el que habla. Hasta los físicos creyeron necesario recordar a los científicos sociales que habían decidido ignorar la esencia de su objeto de estudio que «el principal problema para comprender las acciones de los hombres es comprender cómo piensan, cómo trabajan sus mentes»[161]. Y, como he afirmado en muchos sitios de este libro, ningún electrodo, ningún microscopio, ningún aparato físico puede revelarnos cómo trabajan las mentes de los hombres; únicamente la mente de un hombre puede descubrir cómo trabaja la mente de otro hombre, utilizando el puente proporcionado por las categorías mentales familiares y las propensiones comunes a ambos. El hombre no puede ser un instrumento tan preciso como un microscopio, pero es el único que puede observar lo que no pueden todos los instrumentos físicos juntos; porque, si no fuera así, enviaríamos algunos politoscopios a desvelar qué es lo que otras gentes piensan, sienten y pueden hacer después, y no a embajadores, consejeros, periodistas y otras clases de observadores; y, como todavía no tenemos politoscopios, no deberíamos enviar a nadie.

Pero tal vez un día lleguemos a darnos cuenta de que el hombre es también un instrumento, el único para estudiar las propensiones del hombre. Ese día no habrá más hombres olvidados, olvidados porque en la actualidad se supone que no sabemos cómo estudiarlos y dar cuenta de lo que piensan, sienten y quieren. Un «ejército de paz», no sólo un «cuerpo de paz», es lo que necesitamos. Admito que esto puede ser un pensamiento utópico, con reminiscencias del eslogan de los *Narodniki*: «Para el pueblo». Pero prefiero ser utópico sobre este punto a serlo respecto de la Nueva Jerusalen que el cientifismo acrítico de uno u otro tipo ofrece como promesa al hombre.

[158] Pareto, *Manuel*, p. 27.

[159] Por ejemplo, Planck, *ibid.*, p. 105; Bohr, *ibid.*, p. 78; H. Margenau, *Open Vistas: Philosophical Perspectives of Modern Science* (New Haven, 1961), p. 198.

[160] F. A. Hayek, *The Counter-Revolution of Science* (Glencoe, Ill., 1952), p. 79.

[161] Bridgman, *Reflections*, p. 450.

APÉNDICE A

SOBRE LA CONSISTENCIA DEL CONTINUO ARITMÉTICO

1. Confío en que el hecho de que haya limitado a los números naturales (en el Capítulo III, Sección 2) mi metáfora de las diferentes cuentas de una sarta sin tal sarta no dará lugar a ningún tipo de recelo. Así pues, resta por ver si es posible que alguna circunstancia en la progresiva construcción del continuo aritmético a partir de números enteros ordenados pero no ensartados puede hacer que la metáfora sea ilícita. En este apéndice, voy a examinar este problema desde un ángulo tan amplio como parece justificar su relación íntima con el tema desarrollado en el Capítulo III.

En la actualidad, el hecho de que exista una gran desfase entre un número entero y su sucesor parece el colmo de lo obvio. Sin embargo, podemos imaginar perfectamente cuentas enteras ensartadas entre sí de tal manera que cada una de ellas toque a sus vecinas. ¿Tendríamos que decir entonces que el conjunto de números enteros es un agregado continuo sin vacío alguno entre sus elementos? Posiblemente se pueda haber pensado alguna vez de este modo, aun cuando no en esos términos sofisticados.

Sea como fuere, a fin de hacer más clara mi argumentación, vamos a suponer que las cuentas enteras están ordenadas de ese modo. Al crecer la necesidad de números (racionales) fraccionarios, nada se opuso a separar las cuentas enteras para hacer sitio a las otras. La introducción de nuevas cuentas cambia la estructura del antiguo conjunto en un único aspecto de interés inmediato. En el nuevo conjunto, entre dos cuentas hay un número infinito de las otras; además, es imposible decir qué cuenta precede o cuál sigue a una cuenta determinada. Ahora bien, esta diferencia no significa que *ahora* las cuentas se encuentren estrechamente unidas por una cuerda. Ya desde Pitágoras se ha sabido que el nuevo conjunto está también lleno de vacíos; es decir, sigue sin haber una cuerda; pero, aunque ignorásemos la existencia de esos vacíos –como lo hacían de hecho los pitagóricos antes de su descubrimiento de la inconmensurabilidad–, estaríamos desconcertados por la sugerencia de que el conjunto de «cuentas» racionales no está estrechamente unido por una cuerda. Lo que ha sucedido después es una historia habitual.

Con cada nuevo empleo inventado para los números, se descubrió una nueva serie de vacíos en lo que anteriormente se había creído era una entidad «continua». Y no acababan de descubrirse nuevos vacíos cuando estaban ya ocupados con nuevos números. El proceso se repitió varias veces antes de que el conjunto de números alcanzase su actual extensión conocida como el continuo aritmético. No es preciso añadir que durante cada una de las fases precedentes era posible declarar que no podía imaginarse ninguna continuidad más allá de la representada por el sistema numérico conocido en esa época específica. Aún más, *ex post* sabemos que, cuando se formuló, semejante declaración era errónea. Así pues, ¿por qué tendría que estar libre del mismo tipo de autoengaño la misma declaración hecha por Bertrand Russell acerca del continuo aritmético? Como lo observó Poincaré, «el poder creador de la mente [no se ha] agotado con la creación del continuo matemático»[1]. Por consiguiente, sería insensato esperar que no se encontrasen nuevos empleos para los números y, por tanto, que no se descubriesen nuevos vacíos en el continuo de Dedekind y Cantor.

En realidad, no precisamos esperar ni buscar tales empleos. Para poner un ejemplo familiar al economista matemático, el cálculo de probabilidades enseña que la probabilidad de que una variable absolutamente continua, X, tome un valor dado, x, es cero. Ahora bien, este «cero» abarca una gama infinita de casos distintos y relevantes. Está el caso en el que a X le es *absolutamente imposible* tomar el valor x: la varianza muestral, por ejemplo, no puede tener un valor negativo. Sin embargo, «cero» abarca también el caso en el que X *debe* tomar necesariamente el valor x ahora y siempre: en términos abstractos, hay infinitos hombres cuya estatura exacta es de seis pies. Igualmente, los hombres de seis pies de estatura son, en cierto sentido definido, más frecuentes que los de siete pies de estatura. Como quiera que sea, esas diferencias no pueden expresarse *directamente* con ayuda de los elementos del continuo aritmético. Es cierto que podemos desplazar el problema de una diferencia de probabilidades cero a una de densidades de probabilidad, pero ni siquiera este procedimiento nos permite distinguir entre el caso en el que $X = x$ es imposible y aquel en el que $X = x$, a pesar de ser posible, tiene una densidad de probabilidad cero.

Y si, al igual que la mayoría de nosotros, un profesor de estadística ha luchado con inmensas dificultades para hacer entender del todo a sus estudiantes por qué Prob $[x_1 \leq X \leq x_2]$ puede ser positiva a pesar de que Prob $[X = x] = 0$ para todo x, $x_1 \leq x \leq x_2$, no debería culparse ni a sí mismo ni a sus estudiantes. Esas dificultades son exógenas a la estadística; sus raíces se

[1] Henri Poincaré, *The Foundations of Science* (Lancaster, Pa., 1946), p. 50.

extienden profundamente dentro del propio análisis matemático. Puede ser que los matemáticos, tras haberse convencido de que el sistema de números reales es «perfecto y relacionado», sean actualmente algo reacios a considerar la idea de que podría ser «imperfecto» por seguir teniendo vacíos que podrían llenarse con nuevos números[2].

Sea como sea, las actuales dificultades proceden de dos fuentes (como espero demostrar en este apéndice). Por muy sorprendente que pueda parecer, la primera fuente es la confusión creada por la introducción solapada de la idea de medición en una fase prematura del análisis aritmético. La segunda fuente (de carácter subjetivo) es la imposibilidad de construir una escala satisfactoria para lo infinitamente grande o para lo infinitamente pequeño con la sola ayuda de los números reales.

2. Dos cuestiones han de dejarse perfectamente en claro desde el comienzo. En primer lugar, con el fin de señalar diferencias como las que existen entre la probabilidad de que un hombre tenga menos de seis pies de estatura y la probabilidad de que la estatura de un hombre sea exactamente de seis pies, no es necesario que atribuyamos un número real a cada elemento; sería totalmente suficiente ordenar todas las probabilidades en un agregado de elementos no especificados. En segundo lugar, como lo ponen de manifiesto las consideraciones de la sección precedente, no hay nada que nos impida (es decir, no surge ninguna inconsistencia con respecto al orden) intercalar nuevos elementos en el «vacío» obtenido al cortar en dos un agregado ordenado.

Actualmente es un lugar común que la raíz principal del concepto de número es la operación de ordenar elementos y no la de medir cuantías. Un modo más expresivo de decir lo mismo es que el papel básico del concepto de número es el de permitirnos hablar acerca de los elementos de un agregado ordenado. Cuando se les despoja de su ropaje técnico, los números no son más que nombres que pueden darse a los elementos de un agregado ordenado de una manera consistente con la estructura de ese agregado. Así, por ejemplo, los números reales son los nombres que pueden darse a un agregado ordenado que tiene las propiedades características del agregado conocido como el continuo aritmético. Hay que subrayar que es el agregado en cuestión el que determina la forma en que ha de denominarse a sus elementos, y no al revés.

[2] La definición de un conjunto continuo como conjunto perfecto y relacionado pertenece a G. Cantor, *Contributions to the Founding of the Theory of Transfinite Numbers* (Nueva York, sin fecha), p. 72. En lo que respecta a una definición dedekindiana del continuo lineal, véase R. L. Wilder, *The Foundations of Mathematics* (Nueva York, 1956), pp. 140 y s.

Hay que recordar que, en la jerga matemática, un agregado ordenado es «perfecto» si toda secuencia del agregado tiene un elemento restrictivo dentro del agregado y si todo elemento es un elemento restrictivo de tal secuencia. Un agregado ordenado es «relacionado» si entre cualesquiera dos elementos hay otro elemento.

No constituye novedad alguna el hecho de que, dado cualquier agregado ordenado de elementos discretamente diferenciados, podemos construir otro agregado intercalando algunos agregados ordenados entre los elementos del primero. Incluso podemos usar la forma impresa en mayúsculas, Número, para indicar un miembro de un agregado ordenado derivado de este modo a partir del continuo aritmético. Sin embargo, el aspecto que deseo resaltar es que, cualquiera que sea la frecuencia con la que repitamos esta operación, los elementos del nuevo agregado no pueden perder su carácter de ser discretamente diferenciados. En otras palabras, no podemos lograr una continuidad dialéctica a partir de la base de agregados discretamente diferenciados, por muy «densos» que puedan ser éstos.

En este momento, es oportuno mencionar también que «Cero», en el sentido de Nada, es totalmente ajeno al concepto de orden. En un agregado ordenado, a cualquier elemento se le puede dar el nombre de Cero igual que cualquier otro nombre que pudiéramos inventar; pero si al hacerlo así establecemos alguna relación entre ese elemento y Nada, hemos adulterado implícitamente el concepto de orden por una impureza de medición. Todo lo que podamos decir después acerca de esta estructura no pertenece ya al orden puro. Lo mismo es cierto de cualquier uso de «infinito» en relación con un agregado ordenado, si el término implica infinidad real. Difícilmente puedo exagerar mi opinión de que todas las bases axiomáticas propuestas para el sistema de los números naturales, como la famosa de G. Peano[3], son impuras en este sentido, pues suponen un primer elemento antes del cual hay Nada.

3. Con el fin de actuar de forma sistemática, indiquemos como es habitual el continuo aritmético por R y dividámosle en dos subconjuntos, los números no positivos y los positivos. En el vacío así creado, coloquemos un conjunto $[\alpha]$, consistente en el conjunto ordenado de todos los números positivos. Naturalmente, esta operación lleva al ordenamiento *Cero* $\prec \alpha \prec r$ para todo α y para todo r positivos. Simétricamente, podemos intercalar el conjunto $[-\alpha]$ entre el subconjunto de números negativos de R y su *Cero* con el ordenamiento $-r < -\alpha <$ *Cero.* En términos más generales, vamos a definir una agregado $[p]$ de Números escritos en la forma compleja $p = (r,\gamma)$, donde r y γ son cualesquiera miembros de R, y ordenemos el agregado de acuerdo con las reglas siguientes:

$$(1) \qquad \begin{aligned} &(r_1,\gamma_1) \prec (r_2,\gamma_2) &&\text{si } r_1 < r_2; \\ &(r,\gamma_1) \prec (r,\gamma_2) &&\text{si } \gamma_1 < \gamma_2. \end{aligned}$$

Es evidente que este ordenamiento, que representa el ordenamiento familiar de los puntos del plano Euclideo (r,γ) por medio de la regla lexicográfica, es transitivo. Una operación ampliada de suma en $[p]$ lo sugiere inmediatamente:

$$(2) \qquad (r_1,\gamma_1) + (r_2,\gamma_2) = (r_1 + r_2 + \gamma_1 + \gamma_2).$$

Con respecto a esta operación, $[p]$ es un grupo Abeliano, siendo su módulo (0,0).

Intentemos ahora introducir una medida en $[p]$ que preserve el ordenamiento y la suma (2). No tenemos más que considerar el subconjunto de todos los p tal que $(0,0) \prec p$ o $p = (0,0)$. La condición preservadora del orden exige que

$$(3) \qquad \text{Med } (0,0) \prec \text{Med } (0,\gamma) \prec \text{Med } (r,0)$$

para todo $\gamma > 0$. A partir de (2), obtenemos $n \times (1,0) = (n,0)$, para cualquier numero entero n, y, a partir de aquí, a través de un procedimiento muy conocido, podemos deducir $r \times (1,0) = (r,0)$. Esta relación nos induce a definir

$$(4) \qquad \text{Med } (r,0) = r, \quad r \geq 0,$$

y a sustituir la (3) por

$$(5) \qquad 0 < \text{Med } (0,\gamma) < r$$

para todo $r > 0$. En esta fase, se invoca un principio fundamental de la teoría de la medida, que dice que una medida menor que cualquier número positivo es cero[4]. Sobre esta base, a partir de (5) se concluye que para todo $\gamma > 0$

$$(6) \qquad \text{Med } (0,\gamma) = 0,$$

conclusión a la que me volveré a referir más adelante (Sección 10).

Observemos también que a partir de las ecuaciones (1) y (2) se sigue que, para $\gamma > 0$, $r > 0$,

$$(7) \qquad Si(0,\gamma) = (0,n\gamma) = n(0,\gamma) \prec (r,0), \quad i \in [n]$$

donde la suma abarca tantos términos como la potencia del conjunto $[n] = (1,2,\ldots,n)$. Por consiguiente, frente a lo que sucede en el caso de R, el conjunto $[p]$ no satisface el axioma de Arquímedes; en otras palabras, no pue-

[3] Véase Wilder, *Foundations*, p. 66.

[4] Este principio se ha usado implícitamente en el análisis matemático mucho antes de que Émile Borel inaugurase la moderna teoría de la medición. El propio Borel lo usó implícitamente en su análisis de conjuntos de medida cero (Émile Borel, *Leçons sur la théorie des fonctions*, París, 1898, p. 47), pero fue más explícito en su obra *Les nombres inaccessibles* (París, 1952), p. 128. Por regla general, no obstante, el principio sólo está imperfectamente expresado: por ejemplo, «la medida lineal del conjunto de puntos en un intervalo lineal (a,b) se supone que es $b - a$», sin tener en cuenta si se incluyen o no los puntos finales, o «un intervalo lineal tiene la medida del plano cero». E. W. Hobson, *The Theory of Functions of a Real Variable and the Theory of Fourier's Series* (2 vols., Nueva York, 1957), I, p. 165.

de alcanzarse ningún Número $(r,0)$ sumando repetidamente un $(0,\gamma)$. Podemos expresar este hecho diciendo que, con respecto a $(0,\gamma)$, todo $(r,0)$ es un *Número infinitamente grande*. Para ver la implicación completa de esta conclusión, observemos que para el subconjunto $[(0,\gamma)]$ podemos establecer una escala gracias al mismo procedimiento utilizado para $(r,0)$. En esta *segunda* escala,

$$\text{med}\,(0,\gamma) = \gamma, \qquad \text{med}\,(r,0) = \infty, \tag{8}$$

habiéndose obtenido la última relación a partir de med $(0,\gamma) <$ med $(r,0)$ invocando otro principio de medición, en concreto, que un número mayor que cualquier número positivo es infinito, ∞.

Las relaciones (8) son los correlativos evidentes de (4) y (6). Y, dado que queremos que la medición sea compatible con la operación de suma, tenemos en general

$$\text{Med}\,(r,\gamma) = r \tag{9}$$

y

$$\text{med}\,(r,\gamma) = \infty \text{ ó } \gamma, \tag{10}$$

según sea $r > 0$ ó $r = 0$.

4. No tenemos que añadir nada más para conectar el conjunto $[(0,\gamma)]$ con el concepto de infinitésimo que surgió de las aproximaciones sucesivas de Newton y que ha sido, a lo largo de los años, objeto de controversias en las que han participado algunos de los más grandes matemáticos[5]. No más de diez años tras la muerte de Newton, un famoso filósofo, Bishop Berkeley, denunció los infinitésimos como «fantasmas de cantidades difuntas»; en nuestro siglo, otro famoso filósofo protestó de que «la filosofía del infinitésimo... es fundamentalmente negativa»[6]. A menudo hemos leído también que G. Cantor y G. Peano demostraron «la no existencia de magnitudes en realidad infinitamente pequeñas»[7].

La verdad es que lo que únicamente demostraron fue que esos números infrafinitos –como es preferible llamarlos a fin de evitar cualquier confusión con los infinitésimos del cálculo– no satisfacen todos los axiomas del continuo aritmético. A la vista de esta diferencia, puede estar justificado considerar que «*en el análisis aritmético, no tiene cabida la noción de lo realmente infinitésimo*». Evidentemente no tiene cabida, pero sólo porque (como veremos en la Sección 10 posterior) se le desterró del análisis mate-

[5] Véase Cantor, *Contributions*, p. 81; Hobson, *Theory of Functions*, I, pp. 57 y s.

[6] George Berkeley, «The Analyst or, A Discourse Addressed to an Infidel Mathematician», *The Works of George Berkeley* (4 vols., Oxford, 1901), III, p. 44; Bertrand Russell, *Mysticism and Logic* (Nueva York, 1929), p. 84.

[7] Phipip E. B. Jourdain en el Prefacio a Cantor, *Contributions*, p. 81.

mático a través de un postulado oculto. Afirmar entonces –como lo hizo Hobson a continuación– que el número infrafinito es una «variable en un estado de flujos, nunca un número,... una forma de expresión, atractiva como es para un modo de pensar que es esencialmente no aritmético» es tratar un serio problema de acuerdo con la idea Berkeley-Russell[8].

Es cierto que la sustitución de cualquier número real por un campo infrafinito, como en la construcción de $[p]$, suprime la propiedad de Arquímedes del sistema de números reales[9], pero así lo hace también el perfeccionamiento de R por los números cardinales transfinitos, los números Alfa (Aleph) de Cantor. Estos satisfacen

$$S_i \aleph_k < \aleph_{k+1}, \qquad i \in I, \tag{11}$$

incluso aunque I tenga la potencia de $\aleph_k$. En realidad, en apoyo de lo infrafinito podemos invocar la propia defensa de Cantor de lo transfinito: «Todas las pretendidas demostraciones de la imposibilidad de los números realmente infinitos son... falsas en cuanto comienzan atribuyendo a los números en cuestión todas las propiedades de los números finitos, mientras que los números infinitos... deben constituir por el contrario *una nueva clase de números*»[10].

En consecuencia, tendríamos que esperar que algunas proposiciones sobre lo infrafinito irritasen nuestro normal sentido común del mismo modo que lo hicieron antes otras, referentes a lo transfinito. Así, por ejemplo, las relaciones establecidas en la sección anterior llevan a

$$\text{Med}\,(r,0) + \text{Med}\,(0,\gamma) = \text{Med}\,(r,0). \tag{12}$$

Indudablemente, esto es lo que Johann Bernoulli, S. D. Poisson y muchos otros matemáticos clásicos pensaban al decir que «una cantidad que está aumentada o reducida en una cantidad infinitamente pequeña ni aumenta ni se reduce». Esta forma de expresar sus ideas puede no ser precisamente afortunada, pero denunciar la propia idea por rayar «con la mística y el absurdo»[11] es síntoma de una desafortunada parcialidad, porque actualmente no observamos ya nada místico o absurdo en la relación correlativa de la (12),

$$\text{med}\,(r,0) + \text{med}\,(0,\gamma) = \text{med}\,(r,0). \tag{13}$$

Sobre esta misma idea –que con respecto a una escala infrafinita todos los números finitos tienen una medida infinita de la misma manera que

[8] Hobson, *Theory of Functions*, I, p. 43.

[9] A efectos de completar lo anterior, podemos añadir que también se echa por la borda otra de las propiedades de R –la separabilidad–. Véase Wilder, *Foundations*, pp. 140 y s.

[10] Cantor, *Contributions*, p. 74. Las cursivas son mías.

[11] H. Eves y C. V. Newsom, *An Introduction to the Foundations and Fundamental Concepts of Mathematics* (Nueva York, 1958), p. 186. También E. T. Bell, *The Development of Mathematics* (Nueva York, 1940), p. 263.

todos los números transfinitos tienen una medida infinita en una escala finita– erigió G. Veronese su geometría de lo infrafinito y lo transfinito[12]. En términos más generales, hay –como vamos a exponer ahora– una sucesión infinita (en ambos sentidos) de clases de Números, teniendo cada clase su propia escala; cualquier clase es *finita* con respecto a su escala, *transfinita* con respecto a la de la clase inmediatamente anterior e *infrafinita* con respecto a la de su sucesora. Qué escala pueda elegirse como escala finita es una cuestión tan arbitraria como la elección del origen de coordenadas a lo largo de una línea recta homogénea[13].

5. Está fuera de duda que existen innumerables casos en los que sin el infrafinito sería imposible expresar las diferencias que las sutilezas analíticas han creado y crean continuamente en su marcha aparentemente irreversible. El conjunto $[p]$, aun cuando es solamente un primer e imperfecto paso en el mundo de lo infrafinito, proporciona algunas ilustraciones sencillas. Así, si la gama de posibles valores de una variable estocástica absolutamente continua, X, es (A,B), la probabilidad de $X = x$, $A \leq x \leq B$, viene representada por algún $(0,\gamma)$, en tanto que las probabilidades de $A < x_1 < X \leq x_2 < B$, $A < x_1 \leq X < x_2 < B$ están representadas por dos diferentes p, (r,γ_1) y (r,γ_2). Para una ilustración todavía más instructiva, consideremos el caso sencillo en el que Prob $[X = x] = (0,\gamma)$ para cualquier $A \leq x \leq B$. En este caso, gracias a una fructífera analogía con la integral de Lebesgue, podemos escribir

$$(14) \qquad S_i(0,\gamma) = [\gamma(B-A),0], \qquad i \in (A,B),$$

donde S_i es una suma en la que existe un término para cada elemento del intervalo (A,B). No tenemos más que sustituir cada uno de los lados de la relación (14) por las medidas correspondientes para transformarla en el caso más sencillo de la integral de Lebesgue. Lo que la relación (14) dice es, por otra parte, que, a pesar de que el axioma de Arquímedes no actúa en el caso de una suma *numerable* de números infrafinitos, podría hacerlo si la potencia de la suma es la del continuo aritmético. Que esto no es siempre cierto, se demuestra por el hecho de que, de acuerdo con la misma idea de Lebesgue,

[12] Giuseppe Veronese, *Fondamenti di Geometria* (Padua, 1891), obra de la que existe también una traducción alemana, *Grundzüge der Geometrie* (Leipzig, 1894).

[13] Sobre esta misma base (la homogeneidad de la línea recta), Veronese (*Fondamenti*, pp. 101-103) afirmó que, en contraposición a la secuencia ordinal transfinita de Cantor ω, $\omega + 1$, $\omega + 2$, $\omega + 3$,..., deberíamos concebir el infinito ∞_1 no sólo seguido por $\infty_1 + 1$, $\infty_1 + 2$, $\infty_1 + 3$,..., sino también precedido por ..., $\infty_1 - 3$, $\infty_1 - 2$, $\infty_1 - 1$. «No existe ningún primer número infinito», porque en la infinidad homogénea hay «muchos números $\infty_1 - n$, distintos de ∞_1, entre los números finitos y el número ∞_1». Es instructivo relacionar esta postura con el hecho de que en el sistema cantoriano la proposición de que X_0 es el *primer* número transfinito se ha demostrado únicamente con ayuda del controvertido Axioma de la Elección, de Zermelo (mencionado en la nota 24 del Capítulo III). Véase Hobson, *Theory of Functions*, I, p. 208.

$$(15) \qquad \text{Med } [S_i\,(0,\gamma)] = 0, \qquad i \in \Gamma,$$

donde Γ denota el famoso conjunto ternario de Cantor[14].

6. No puede negarse que el velo de la medición numérica oculta el infinito espectro de diferencias que realmente existen. Así es, por ejemplo, la diferencia entre $S_i\,(0,\gamma)$, $i \in \Gamma$, y $S_i\,(0,\gamma)$, $i \in N$, donde N es el conjunto de todos los números enteros positivos. El problema consiste en saber si hay alguna escala sobre la que puedan representarse sistemáticamente esas diferencias.

El primer paso hacia la construcción de una escala semejante lleva consigo la solución de un problema relativamente sencillo: qué tipo de número infrafinito tendría que sustituir a $(0,\gamma)$ en la expresión (15) o en $S_i\,(0,\gamma)$, $i \in N$, para que esas sumas tuviesen una medida finita. En otras palabras, ¿hay un número infrafinito π tal que Med $(S_i\,\pi) = 1$ para $i \in N$? Aun cuando esta última cuestión pueda parecer que es la más sencilla de su tipo, es especialmente adecuada para mostrar lo inmensamente complejos que son los problemas planteados por el concepto de lo infrafinito.

La cuestión se encuentra relacionada con el problema de elegir un número entero positivo *completamente al azar*, es decir, por un procedimiento tal que la probabilidad de elegir cualquier número entero sea la misma. Dentro del análisis aritmético, la respuesta es que esta probabilidad, π, es cero. Se trata de una respuesta paradójica, porque, si $\pi = 0$, la probabilidad de elegir un número que no sea mayor que n es $S_i\,\pi = 0$, $i \in [n]$, para cualquier n. Como observa Borel, hemos de concluir que el número elegido será seguramente un «número inaccesible», es decir, un número que supera el límite, no de nuestro poder de imaginación, sino de nuestra capacidad de tratar con él en la realidad. Sobre la base de esta paradoja, Borel afirma que la distribución uniforme a lo largo de un conjunto numerable es un absurdo matemático. Así, hay que asignar necesariamente menores probabilidades a los números inaccesibles, de modo que, cuanto más inaccesibles son, menores sus probabilidades[15]. Borel admite ciertamente que esta conclusión se basa en gran medida en consideraciones *prácticas*. Ahora bien, hay que subrayar que, si las consideraciones prácticas se convirtiesen en el criterio de separación entre sentido y sin sentido, mucho de lo que pasa por alta matemática no sería más que un sin sentido.

Ahora bien, la argumentación de Borel puede aplicarse incluso aunque

[14] A este respecto, véase, por ejemplo, B. R. Gelbaum y J. M. H. Olmsted, *Counterexamples in Analysis* (San Francisco, 1964), pp. 85-87.

[15] Borel, *Les nombres inaccessibles*, pp. 37-42. A propósito, el concepto de Borel de un número inaccesible proporciona una ilustración especialmente interesante de los conceptos omnipresentemente dialécticos en mi propio sentido. Existe únicamente un límite imperfectamente definido (es decir, una penumbra limitada por otras penumbras) que separa los números accesibles de los inaccesibles. Sin embargo, esto no nos impide estar seguros de que 11 es un número accesible y $1.000^{1.000}$ un número inaccesible (Borel, *ibid.*, p. 4).

se asuma la existencia de un número infrafinito, π, tal que Med $(S_i\ \pi) = 1$, $i \in N$: porque, entonces, necesariamente Med $(S_i\ \pi) = 0$, $i \in [n]$. Y esto no constituye en absoluto una paradoja distinta de la que surge del contraste entre (14) y (15). Así pues, el caso de las probabilidades numerables no puede particularizarse sobre esta base. En mi opinión, lo que individualiza una suma numerable con respecto a una suma continua, como la (14) o la (15), es el hecho de que todavía no existe ningún concepto de medición dentro de un conjunto numerable. Esta es la explicación del hecho, resaltado por Borel, de que, aun cuando la noción de los matemáticos de una serie ilimitada de números enteros es «aparentemente tan clara y precisa», el problema de la probabilidad en lo numerable es más complicado que en la compleja estructura del continuo[16].

De acuerdo con una idea de Borel, podemos admitir que el conjunto de puntos de la abscisa $(10,10^2,10^3,\ldots)$ es más enrarecido que el de la abscisa $(1,2,3,\ldots)$[17]. En el caso en que el orden de los elementos viene dado por los datos objetivos del problema, podemos utilizar fácilmente la observación de Borel para construir una medición. La medición del subconjunto $[mn]$ de N para un m dado es $1/m$, y la de $[10^n]$ es cero. Sin embargo, el concepto de un conjunto numerable implica únicamente que existe *cierta* forma, no una *determinada*, de ordenar los elementos exhaustivamente en una secuencia. Casi todos los conjuntos numerables de puntos en un espacio de mayor dimensión que la primera no están asociados con un orden «natural». Este aspecto tiene una ilustración excelente en una paradoja usada por Borel contra la distribución uniforme a lo largo de un conjunto numerable y que es independiente de que se admita o no la existencia del infrafinito π.

F. Hausdorff ha demostrado que en una esfera podemos construir tres conjuntos numerables inconexos A, B, C, tales que una rotación haga que A coincida con $C + B$, y que otra rotación haga que A coincida con B, B con C y C con A. A partir de la idea aparentemente inevitable de que las probabilidades de elegir un punto entre los conjuntos de puntos que son

[16] Émile Borel, *Probability and Certainty* (Nueva York, 1963), p. 79; Borel, *Les nombres inaccessibles*, p. 100. La cuestión es más desconcertante de lo que sugiere la observación de Borel. Como lo señaló Paul Lévy, en *Théorie de l'addition des variables aléatoires* (París, 1937), p. 25, en el caso de las distribuciones de probabilidad a lo largo de un conjunto con una potencia mayor que la del continuo «hay un caos completo». Esta situación se debe indudablemente –admitámoslo– al hecho de que el concepto de medida se basa en nuestras nociones intuitivas de longitud, superficie y volumen. Posiblemente, el rechazo a tener nada que ver con lo infrafinito nos impide enfocar el problema desde la dirección sugerida anteriormente: ¿qué tipo de infrafinito σ corresponde a Med $(S_i\ \sigma) = 1$, $i \in F$, si F es un conjunto con una potencia determinada mayor que la del continuo?

[17] Borel, *Les nombres inaccessibles*, pp. 85 y s. Naturalmente, si aceptamos que los conjuntos se «reorganicen» en el sentido usado con frecuencia en el análisis de las potencias numerables, la rarefacción pierde todo su sentido. Pero no debemos olvidar que la reorganización destruiría la medida incluso en el continuo: al reorganizar Γ, podemos atribuirle cualquier medida que queramos. Se trata aquí de cuestiones matemáticas que siguen sin resolverse.

congruentes en la geometría euclideana son iguales, se deduce que Prob $[x \in A]$ = Prob $[x \in B]$ + Prob $[x \in C]$, en la primera rotación, y Prob $[x \in A]$ = Prob $[x \in B]$ = Prob $[x \in C]$, en la segunda. Todas esas probabilidades deben ser cero, pero si A, B, C, son los únicos conjuntos de entre los que se elige un punto, las mismas probabilidades deben sumar la unidad. De aquí surge la paradoja[18].

7. Es elemental que el número infrafinito τ que satisficiera Med $(S_i\ \tau) = 1$, $i \in \Gamma$, debe pertenecer a una clase diferente de la de $(0,\gamma)$ en (14) o de la de π en Med $(S_i\ \pi) = 1$, $i \in N$. El hecho de que el producto de dos números infrafinitos $(0,\gamma_1) \times (0,\gamma_2)$ no pueda pertenecer a $[p]$ sin contradecir su estructura no arquimediana nos lleva también a la misma conclusión. Estamos así inducidos a definir $(0,\gamma_1) \times (0,\gamma_2) = (0,0,\gamma_1\gamma_2)$, donde $(0,0,\gamma_1\gamma_2)$ es un número infrafinito de segundo orden. Obviamente, esto nos vuelve a llevar a la misma regresión infinita que la tan minuciosamente descrita por Cantor: «en la sucesiva formación de las clases de números, podemos ir siempre más lejos, sin alcanzar nunca un límite que no pueda sobrepasarse, de modo que nunca alcanzamos ni siquiera una [comprensión] aproximada de lo Absoluto... Lo Absoluto sólo puede [concebirse], pero nunca [realizarse] ni siquiera aproximadamente»[19]. La diferencia es que lo infrafinito se mueve en el sentido opuesto, desde lo finito a la Nada Absoluta que, si uno se para a pensar sobre ello, es una noción filosófica tan desconcertante como el Infinito Absoluto. Borel llegó incluso a pensar que «lo infinitamente pequeño, aun cuando aparentemente más cercano a nosotros y más familiar que lo infinitamente grande, es, hablando en términos relativos, más difícil de medir y de entender»[20].

Una idea muy sencilla para resolver la regresión infinita de las clases infrafinitas consiste en definir un Número

$$\rho = (r_1, r_2, r_3, \ldots) \tag{17}$$

[18] Borel, *Les nombres inaccessibles*, pp. 95-100 y 124-126; Borel, *Les paradoxes de l'infini* (2ª edic., París, 1946), pp. 198-210. Borel complica aún más la paradoja utilizando el Axioma de la Elección, de Zermelo, para dividir la totalidad de la esfera en tres conjuntos que tienen las mismas propiedades que A, B, C. Sin embargo, en este caso podría objetarse que, dado que esos conjuntos no son mensurables, no debería hablarse de probabilidad en relación con ellos. En matemáticas, «probabilidad» y «medida» son conceptos intercambiables. En consecuencia, está lejos de ser cierto que la paradoja de Hausdorff demuestre –como afirma Borel en *Les paradoxes*, p. 210– que ha de descartarse el axioma de Zermelo.

[19] Cantor, *Contributions*, p. 62*n*. Evidentemente, Cantor tenía en mente el Absoluto Infinito de Hegel, es decir, el infinito, Ω, para el que no es permisible escribir $\Omega + 1$, porque no hay nada que no esté ya cubierto por Ω. En consecuencia, ni la antinomia de Burali-Forti ni la de Bertrand Russell pueden actuar en el caso de Ω. En realidad, la solución propuesta por Bertrand Russell para su antinomia –en concreto, eliminar de la Lógica el concepto de «la clase de todas las clases»– equivale a decir que la Lógica debería reconocer que Ω es la única clase que no puede ser miembro de ninguna clase que contenga a otros miembros. (En lo que respecta a las antinomias mencionadas, véase Wilder, pp. 55 y s. y 124).

[20] Émile Borel, *Probability and Certainty*, p. 84.

tal que r_i sea un número infrafinito de primer orden con respecto a r_{i-1}, de segundo orden con respecto a r_{i-2},... y un número transfinito de primer orden (no en el sentido de Cantor) con respecto a r_{i+1}, y así sucesivamente. La idea de ordenar este nuevo agregado de acuerdo con la regla lexicográfica es totalmente natural. Tanto las operaciones de adición y multiplicación como las definiciones de medida en varias escalas se extienden fácilmente a $[\rho]$ de la misma manera que la utilizada para $[p]$. La operación de división es, por tanto, sencilla, como puede ilustrarse con el sencillo caso de

$$\rho : (1,\gamma) = \rho \times (1, -\gamma, \gamma^2, -\gamma^3, \ldots). \tag{18}$$

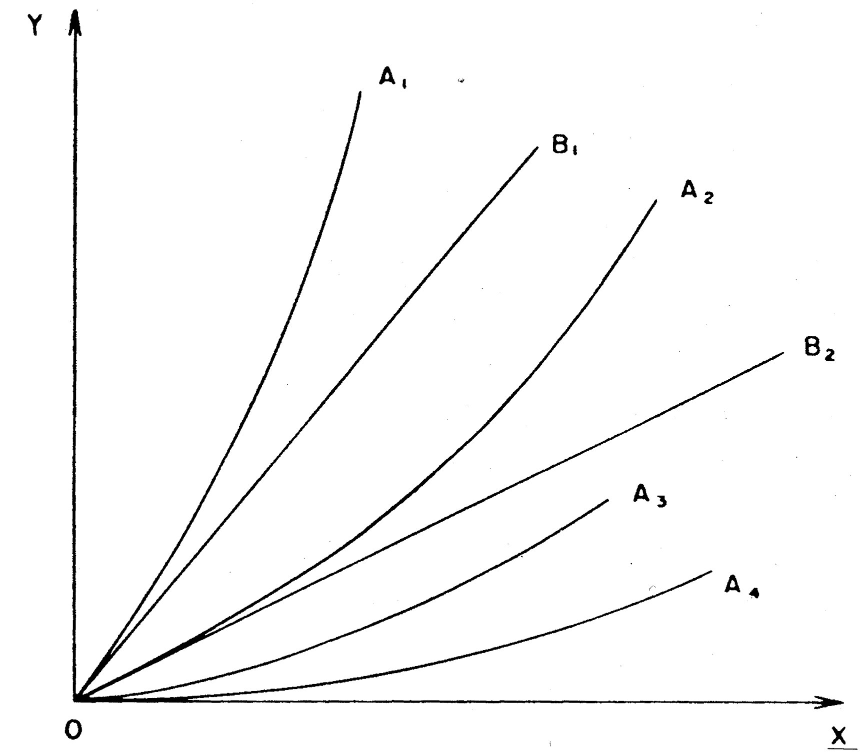
Lo novedoso de la idea de Veronese con respecto a la de Dedekind es que la línea geométrica se compone de todos los puntos que tienen un $[\rho]$ como abscisa, no sólo de aquellos cuyas abscisas son números reales. Sin embargo, estamos condicionados tan a fondo a pensar únicamente de acuerdo con la línea del postulado de Dedekind[21] que somos capaces de tratar la representación de la línea a través de un continuo de Veronese como absurdo matemático. Con todo, nadie más que David Hilbert, el fundador de la geometría axiomática pura, demostró que una geometría no arquimediana, como la de Veronese, funciona tan bien como la de Dedekind[22].

8. A la vista de las numerosas proclamaciones dogmáticas de que sencillamente no existe el infinitésimo *real*, no deberíamos dejar de mencionar una aplicación muy elemental de $[\rho]$, en la que el infrafinito es tan real que hasta un alumno de grado medio sería capaz de dibujarla en un papel. Los ángulos llanos normales, como el B_1OX en la Figura 3, se miden por una escala de números reales finitos. Ahora bien, pueden considerarse también –como lo hicieron hasta los geómetras griegos– los ángulos curvos, formados por una línea recta y una curva o por dos curvas. A_1OB_1 y A_2OA_3 son ejemplos de tales ángulos. Es lógico que el ángulo curvo A_1OX pueda considerarse mayor que el A_2OX: los correspondientes ángulos llanos formados por las tangentes a OA_1 y OA_2 satisfacen la desigualdad $B_1OX > B_2OX$. Es igualmente lógico que el ángulo curvo A_1OX pueda considerarse mayor que el ángulo llano B_1OX. El problema pone de manifiesto sus aspectos más profundos si consideramos los ángulos curvos A_3OX y A_4OX (siendo OA_3 y OA_4 tangentes a OX en O). Dado que para esos dos ángulos los correspondientes ángulos llanos son cero, su diferencia puede

[21] R. Dedekind, *Essays on the Theory of Numbers* (Chicago, 1924), pp. 6-8; véase también la nota 22, Capítulo III.

[22] David Hilbert, *The Foundations of Geometry* (2.ª edic., Chicago, 1921), pp. 34-36. En relación a mi tesis sobre la textura granular del continuo aritmético, es muy importante mencionar que Dedekind, *Essays*, pp. 37 y s., conjeturó correctamente que la validez de los *Elementos* de Euclides no se ve afectada si, por el contrario, «afinamos» la línea excluyendo de ella todos los puntos cuyas abscisas son números trascendentales.

Figura 3



mostrarse únicamente en otra escala, una escala infrafinita entre lo finito y cero. Y en esta escala, a su vez, no es posible representar todas las diferencias de segundo orden, es decir, las diferencias entre las diferencias mostradas por ángulos tales como A_3OX y A_4OX.

La forma en que puede resolverse, en parte, esta desconcertante cuestión con ayuda de $[\rho]$ se hace perfectamente evidente si consideramos sólo las curvas, (OA), que en las proximidades del origen son convexas y que están representadas por una función analítica

$$y = r_1x + r_2x^2 + r_3x^3 + \ldots, \qquad r_1 \geq 0,\ r_2 \geq 0. \tag{19}$$

El agregado de los ángulos curvos constituidos por esas curvas con OX en O constituye un ejemplo palmario de un agregado ordenado de elementos cuantitativos, debido a que este continuo no es lo suficientemente rico en «cuentas». Volviendo a una geometría sencilla, si para dos curvas, C' y C'', tenemos $r_1' > r_1''$, el mayor ángulo curvo formado con OX es el de C'. Si, como sucede en el caso de OA_3 y OA_4, $r_1' = r_1''$, pero $r_2' > r_2''$, el mayor ángulo es el de C', a pesar de que la diferencia entre los dos ángulos, medida en la misma escala que en el caso precedente, es cero. Cualquier ángulo curvo, como el A_3OX, representa algún número infrafinito. Por consiguiente, lo infrafinito, lejos de ser un fantasma de cantidades difuntas, está «en carne y hueso» directamente ante nosotros para todo aquel que quiera verlo.

La clase de funciones (19) puede extenderse, en primer lugar, para incluir hasta funciones no analíticas siempre que tengan derivadas de cualquier orden para $x = 0$. En este contexto, Felix Klein hace la muy interesante observación de que el ángulo curvo constituido por la función $y = Ae^{-1/x^2}$, $A > 0$, cuyas derivadas de todos los órdenes son cero para $x = 0$, es menor que cualquier ángulo formado por una curva (19)[23]. Sin embargo, constituiría un grave error por nuestra parte concluir a partir de esta observación que las funciones de Klein llenan todo el «espacio» entre lo finito y cero, y sería una equivocación colosal pensar que podemos alcanzar el cero a través de tales funciones. El cero corresponde solamente a $y = 0$; y $Ae^{-1/x^4} > 0$, para todo $A > 0$, $x > 0$. Un aspecto todavía más sutil es que hay funciones ($y = Ae^{-1/x^4}$, por ejemplo) que forman un ángulo curvo menor que los de las funciones de Klein y además otros que forman un ángulo menor que los formados por las funciones «súper Klein», y así *ad infinitum.* Vemos así que $\rho = [0,0,0,...]$, abarca también una infinita variedad de ángulos curvos, no sólo el ángulo que no tiene absolutamente contenido alguno. Además, los ángulos curvos de muchas curvas, aunque diferentes, están representados por la misma ρ, siendo ahora la razón de distinta esencia. Por ejemplo, los ángulos formados con OX por las curvas de $y = x^{3/2}$ e $y = x^{4/3}$ corresponden a la misma $\rho = (0, +\infty, -\infty, +\infty,...)$. Nos vemos así obligados a ver cómo podemos distinguir entre un ∞ y otro ∞. Todo esto demuestra que ni siquiera $[\rho]$ es lo suficientemente rico como para describir todas las posibles diferencias existentes entre los ángulos curvos.

9. El aspecto correlativo para el paso de lo finito al Infinito Absoluto se conoce desde hace tiempo. Me refiero al descubrimiento por Paul du Bois-Reymond de una secuencia infinita de funciones cada vez más crecientes

$$\varphi_1(x) \prec \varphi_2(x) \prec \varphi_3(x) \prec \ldots \prec \varphi_n(x) \prec \ldots, \tag{20}$$

donde la expresión «cada vez más crecientes» significa que, dado cualquier $K > 0$, hay un $X_n > 0$ tal que para todo $x > X_n$ tenemos $\varphi_{n+1}(x) > K\varphi_n(x)$. El famoso teorema demostrado por Bois-Reymond dice que para cualquier secuencia (20) podemos encontrar una función aún más rápidamente creciente, φ, es decir, tal que $\varphi_n \prec \varphi$ para todo n. Un segundo teorema dice que, para cualquier φ que satisfaga esta condición, podemos encontrar una función ψ_1 tal que $\varphi_n < \psi_1 < \varphi$ para todo n. Por reiteración, obtenemos el modelo ordinal

[23] Felix Klein, *Elementary Mathematics from an Advanced Standpoint: Geometry* (Nueva York, 1939), p. 206.

$$\varphi_1 \prec \varphi_2 \prec \varphi_3 \prec \ldots \prec \psi_3 \prec \psi_2 \prec \psi_1 \prec \varphi. \tag{21}$$

Este modelo demuestra que no hay escala de Arquímedes para todas las clases de lo infinitamente grande[24].

Indudablemente, las vías de lo finito al Cero Absoluto o al Infinito Absoluto a través del campo del Análisis son igualmente largas y nunca alcanzan sus destinos finales. Puede sernos difícil comprender este hecho si insistimos en mirar sólo los escasos mojones colocados en esas vías por el continuo aritmético.

10. Para poder resumir la argumentación técnica de este apéndice, precisamos volver al punto, antes mencionado, relativo al postulado oculto en cuya virtud lo infrafinito está excluido *ab initio* del análisis aritmético. Una fábula nos permitirá acallar algunas de las ideas preconcebidas que nuestro hábito aritmético de pensar puede hacer que nos pillen desprevenidos.

Imaginemos una lámpara maravillosa que se «enciende» y se «apaga» por sí sola indefinidamente de acuerdo con el calendario siguiente. La lámpara está encendida en $t = 0$; por consiguiente, en cada instante

$$t_n = 1 + \frac{1}{2} + \frac{1}{2^2} + \ldots + \frac{1}{2^{n-1}}, \ (n = 1,2,3,\ldots), \tag{22}$$

la lámpara se «apaga» o se «enciende» por sí sola según que n sea un reloj extraño o incluso un reloj con sentido del tiempo[25]. Varios aspectos tendrían que quedar ahora en claro, si bien no todos pueden ser inmediatamente evidentes.

[24] G. H. Hardy, *Orders of Infinity: The "Infinitärcalcül" of Paul du Bois-Reymond* (Cambridge, Ingl., 1924), pp. 11 y s. Dado que la ordenación de (21) recuerda las secuencias ascendente y descendente por las que se definen los números irracionales en un enfoque familiar, es oportuno señalar –para su uso posterior– que la (21) no define necesariamente una función de corte X, es decir, tal que $\varphi_n \prec X \prec \psi_m$ para todo n y todo m. En efecto, sea (ψ) la clase de funciones para las que existe $\int^{\infty} \psi^{-1}\, dx$, y (φ) la clase de aquellas para las que no existe la misma integral. Evidentemente, ninguna función corresponde a esta división. En lo que se refiere a estas y otras observaciones altamente interesantes en relación con el problema de la escala, véase Borel, *Leçons sur la théorie des fonctions*, pp. 111-119.

[25] La cuestión aparentemente obvia de que ninguna lámpara real podría verdaderamente parpadear de la forma descrita, cuestión que en su esencia se remonta a Aristóteles, *Physics*, 263^b 15-35, no se ha dado por sentada por todos los filósofos. Fue Bertrand Russell, en «The Limits of Empiricism», *Proceedings of the Aristotelian Society*, XXXVI (1935/6), pp. 143 y s., quien lanzó la idea de que realizar un número infinito de diferentes tareas en un intervalo de tiempo finito no es un absurdo lógico. Sin embargo, la lámpara maravillosa permitió a J. F. Thomson, en «Tasks and Super-Tasks», *Analysis*, XV (1954), pp. 1-13, refutar la idea. Recientemente, empero, parece existir un creciente entusiasmo por proyectos de máquinas que, supuestamente, realizan supertareas del tipo de imprimir todos los decimales de π o de enumerar todos los números enteros dentro de un intervalo temporal finito. Véase A. Grünbaum, «Are "Infinity Machines" Paradoxical?», *Science*, 26 de enero de 1968, pp. 396-406.

En primer lugar, dado que t_n es un número racional *más pequeño* que $t^* = 2$, ningún tipo de lógica puede hacer que consideremos que t^* es miembro del agregado $(t_1, t_2, t_3, \ldots)$. En segundo lugar, el estado de la lámpara –hay cuatro estados en total– está completamente determinado en cada instante t si t es un número real tal que $0 < t < t^*$. En tercer lugar, *sin información adicional* es imposible determinar el estado de la lámpara en $t = 100$ o ni siquiera en t^*: la fábula no dice una sola palabra sobre esos estados[26]; así, por ejemplo, la lámpara podría «desaparecer» en t^* sin contradecir la fábula. En cuarto lugar –lo que constituye el aspecto más importante–, la lámpara podría desaparecer también incluso en un instante t' anterior a t^*.

Debemos esperar que casi todo el mundo denuncie al momento la cuarta afirmación como totalmente absurda. Si t' es un instante anterior a t^*, entonces –se replicará– hay un k tal que t_k es un instante posterior a t'; y, dado que de acuerdo con la fábula la lámpara debe continuar parpadeando después de t_k, es absurdo suponer que puede desaparecer antes.

Ahora bien, al replicar de este modo se ignora un aspecto esencial: a partir del hecho de que t' es anterior a t^*, es decir, $t' < t^*$, no se sigue que $t' < t_k$ para cualquier k. El modelo analítico mencionado en la sección precedente en relación con el teorema de Bois-Reymond demuestra que la ordenación $t_1 \prec t_2 \prec t_3 \prec \ldots \prec t^*$ es totalmente compatible con la existencia de un t' tal que $t_k \prec t' \prec t^*$ para cualquier k. Para evitar la posibilidad de que la importancia de esta observación se ponga en duda para el caso específico en discusión, podemos subrayar, en primer lugar, que la fábula no dice nada acerca de la textura del continuo temporal y, en segundo lugar, que si este continuo es similar al de $[p]$ de la Sección 2 anterior, hay instantes $t' = (t^*, \gamma)$, $\gamma < 0$. Para cualquiera de esos instantes, $t_k < t' < t^*$, para cualquier k.

Lo que elimina esta última alternativa es una serie de postulados. El primero proclama la geometrización del tiempo:

El continuo del tiempo y el de la línea son idénticos.

La lógica de medir el tiempo por la posición de la punta de una aguja de reloj se basa en este postulado. El segundo postulado es el de Dedekind:

El continuo de la línea y el continuo de los Números son idénticos.

Queda por demostrar por qué Números como (r, γ) no se incluyen en la categoría admitida por el análisis aritmético. La exclusión es el resultado de un postulado que, si bien usado implícitamente una y otra vez en el

[26] Véase Thomson, obra recién citada, pp. 5 y s.

análisis aritmético, no se encuentra, se busque como se busque, explícitamente establecido en la literatura. Podemos llamarle el Postulado de la Negación de lo Infrafinito:

Si un Número (considerado sin su signo) es menor que cualquier número real positivo, ese Número es cero.

A la luz de la Sección 3 anterior, podemos ver que este postulado tiene sus raíces en el concepto de medida, no en el de orden puro. Lo que dice en esencia es que, dado que Cero es el único Número (nótese la mayúscula N) con una medida cero, la relación (6) significa que $(0,\gamma)$ no puede ser más que cero. Por consiguiente, lo infrafinito no existe. La afirmación general de que la medida es un proceso extraño a la ordenación es ciertamente válida, pero, como resultado de la infiltración del último postulado en los desarrollos preliminares del análisis aritmético, precisa cierta revisión la afirmación igualmente general de que el análisis aritmético es completamente independiente del concepto de medida.

11. El aspecto que ha de subrayarse ahora es que los postulados recién explicados siguen sin permitirnos determinar el estado de nuestra lámpara maravillosa en t^*. El reconocimiento de este hecho tiene su lugar preciso en el análisis matemático, a saber, en el concepto de discontinuidad simple; y, como lo puso de manifiesto J. F. Thomson, representa también el golpe mortal a la afirmación de que puede realizarse una infinidad de tareas dentro de un intervalo temporal finito.

Sobre esta base –y teniendo en cuenta el paralelismo entre la forma en que se lleva a cabo la conexión de la lámpara y aquella en la que Zenón describe la carrera de Aquiles tras la tortuga–, podemos vernos tentados a afirmar que hasta el análisis matemático confirma a Zenón. Sin embargo, entre las dos fábulas existe una diferencia fundamental que fue claramente señalada por Aristóteles. La locomoción es el Cambio continuo por excelencia; en la jerga de la cinemática, el movimiento es una función continua entre dos variables continuas, tiempo y distancia. Por consiguiente, si $t^* - t = 0$, la distancia entre Aquiles y la tortuga debe ser también cero, no un número indeterminado. Todas las otras formas de Cambio, como la de conectar la lámpara o la de imprimir otro decimal de π, consisten en «tareas» precisas (en la jerga de Thomson) o en unidades «reales» (en la de Aristóteles)[27]. Como lo resaltó Aristóteles, la ingeniosidad de Zenón consistió en describir la locomoción de Aquiles como si consistiese en un número infinito de tareas o de unidades, en una súper-tarea, y en exclamar después «paradoja»[28]. Sin embargo, ir de A a B no es una tarea diferente de conti-

[27] Aristóteles, *Physics*, 260b-261a y 263b 4-5.
[28] *Ibid.*, 263a 22-23.

nuar desde B hasta C, a no ser que uno *se pare realmente* en B, hecho que introduciría la necesaria discontinuidad.

12. La popular refutación de la paradoja de Zenón se basa en la idea de que la *suma* de la serie infinita (obtenida a partir de t_n para $n \to \infty$)

$$1 + \frac{1}{2} + \frac{1}{4} + \dots + \frac{1}{2^n} \dots$$

es $t^* = 2$, entendiéndose la *suma* en el sentido ordinario en el que se aplica a un número finito de términos[29]. Algunos llegan a afirmar que lógicamente no hay absolutamente nada erróneo en la súper-tarea de muchas e infinitamente distintas operaciones de adición[30]. Creo que tales ideas nacen de ciertas expresiones poco precisas que desdibujan el carácter esencial del concepto de límite y que dejamos deslizar con frecuencia en los textos matemáticos. Incluso aunque esas expresiones casi no causasen *per se* daño alguno al análisis aritmético, habría que evitarlas debido a que pueden generar ideas confusas en algunas cuestiones relacionadas, en especial en la de la existencia y la esencia de los números infrafinitos.

El concepto de límite es una *asociación* legítima (e igualmente fecunda) entre una secuencia infinita y un número. Cualquier cosa que pueda sugerir una relación más íntima entre los dos términos –especialmente, una relación de identidad– fomenta la confusión analítica. A fin de poner un ejemplo concreto y destacado, consideremos la asociación del número cero con una secuencia de números *positivos*

(23) $$(a_1, a_2, a_3, \dots, a_n, \dots),$$

tal que, para cualquier $\varepsilon > 0$, tenemos $a_n < \varepsilon$ para todo $n > N(\varepsilon)$. Es perfectamente legítimo expresar esta asociación a través de algún simbolismo, como «$\lim a_n = 0$ para $n \to \infty$», o a través de alguna otra forma, como «a_n es una aproximación de cero»; pero hay que tener en cuenta que se introduce cierta confusión si no se hace claramente hincapié en que 0,999..., por ejemplo, es solamente una notación oportuna en lugar de «lim ($b_n = 9 \Sigma_1^n 10^{-1}$) *para* $n \to \infty$». Por desgracia, hasta las obras de muchas autoridades matemáticas no se pronuncian sobre la diferencia esencial que existe entre las representaciones decimales de 1/4 a través de 0,25 y de 0,24999...[31]. Ahora bien, la mayor fuente de confusión es la muy utilizada expresión «en el límite» o, como diría Bertrand Russell, «tras un número infinito de

[29] Encontramos que esta postura se adoptó incluso por Alfred North Whitehead, *Process and Reality: An Essay in Cosmology* (Nueva York, 1929), p. 107.

[30] Véase, en concreto, J. Watling, «The Sum of an Infinite Series», *Analysis*, XIII (1952), pp. 39-46.

operaciones, a_n se hace cero». La verdad es que a_n *nunca* se hace cero, pues, por mucho que viajemos a lo largo de la secuencia (23), no encontraremos más que números positivos. Siendo esto así, debería ser evidente para todo el mundo que es el límite de esa secuencia, es decir, cero, lo que constituye el fantasma de los difuntos números positivos, a_n. Se utiliza aquí la ocurrencia de Bishop Berkeley no como burla del concepto de límite sino para acentuar que este concepto implica un salto transfinito. Cantor, a quien podemos recurrir una vez más a fin de obtener nueva luz, no identifica su primer número transfinito ω con la secuencia eterna de números enteros; tampoco dice que el número entero finito n se haga ω en el límite. En lugar de ello, coloca a ω en el final transfinito de la secuencia de números enteros, es decir, una vez que todos estos números enteros son difuntos.

La confusión, en estado latente, entre lo infrafinito y una secuencia con el límite cero domina muchas ideas en el análisis infinitesimal. A efectos de proporcionar un ejemplo destacado, vemos que ocasionalmente nos encontramos con la argumentación de que, debido a que $1 - b_n$ puede ser menor que cualquier número positivo arbitrario ε, $1 - 0{,}999... = 0$ es válido en el sentido puramente aritmético en base a la autoridad del Postulado de Negación de lo Infrafinito. Pero esto no es así. El postulado se refiere a un número, no a una variable en «estado de flujo», como es $1 - b_n$. Podemos ver ahora que Hobson vinculó este estado al objeto erróneo: concebido adecuadamente, un número infrafinito es una entidad tan fija y precisa como cualquier número finito o transfinito.

13. Precisamente debido a esta última propiedad es por lo que lo infrafinito no puede permitirnos (no más de lo que puede hacerlo el número ordinario) llegar a una representación aritmomórfica del Cambio. Sin embargo, no sólo ilumina algunas de las insospechadas imperfecciones inherentes al agregado «perfecto» de números ordinarios sino que nos ayuda también a eliminar alguna de esas imperfecciones. Como ejemplo, volvamos al problema de los acontecimientos que pese a todo es posible que tengan una medida de probabilidad igual a cero. Con la ayuda de lo infrafinito, podemos dar una definición precisa de la cuasi-imposibilidad o de la cuasi-certeza, de modo que se distinga claramente entre esas situaciones, por un lado, y la imposibilidad y la certeza, por el otro. Como ya he afirmado en otro lugar[32], semejante distinción es indispensable para un análisis completo de las expectativas, incluso aunque no tenga una relación continua con nuestro comportamiento frente a la incertidumbre. Es úni-

[31] Así, por ejemplo, en Borel, *Les paradoxes de l'infini*, p. 118, encontramos la extraña argumentación de que el límite de la secuencia 0,19, 0,199, 0,1999,... debe escribirse 0,1999..., no 0,2, si «no se quiere eliminar la continuidad».

[32] Véase mi artículo «The Nature of Expectation and Uncertainty» (1958), reimpreso en *AE*, pp. 251-253. Sobre esta cuestión, véase también Richard von Mises, *Probability, Statistics and Truth* (2.ª edic., Londres, 1957), pp. 33 y s.

camente gracias al infrafinito como podemos evitar el embrollo analítico que supone el consejo de Borel «no tener miedo a usar la palabra *certeza* para describir una probabilidad que no llega a la unidad por una cantidad suficientemente pequeña»[33]. Que tal probabilidad, siendo finita, es menor incluso que una cuasi-certeza es un aspecto secundario; pero igualarla a la certeza conduce directamente a la doctrina de Azaïs y Marbe[34].

14. Debiéramos estar ahora en situación de ver que, por encima de las complejidades técnicas de la cuestión discutida en este apéndice, brillan las cristalinas enseñanzas de Aristóteles sobre el concepto general de continuo; porque, cuando todo está ya dicho y hecho, no podemos dejar de reconocer que un continuo, pese a estar definido o aprehendido, se encuentra indisolublemente ligado a una idea fundamental sobre la que Aristóteles insistió concienzudamente. Lo que es infinitamente divisible sigue siendo así en todo tiempo; y, dado que un punto es indivisible, no existe puente alguno entre una *parte propia* de una línea, esto es, una parte que posee la cualidad *específica* de línea, y un *punto.* Al igual que todo lo que es continuo, la línea «es divisible [sólo] en divisibles que son infinitamente divisibles»[35]. Sería duro incluso para un detractor de Aristóteles –lo que no constituye actualmente un raro fenómeno– cerrar sus ojos a la confirmación de esta idea por el propio análisis aritmético. En efecto, si en la expresión (23) los a_n representan las partes sucesivas de una división infinita de un segmento lineal, el propio concepto de límite proclama (si no se abusa de él) que el proceso nunca producirá un punto, es decir, un «segmento» sin longitud.

Sin embargo, Aristóteles sostuvo también que «nada continuo puede estar compuesto de indivisibles: por ejemplo, una línea no puede estar compuesta de puntos, siendo la línea continua y el punto indivisible»[36]. Por otra parte, el continuo aritmético se ha creado exclusivamente de indivisibles, las cuentas sin sarta individualmente distintas de mi metáfora. La contradicción se ha puesto de manifiesto por muchas autoridades de las matemáticas: «la distinción *genérica* entre un objeto geométrico continuo y un punto... situado en ese objeto no es susceptible de representación aritmética directa»[37]. El continuo aritmético concebido exclusivamente como un agregado de indivisibles no ofrece sitio alguno a las propiedades métricas del espacio o de cualquier otra estructura continua de esa materia. Es posible que –como dijo en una ocasión Bertrand Russell–[38], la geometría métrica sea un «pequeño rincón de la Geometría», pero su papel en la cuestión es primordial: proporciona la única prueba de fuego de la importancia del continuo aritmético fuera del propio análisis aritmético. Y, puesto que para la identificación de un punto en un espacio métrico tenemos que emplear coordenadas métricas, esto es, longitudes, la noción de medida tuvo que entretejerse consecuentemente en la trama original del continuo aritmético. De este modo, se dotó al continuo aritmético de las partes infinitamente divisibles que –como dijo Aristóteles– ha de poseer todo continuo. Divorciado del concepto de medida, es muy probable que el puesto del continuo aritmético (incluso con respecto a las restantes ramas de las matemáticas) hubiese estado en una caja de cristal para ser admirado exclusivamente como la creación más sublime, aunque perfectamente inútil, de la mente humana.

Otro desarrollo, que ha surgido recientemente a partir de algunas ideas latentemente arraigadas en el análisis aritmético, reivindica igualmente la postura de Aristóteles. Se trata del análisis de la dimensionalidad, que abiertamente reconoce la distancia insalvable que existe entre punto, línea, superficie, etc. Indudablemente, Aristóteles hubiera dicho no solamente que el punto no es parte de la línea sino también que la línea es el fin de una superficie, no parte de ella. O, para expresarlo de forma diferente, al nivel de la superficie la línea aparece como un indivisible. Si esta enseñanza fuese errónea, tendríamos que reflexionar sobre la cuestión de por qué el análisis aritmético no ha sido todavía capaz de transformar la propia dimensión en un concepto continuo, de modo que la dimensión $\sqrt{2}$, por ejemplo, existiese también como concepto significativo. Posiblemente, el mismo absurdo de una imagen con una dimensión entre la del punto y la de la línea explica por qué el análisis de la dimensionalidad ni siquiera se ha aventurado en esa dirección.

Por lo general, parece que el máximo obstáculo que se opone a todos nuestros intentos de adentrarnos en el misterio del concepto general de continuo con la ayuda de una estructura numérica es la imposibilidad de evitar por completo la discontinuidad. Como acabamos de ver, la discontinuidad aparece de forma inevitable en el análisis de la dimensionalidad; tampoco puede evitarse en ningún modelo aritmomórfico dirigido a colocar lo transfinito o lo infrafinito sobre una base aceptable para la Lógica. Entre los elementos sobre los que Cantor erigió su idea de lo transfinito, así como entre aquellos sobre los de descansa el sistema [ρ], ocupa un lu-

[33] Borel, *Probability and Certainty*, p. vii.

[34] Sobre ello, véase el Apéndice C en el presente volumen. A este respecto, voy a resaltar otro embrollo. Todos los tests estadísticos se basan en la hipótesis de que el coeficiente de probabilidad es un número *positivo*; no existe ningún test para el caso de acontecimientos cuasi-ciertos. Por consiguiente, las argumentaciones que invocan el fracaso de los tests estadísticos para apoyar la existencia de la Percepción Extra Sensorial son vanos: la Percepción Extra Sensorial puede ser posible, aunque con una probabilidad cero. (Huelga añadir que mi opinión no implica que este sea el caso real).

[35] Aristóteles, *Physics*, 231^b 15-16.

[36] *Ibid.*, 231^a 24-25. Véase también Immanuel Kant, *Critique of Pure Reason* (Edic. Everyman's Library, Nueva York, 1934), p. 136.

[37] Hobson, *Theory of Functions*, I, p. 89.

[38] Russell, *Mysticism and Logic*, p. 91.

gar destacado el tipo más diáfano de discontinuidad: el del sistema de números enteros. Posiblemente, este hecho es la consecuencia inevitable del pecado original de todos los intentos efectuados por reducir el continuo a su opuesto, el Número discretamente diferenciado. Quien insista en sembrar exclusivamente semillas de discontinuidad no debería maravillarse de la discontinuidad inherente a su cosecha, pero tampoco tendría que intentar negar su existencia.

APÉNDICE B

IGNORANCIA, INFORMACIÓN Y ENTROPÍA

1. Como ya hemos visto en el Capítulo V, Sección 4, la entropía de un sistema fue definida en primer lugar por Clausius como función de otras macrocoordenadas que pueden *medirse* directamente. Esa definición sigue siendo la única que nos permite determinar la entropía de un sistema real. Con el advenimiento de la termodinámica estadística, la entropía se definió de nuevo como función de las posiciones y velocidades de todas las partículas incluidas en el sistema (Capítulo VI, Sección 1). De acuerdo con esta nueva definición, la entropía puede *calcularse* a partir del conocimiento de esas microcoordenadas. Naturalmente, lo contrario no es cierto: dado el valor de la entropía de un sistema, no podemos deducir las posiciones y velocidades individuales. Sin embargo, nuestra ignorancia sobre el microestado real no es total, ni tampoco el grado de esta ignorancia es el mismo para cada valor de la entropía.

Tomemos un ejemplo muy sencillo, de cuatro partículas denominadas U, X, Y, Z y dos estados A y B. Consideremos el microestado U, X, Y en A y Z en B, y llamemos S a la entropía de este microestado. Dado que las microcoordenadas no dependen de las partículas específicas que se encuentren en cada estado, cada microestado en el que tres partículas están en A y la otra en B debe tener la misma entropía S. A partir del conocimiento de S, conocemos por tanto el macroestado; es decir, sabemos que hay tres partículas en A y una en B, pero no qué partícula específica hay en cada estado; sin embargo, sabemos que hay cuatro microestados que son compatibles con S. Y si tuviéramos que ocuparnos del microestado en el que U y X se encuentran en A e Y y Z en B, a partir del conocimiento de la correspondiente entropía S' sabríamos que hay seis microestados compatibles con S'. La idea trascendental de Boltzmann es que $S = k \ln 4$ y $S' = k \ln 6$[1].

Sobre esta base, G. N. Lewis afirmó en un trabajo de 1930 que la entropía de un sistema constituye un índice de nuestro grado de ignorancia

[1] Véase el Capítulo VI, fórmula (2).

sobre la microestructura del sistema. La idea es perfectamente razonable. Conociendo S, nos preguntamos de cuál de los *cuatro* microestados compatibles se trata realmente; conociendo S', el espectro de posibilidades aumenta a *seis* microestados[2]. Es evidente así que, al aumentar la entropía de nuestro sistema desde S a S', aumenta también nuestro grado de ignorancia –o nuestro grado de incertidumbre– sobre el microestado real. Como lo expresó Lewis: «El aumento en la entropía tiene lugar cuando una distribución *conocida* pasa a ser una distribución *desconocida*. La pérdida, característica de un proceso irreversible, es *pérdida de información*»[3].

Hay que resaltar bien ahora varios aspectos que son cruciales para la argumentación que sigue.

El primero es que el análisis precedente no implica en modo alguno que exista un índice del grado de ignorancia en cualquier otra situación, por ejemplo, si nos preguntamos si hay vida en Marte o si un producto químico recién sintetizado puede curar la fiebre del heno. Lo máximo que podemos deducir de ello es que un índice semejante puede construirse también para aquellos casos en los que podemos establecer cierto tipo de medida para cada alternativa posible.

En segundo lugar, no debemos ignorar el hecho de que, incluso si se cumple esta última condición, el grado de ignorancia no es una variable mensurable. El grado de ignorancia comparte las mismas dificultades analíticas de las nociones de orden (o desorden) en la termodinámica estadística o del nivel de precios y del producto nacional en economía. Todas esas variables no son mensurables ni siquiera en sentido ordinal; aceptan la relaciones «más» o «menos», pero únicamente si esas relaciones se toman dialécticamente. Como resultado de todo ello, lo más que podemos hacer es establecer pseudo medidas para cada una de ellas. Además, la gama de esas pseudo medidas es tan ilimitada como la de las medias, ya que la elección de una pseudo medida se encuentra igualmente limitada sólo por unas pocas condiciones. Y, debido precisamente a la esencia dialéctica de las pseudo medidas, no hay forma alguna de eliminar los casos en los que dos pseudo medidas de la misma variable arrojan clasificaciones totalmente distintas.

Una ilustración instructiva de las últimas observaciones es la sugerencia de O. Onicescu[4] de que se mida el orden (o la información) por lo que califica de «energía informativa»:

[2] En este punto, es importante recordar que sólo deben tenerse en cuenta los microestados que tienen la misma energía total (véase la nota 5, Capítulo VI). La cuestión es que, aunque la fórmula de Boltzmann da también el mismo S, $S = k \ln 4$, para cualquiera de los cuatro microestados en los que una partícula está en A y tres partículas en B, han de ignorarse esos microestados si el otro macroestado (consistente en tres partículas en A y una en B) representa la energía total dada.

[3] G. N. Lewis, «The Symmetry of Time in Physics», *Science*, 6 de junio de 1930, p. 573.

[4] Octav Onicescu, «Énergie informationnelle», *Comptes Rendus de l'Académie des Sciences*, Series A, CCLXIII (1966), pp. 841 y s. Véase la fórmula (8), Capítulo VI.

$$\mathscr{E} = \sum_1^s (N_i/N)^2 = \sum_1^s f_i^2, \tag{1}$$

donde $f_i = N_i/N$. Evidentemente, esta es una pseudo medida de orden tan buena que la podemos denominar ahora *negentropía* por partícula[5]:

$$\sum_1^s (N_i/N) \ln (N_i/N) = \sum_1^s f_i \ln f_i . \tag{2}$$

Al igual que H, $\mathscr{E}$ alcanza su mínimo para el microestado de menor orden, $f_1 = f_2 = \ldots = f_s$ (y sólo para éste), y su máximo para el microestado de mayor orden, $f_k = 1$ (y sólo para éste). Pero, como ya se ha dicho, no ordena en la misma forma en que lo hace H[6]. Sin embargo, como demostró Onicescu, $\mathscr{E}$ tiene unas propiedades tan interesantes como las de H[7]. Por ejemplo, si $f_{ik} = f_i' f_k''$ es una estructura compuesta, $\mathscr{E}(f) = \mathscr{E}(f')\,\mathscr{E}(f'')$. Por consiguiente, log $\mathscr{E}$ tiene la misma propiedad aditiva que H.

Una sugerencia interesante se deriva de la sencilla relación existente entre energía informativa y la desviación estándar de $(f_1, f_2, \ldots, f_s)$:

$$\mathscr{E} = \sum_1^s \left(f_i - \frac{1}{s}\right)^2 + \frac{1}{s}. \tag{3}$$

Esta relación nos lleva a recordar que el proceso a través del cual se alcanza el equilibrio termodinámico consiste en una progresiva difusión del calor disponible y, por tanto, de una progresiva reducción de las diferencias existentes entre los niveles de energía. Así pues, casi toda pseudo medida de dispersión puede servir de pseudo medida de orden[8]. De hecho, la propia función-H de Boltzmann es una pseudo medida de dispersión. Un álgebra relativamente sencilla demostrará estos puntos.

Sea $g(x)$ una función estrictamente convexa a lo largo del intervalo cerrado $[a,b]$, $a < b$, es decir, una función definida para cada $x \in [a,b]$ y tal que para todo $x, y \in [a,b]$ y $\alpha \in [0,1]$ tenemos

$$g[\alpha x + (1-\alpha)y] \leq \alpha\, g(x) + (1-\alpha)\, g(y), \tag{4}$$

donde la igualdad prevalece si, y sólo si, $x = y$ o si $\alpha = 0$. Sea $\alpha \leq x_1 \leq x_2 \leq \ldots \leq x_s \leq b$, y hagamos que $G = \Sigma_1^s g(x_i)$, $X_k = \Sigma_1^k x_i$ y $M_k = X_k/k$. A partir de (4), se deduce que

[5] Véase la fórmula (6), Capítulo VI.

[6] Dado que este punto es importante para comprender las peculiaridades de las pseudo medidas, puede ser útil proporcionar una demostración del mismo. Dado que $\Sigma p_i = 1$, tenemos:

$$d\mathscr{E} = \Sigma(p_i - p_s)\, dp_i, \qquad dH = \Sigma(\ln p_i - \ln p_s)\, dp_i.$$

Para que $d\mathscr{E}$ y dH tengan el mismo signo para cualquier dp_i, es necesario que los coeficientes de dp_i en esas sumas sean siempre proporcionales. Evidentemente, salvo en el caso de $s = 2$, esto no puede ser cierto para todos los valores p_i.

[7] Para las propiedades analíticas de H, véase C. E. Shannon y W. Weaver, *The Mathematical Theory of Communication* (Urbana, Ill., 1949), pp. 19-22.

[8] Digo «casi» porque $s > 3$ y el mayor orden que alcanzan los intercuartiles es cero.

$$g(x_i) \leq \frac{b - x_i}{b - a} g(a) + \frac{x_i - a}{b - a} g(b), \tag{5}$$

y, por adición,

$$\frac{G}{s} \leq \frac{b - M_s}{b - a} g(a) + \frac{M_s - a}{b - a} g(b), \tag{6}$$

donde la igualdad es válida si, y sólo si, es válida en (5) para todo i. A su vez, esta última condición es equivalente a

$$a = x_1 = x_2 = \ldots = x_j \quad x_{j+1} = x_{j+2} = \ldots x_s = b, \tag{7}$$

para cualquier j. También a partir de (4) y de

$$M_{k-1} \leq M_k = \frac{k-1}{k} M_{k-1} + \frac{1}{k} x_k \leq x_k, \qquad 1 < k \leq s, \tag{8}$$

obtenemos

$$g(M_k) \leq \frac{k-1}{k} g(M_{k-1}) + \frac{1}{k} g(x_k), \tag{9}$$

donde la igualdad prevalece si, y sólo si, $M_{k-1} = x_k$, en consecuencia, si, y sólo si,

$$x_1 = x_2 = \ldots = x_k. \tag{10}$$

Por simple inducción, la (9) da como resultado

$$g(M_s) \leq G/s, \tag{11}$$

donde la igualdad es válida si, y sólo si,

$$x_1 = x_2 = \ldots = x_s. \tag{12}$$

A partir de (6) y (11), obtenemos inmediatamente el siguiente resultado.

LEMA A: *Si los x_i están sujetos a la restricción $X_s = ta + (s - t)b$ para números enteros positivos dados s y t, $t < s$, G alcanza su máximo para (7) y su mínimo para (12).*

Consideremos el caso de $0 \leq x \leq 1$ y $X_s = 1$. Para $g(x) = x \log x$ (con $g(0) = 0$), obtenemos la propiedad de los extremos de H, y para $g(x) = x^2$, la de $\mathscr{E}$. Una extensa clase de funciones que tiene la misma propiedad corresponde a $g(x) = x^\alpha$, $\alpha > 1$. El caso de $g(x) = |x - (1/s)|^\alpha$, para $\alpha \geq 1$, da como resultado la clase familiar de momentos en torno a la media[9].

[9] El caso de $\alpha = 1$ puede demostrarse directamente examinando el signo de la diferencial total de G (como en la nota 6 anterior).

Sea $(x^0) = (x_1^0, x_2^0, \ldots x_s^0)$ un conjunto tal que $\Sigma_1^s x_1^0 = 1$ y $0 \leq x_i^0 \leq 1$. Dado k, $0 \leq k \leq s$, denotemos por (x^k) el conjunto tal que $x_i^k = x_i^0$ para $i \leq k$, y $x_i^k = (\Sigma_{k+1}^s x_j^0)/(s - k)$ para $i > k$. Evidentemente, $\Sigma_k^s x_j^k = \Sigma_k^s x_j^{k-1}$. De acuerdo con el Lema A, tenemos:

$$\sum_k^s g(x_j^{k-1}) \leq \sum_k^s g(x_j^k). \tag{13}$$

Si hacemos $G_k = \Sigma_1^s g(x_j^k)$, la (13) da como resultado

$$G_0 \leq G_1 \leq \ldots \leq G_{s-1} = G_s \tag{14}$$

relación que más tarde demostrará ser muy útil.

Si aceptamos la teoría estadística del calor y, sobre todo, si aceptamos el principio de que la fórmula de Boltzmann proporciona en cada caso el mismo valor que el valor de la entropía determinada experimentalmente por la fórmula de Clausius, es evidente que tenemos que preferir la función-H de Boltzmann a todas las restantes pseudo medidas de orden de la termodinámica. Pero, además, la función-H tiene una clara ventaja sobre las demás también en la teoría de la información, donde, como veremos ahora, se encuentra directamente relacionada con la frecuencia relativa de un acontecimiento específico[10].

2. Entre entropía e información existe, no obstante, una relación de carácter diferente al analizado antes. Se deriva del hecho de que no podemos obtener, transmitir o incluso almacenar información de cualquier tipo sin aumentar la entropía total del sistema aislado en el que actuamos. Para determinar la velocidad de una partícula, debemos proyectar sobre ella un rayo de luz, lo que producirá necesariamente una disipación de energía disponible y, por tanto, un aumento de la entropía total. El mismo resultado se produce con el ladrido de un perro que quiere informar a su amo que desea que se le deje en casa. Igualmente, un cajista aumenta la entropía total cuando compone, incluso aunque componga una secuencia de letras totalmente incomprensible. En general, si entre dos instantes de tiempo, $t_0 < t_1$, se obtiene o transmite cierta información de *cualquier tipo*, el aumento en la entropía, $S_1 - S_0$, puede descomponerse en dos partes: $S - S_0$, el aumento que se habría producido si no se hubieran realizado las operaciones necesarias para obtener o transmitir la información, y $S_1 - S$, que por definición representa el aumento en la entropía originado por tales operaciones. La relación

$$S_1 - S_0 = (S - S_0) + (S_1 - S) \tag{15}$$

es una consecuencia tautológica de la Ley de la Entropía. Este punto fue utilizado en un trabajo de L. Szilard de 1929 para refutar la paradoja del

[10] Relación (42) posterior.

demonio de Maxwell. Szilard afirmó que ningún demonio semejante podría llevar a cabo la tarea asumida sin obtener primero cierta información sobre las partículas objeto de su atención, es decir, sin aumentar primero la entropía total del sistema[11].

En cualquier caso, es preciso subrayar ahora un aspecto elemental: la relación (15) deja de ser una tautología si le damos la vuelta y decimos que $S_1 - S$ es una medida de «la cantidad de información» obtenida en virtud de ese aumento adicional de la entropía total. Evidentemente, podemos volver a convertirla en una tautología definiendo implícitamente «la cantidad de información» igual a $S_1 - S$, pero semejante definición implícita plantea numerosos problemas espinosos.

En primer lugar, nos obligaría prácticamente a decir que todos los términos en la relación (15) representan cantidades de información. Así, llevó a Lewis a concluir que «ganancia en entropía significa pérdida de información, y nada más»[12]. Ahora bien, por más que sus lazos sean fuertemente antropomórficos, la Ley de la Entropía es una ley de la Naturaleza expresable en términos puramente físicos. No obstante, de acuerdo con la vuelta dada por Lewis, deberíamos renunciar a contrastar esta ley en el laboratorio de los físicos midiendo las variables físicas implícitas en su definición clásica, recomendación que es difícil de aceptar. No puedo imaginar que un físico-químico discutiendo la estructura de una molécula de cierto compuesto químico o un ingeniero analizando un motor térmico encontrasen correcto decir que la entropía del correspondiente sistema no significa más que su propio grado de ignorancia.

En segundo lugar, la definición implícita modificaría la noción básica de información más allá del reconocimiento, mejor dicho, más allá de toda utilidad práctica. A saber, la transmisión de un mensaje totalmente sin sentido puede causar perfectamente un aumento de la entropía total mayor que el de un descubrimiento muy importante.

Por consiguiente, sería instructivo examinar con algún detalle el curso de las ideas que han conducido gradualmente a la postura de que la entropía y la información son entidades equivalentes.

3. En 1948, Norbert Wiener introdujo una definición específica de «la cantidad de información» con respecto a una distribución de probabilidad, contemplando el problema no *ex ante* (como lo hizo Laplace) sino *ex post*[13]. Como lo explicó, «si sabemos *a priori* que una variable se encuentra

[11] L. Szilard, «Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen», *Zeitschrift für Physik*, LIII (1929), pp. 840-856. En lo que se refiere al demonio de Maxwell, véase el Capítulo VII, Sección 7, anterior.

[12] Lewis, «Symmetry», p. 573.

[13] Norbert Wiener, *Cybernetics* (2.ª edic., Nueva York, 1961), pp. 61 y s. La idea básica se presentó mucho antes, en una reunión celebrada en 1927. Véase R. V. L. Hartley, «Transmission of Information», *Bell System Technical Journal*, VII (1928), pp. 535-544.

entre 0 y 1, y *a posteriori* que está en el intervalo (a,b) dentro de (0,1)», es bastante razonable considerar que cualquier función positiva y monótonamente decreciente de [medida de (a,b)/medida de (0,1)] es una medida ordinal de la cantidad de información *a posteriori*. Sucintamente,

$$\text{Cantidad de información} = F\left[\frac{medida\ de\ (a,b)}{medida\ de\ (0,1)}\right], \tag{16}$$

donde $F(x)$ es estrictamente decreciente con x. Sin embargo, dado que es razonable esperar que la (16) de como resultado el mismo valor para todos los intervalos iguales a (a,b), es necesario suponer que la variable relacionada con la (16) está uniformemente distribuida a lo largo de (0,1), en cuyo caso [medida de (a,b)]/[medida de (0,1)] es la probabilidad de que la variable se encuentre dentro de (a,b).

Otra forma de contemplar el problema es la siguiente. Se va a extraer al azar una carta de una baraja. En ese momento, hay cincuenta y dos interrogantes en nuestra mente. Si estamos previamente advertidos de que la carta extraída es una figura, desaparecen treinta y dos de tales interrogantes; sólo quedan veinte. Si se nos ha indicado que la carta es una figura de picas, quedan exclusivamente cinco interrogantes. Así, cuanto menor es la proporción de los interrogantes iniciales que quedan después de facilitar cierta información a una persona, mayor es la importancia (o la cantidad) de esa información. El principio general se hace así evidente: la cantidad de información $I(E)$ que ha producido el acontecimiento E de probabilidad p se mide ordinalmente por la fórmula

$$I(E) = F(p), \tag{17}$$

donde F es una función estrictamente decreciente que, por razones obvias, puede suponerse que satisface la condición $F = 0$ para $p = 1$. Podemos tomar, por ejemplo, $F = 1 - p^{\alpha}$. Wiener ha elegido el logaritmo negativo

$$I(E) = -\log p. \tag{18}$$

La elección tiene ventajas evidentes. Si en la (16) suponemos $a = b$, la información es extraordinariamente valiosa debido a que determina por completo la variable. Con la (18), el valor de la (16) es infinito. Si, por otra parte $(a,b) = (0,1)$, la información no nos dice nada que no sepamos ya. El valor de la relación (18) es en este caso cero, y todo está en orden[14].

[14] Voy a apuntar ahora que, en la medida en que hablamos del grado de la creencia ex ante en que ocurra un acontecimiento E de probabilidad p, toda función estrictamente creciente de p proporciona una medida ordinal de esa creencia. Además, como afirmó G. L. S. Shackle en *Expectations in Economics* (Cambridge, Ingl., 1949), cuanto mayor es el grado de la creencia ex ante, menor es el grado de sorpresa *tras* ocurrir E. El estrecho parentesco entre el grado de sorpresa ex post y la cantidad de infor-

Sin embargo, la ventaja más destacada de introducir el logaritmo se deriva de la transformación de la fórmula clásica para combinar acontecimientos

$$p(A \cap B) = p(A) \times p(B \mid A) \tag{19}$$

en una suma

$$\log p(A \cap B) = \log p(A) + \log p(B \mid A). \tag{20}$$

A partir de la (18), se deduce que las cantidades de información que vienen en sucesión son aditivas

$$I(A \cap B) = I(A) + I(B \mid A). \tag{21}$$

4. Todo está en orden, pero Wiener, con una argumentación muy oscura (en la que reconoció una sugerencia de J. von Neumann), concluyó que «una medida razonable de la cantidad de información» asociada a la densidad de probabilidad $f(x)$ es

$$\int_{-\infty}^{+\infty} [\log f(x)] f(x)\, dx, \tag{22}$$

y afirmó además que esta expresión es «el negativo de la cantidad habitualmente definida como entropía en situaciones similares»[15]. En la argumentación de Wiener hay tanto una analogía espuria como un error elemental de análisis matemático. No hay que sorprenderse, por tanto, de que el problema de la relación entre la función-H de Boltzmann y la cantidad de información se encuentre lejos de estar resuelto incluso tras tantos años transcurridos.

El hecho de que la función logarítmica aparezca tanto en la relación (18) como en la (22) no constituye motivo suficiente para considerar que la (22) representa, también, una medida de cantidad de información. Curiosamente, Wiener no vio que en la (18) tenemos el logaritmo de una *probabilidad*, mientras que en la (22) el logaritmo se aplica a la *densidad de probabilidad*, y, como voy a mostrar ahora, la (22) no puede considerarse en modo alguno la forma continua de la función-H. Más aún, el concepto de entropía tal como lo definió Boltzmann –es decir, por la función-H (2)– no puede extenderse a una distribución continua.

Podemos comenzar resaltando que, de acuerdo con la definición de Wiener (18), en la actualidad generalmente aceptada, no tenemos derecho a hablar de cantidad de información si no nos referimos *al acontecer de un acontecimiento estocástico*: la extracción de una figura de una baraja, el disparar dentro del segundo círculo que rodea a una diana, etc. Así pues, debemos preguntar: ¿cuál es el acontecimiento *acaecido* que puede asociarse a

mación es obvio. Por consiguiente, cualquier fórmula de la cantidad de información es también una medida del grado de sorpresa, y viceversa.

[15] Wiener, p. 62.

una distribución de probabilidad? La respuesta es que no hay ninguno. Sin embargo, existen diferentes vías por las que puede establecerse una relación entre información –entendida igualmente en sentido muy restringido, como en el caso de la (18)– y la función-H. En realidad, hay una amplia clase de funciones para las que esto es cierto. Voy a proseguir todavía la argumentación relativa al caso general a fin de dejar perfectamente claro que el problema no implica necesariamente el concepto de entropía de Boltzmann.

Sea $E_1, E_2,\ldots,E_s$ un conjunto de acontecimientos mutuamente excluyentes y completamente exhaustivos, de probabilidades $p_1, p_2,\ldots,p_s$, $\Sigma\, p_i = 1$. En virtud de la (17), cuando (y si) tiene lugar el acontecimiento E_i, la cantidad de información que se ha producido será $F(p_i)$. Ahora bien, puesto que no sabemos todavía que acontecimiento tendrá lugar, no podemos recurrir más que a un cálculo racional de la cantidad *futura* de información. Un camino muy trillado nos lleva a la cantidad *esperada* de información:

$$\phi_F(p) = \sum_1^s p_i F(p_i). \tag{23}$$

De forma alternativa, podemos interpretar ϕ_F como el grado esperado de sorpresa causado por uno de los futuros acontecimientos de la misma distribución de probabilidad[16]. En ambos casos, ϕ_F es una estimación *ex ante* de una coordenada *ex post*. Analíticamente, $\phi_F(p)$ es de hecho un *estadístico* peculiar de una distribución y no, como ya se ha dicho, una característica de un simple acontecimiento. Y digo «peculiar», porque ϕ_F es un estadístico que implica *exclusivamente* a las probabilidades de una distribución.

Dado un campo de acontecimientos estocásticos, podemos dividirlo en un gran número de formas; y es evidente que el valor de ϕ_F depende de la manera en que hayamos dividido el campo: por consiguiente, *ϕ_F no es una característica invariante de un campo estocástico*. A efectos de sencillez, supongamos que $F(p) = 1 - p$ y consideremos el campo de cartas extraídas de una baraja estándar en virtud de un mecanismo *aleatorio insesgado*[17]. Si se considera a cada carta como un acontecimiento separado,

$$\phi_F(p) = 1 - \left(\frac{1}{52}\right) = \frac{51}{52}, \tag{24}$$

y si el campo está dividido en «figuras» y «cartas bajas»,

$$\phi_F(p) = \frac{80}{13^2}. \tag{25}$$

[16] Véase la nota 14 anterior.

[17] Puesto que $F = 1 - p$, tenemos $\phi = 1 - \mathscr{E}$.

El hecho de que ϕ_F es mayor para la primera y más fina división es una propiedad común a la amplia clase de funciones $F(p)$. Únicamente necesitamos suponer que $h(p) = pF(p)$ es estrictamente cóncava[18] y que $h(0) = 0$. Transformemos ahora la división (p) en (p') de modo que todo p_k esté dividido en una suma

(26) $$p_k = p'_j + p'_{j+1}, + \dots + p'_{j+i}, \qquad 0 \leq i.$$

Sean x e y tales que $0 < x < y < x + y < 1$. A partir de esta ordenación y de la condición de estricta concavidad, obtenemos

(27) $$h(x) \geq \frac{y-x}{y} h(0) + \frac{x}{y} h(y),$$

$$h(y) \geq \frac{x}{y} h(x) + \frac{y-x}{y} h(x+y).$$

Estas desigualdades dan como resultado

(28) $$h(x) + h(y) \geq h(x+y),$$

que paso a paso conduce a

(29) $$\Phi_F(p) \leq \Phi_F(p').$$

Esta propiedad (compartida en especial por $-H$ y $1 - \mathscr{E}$) expresa el hecho de que una clasificación más fina es siempre capaz de obtener mayor información[19], pero no debiéramos confundirnos con nuestras operaciones teóricas sobre el papel, aquí o en cualquier otra parte: el hecho no está demostrado por la (29); ésta únicamente confirma lo apropiado de nuestras formalizaciones.

Consideremos ahora la densidad de probabilidad $f(x)$ de una distribución absolutamente continua. Dado que

(30) $$\int_{-\infty}^{+\infty} f(x)\,dx = 1,$$

podemos encontrar n intervalos $-\infty < x_1 < x_2 < \dots < x_{n-1} < +\infty$ tales que la probabilidad en cada uno de ellos sea $1/n$. Para esta división, la (23) se convierte en

(31) $$\Phi_F(n) = F\left(\frac{1}{n}\right).$$

[18] Una función $h(x)$ es estrictamente cóncava si $g(x) = -h(x)$ es estrictamente convexa.

[19] Por una clasificación más fina entendemos aquí no sólo un mayor número de clases (que pueden solaparse con las iniciales) sino también una mayor división de las clases iniciales, como lo pone de manifiesto la (26).

A partir de la (27), se deduce que

(32) $$\frac{h(x)}{x} > \frac{h(y)}{y}$$

para cualquier x, $0 < x < y$. Por lo tanto, $F(1/n)$ tiene un límite, finito o infinito, para $n \to \infty$. Si lo designamos por F_0, la (31) da

(33) $$\lim_{n \to \infty} \Phi_F(n) = F_0.$$

Se trata aquí de un interesante resultado, pues demuestra que *la cantidad esperada de información en una distribución absolutamente continua depende exclusivamente de la medida ordinal adoptada –más exactamente, del* lim *$F(p)$ para $p \to 0$- y no de la propia distribución.*

Por ejemplo, si $F(p) = 1 - p$ (la fórmula de Onicescu modificada), resulta

(34) $$\lim_{n \to \infty} \Phi_F(n) = 1.$$

Para la entropía, es decir, para $F(p) = -\ln p$, tenemos

(35) $$\lim_{n \to \infty} \Phi_F(n) = +\infty,$$

lo que demuestra mi anterior opinión de que la función-H de Boltzmann no puede extenderse a una distribución continua.

5. Existe una segunda (y creo que nueva) forma de relacionar la función-H o su generalización con la información. Sean $A_1, A_2, \dots, A_s$, s individuos, cada uno de los cuales posee su propia cantidad de tierra, x_i. Supongamos que primeramente no conocemos más que la cantidad total de tierra, $X = \Sigma x_i$. La única imagen que podemos tener de la distribución de la tierra es aquella a la que llegamos a través del Principio de la Razón Insuficiente, esto es, que cada uno posee la misma cantidad de tierra, X/s. Si posteriormente se nos informa de la cantidad que posee A_1, la imagen correspondiente a los otros será aquella en la que cada uno posee $(X - x_1)/(s - 1)$. Si se llega a conocer también x_2, nuestra conjetura racional es que cada uno de los otros posee $(X - x_1 - x_2)/(s - 2)$. Cuando por último se llega a conocer x_{s-1}, se conoce toda la distribución.

Lo que necesitamos ahora es una función de la distribución de X tal que aumente (o disminuya) según se conozca progresivamente la distribución real. En virtud de la (14), la función $G = \Sigma x_i g(x_i)$ satisface esta condición. Por consiguiente, puede tomarse como una medida de la información que tenemos sobre la distribución de X. Si la distribución real es *uniforme*, entonces $G_0 = G_s$, lo que es totalmente razonable: la información completa no modifica en modo alguno la imagen que teníamos en mente antes de disponer de cualquier información[20].

[20] No es preciso añadir que H y $\mathscr{E}$ poseen la propiedad (14) y, por tanto, las consideraciones anteriores son aplicables a la información disponible sobre una distribución *discreta* de probabilidades.

Es obvio que los resultados precedentes son consistentes si cada x_i se sustituye por la participación relativa $\xi_i = x_i/X$, en cuyo caso $\Sigma\ \xi_i = 1$. Y si recordamos las proposiciones demostradas en la Sección 1 anterior, relativas a los extremos de $G = \Sigma\ \xi_i\ g(\xi_i)$, vemos que esta forma generalizada de la función-H puede servir también como medida de *concentración.*

6. Otra nueva forma de conectar la entropía con la información se debe a C. E. Shannon, quien, casi al mismo tiempo que Wiener, la presentó en una memoria clásica sobre teoría de la *comunicación.* A diferencia de Wiener, Shannon quería llegar a una medida de la capacidad (o de la potencia) de un sistema de códigos para transmitir o almacenar *mensajes.* También a diferencia de Wiener, Shannon no se ocupó de si un mensaje contiene información valiosa alguna[21]. Para un especialista en comunicación, esto es perfectamente comprensible: a efectos prácticos, el coste de transmitir un mensaje es independiente de si el mensaje tiene una importancia vital para todo el mundo o carece completamente de sentido. El problema básico de la comunicación es saber qué código tiene la mayor capacidad «de transmitir información»[22]. Shannon acentúa desde el principio el cambio en el significado de «información»: «el número de mensajes... o cualquier función monotónica de este número puede contemplarse como una medida de la información producida cuando se elige un mensaje a partir del conjunto» de todos los mensajes de la misma longitud[23].

El número de mensajes diferentes consistentes en N señales del código binario –N puntos o rayas– es 2^N. En general, si el código consiste en s señales diferentes, el número de mensajes es S^N. Siguiendo la sugerencia hecha por R. V. L. Hartley (citada en la nota 13), Shannon tomó el logaritmo de este número como medida de la capacidad de información. Además, en vista del importante papel desempeñado por el sistema binario en los sistemas electrónicos de transmisión y almacenamiento, parecía natural elegir el logaritmo en base 2. Así pues, la información de Shannon para un mensaje de N señales binarias es sencillamente

$$\log_2 2^N = N \tag{36}$$

en unidades denominadas «bits», abreviatura de «unidad binaria»[24]. Para $s > 2$, la misma información se mide por $N \log_2 s > N$. Así, la información de Shannon por señal es un bit para el código binario y $\log_2 s$ bits para el caso general. Su importante papel en la teoría de la comunicación se deriva del hecho de que es independiente de la longitud del mensaje.

[21] Shannon y Weaver, *Mathematical Theory of Communication*, p. 3.
[22] *Ibid.*, especialmente pp. 7 y 106.
[23] *Ibid.*, p. 3.
[24] *Ibid.*, pp. 4 y 100.

El caso de mensajes transmitidos en un lenguaje ordinario es más complicado, ya que no todas las secuencias de signos constituyen mensajes. Una larga secuencia de la misma carta, por ejemplo, no tiene significado alguno en ningún lenguaje; por consiguiente, no debe incluirse al medir la capacidad de información de un lenguaje. Para llegar a la fórmula correspondiente a este caso, Shannon buscó una función que satisficiera ciertas condiciones analíticas razonables[25]. Sin embargo, la misma fórmula puede alcanzarse por una vía directa que tiene el mérito de precisar con exactitud la razón por la que esa fórmula es idéntica a la función-H de Boltzmann.

Aceptamos como un hecho dado que la frecuencia relativa con la que cada signo escrito (una letra, un signo de puntuación, o un espacio en blanco) aparece en todo lenguaje tiene un límite pseudo *ergódico.* Si p_1, p_2, ..., p_s designan estos límites de frecuencia[26], un mensaje *típico* de N signos debe contener $N_1 = p_1N$, $N_2 = p_2N$, ..., $N_s = p_sN$ signos de cada tipo. El número total de mensajes típicos viene dado por la conocida fórmula combinatoria

$$W = \frac{N!}{N_1!N_2!\ldots N_s!}. \tag{37}$$

El misterio se ha revelado ahora: la relación (37) es la misma fórmula a partir de la cual dedujo Boltzmann su función-H para N muy grande:

$$\ln W = -N\ \Sigma\ p_i \ln p_i \tag{38}$$

que es la relación (4) del Capítulo VI anterior. Así pues, la información de Shannon por señal es

$$(\ln W)/N = -H, \tag{39}$$

que nuevamente se observa es independiente de N.

Debemos tener en cuenta también que la frecuencia relativa de los mensajes típicos entre todos los mensajes de longitud N es

$$P = Wp_1^N 1 p_2^N 2 \ldots p_s^N S = Wp, \tag{40}$$

donde p es la frecuencia de cualquier mensaje típico dado. Esto da como resultado

$$\ln P = \ln W + \ln p. \tag{41}$$

[25] *Ibid.*, pp. 18-20 y 82 y s.
[26] El motivo por el que rehuso referirme a estos coeficientes como «probabilidades» debería quedar claro a partir de lo que dije en el Capítulo VI, Sección 3. Es cierto que en un lenguaje las letras no se siguen unas a otras de acuerdo con una norma fija, pero tampoco tienen lugar de forma aleatoria, como los puntos en un lanzamiento de dados.