
POPPER, Karl R. *La
responsabilidad de vivir.
Escritos sobre política,
historia y conocimiento.*
Paidós, Barcelona, 1995
pp. 17-79

PRÓLOGO

La presente colección de artículos y conferencias se puede considerar como una continuación de mi libro *En busca de un mundo mejor* [Barcelona, Paidós, 1994]. Ambos libros contienen algunas colaboraciones drásticamente orientadas al campo de las ciencias naturales y otras que están orientadas histórica o políticamente. El título de este libro, *La responsabilidad de vivir*, coincide con el título de su capítulo 12 -un capítulo que influyó de manera decisoria en el corto pero pertinentemente importante «resumen como prólogo» que incluí en un volumen colectivo anterior.

Me he esforzado por conceder también al prefacio de la presente recopilación más peso del que generalmente suelen tener los prólogos.

La selección de los capítulos es fruto del trabajo conjunto de mi asistente, la señora Melitta Mew, y del doctor Klaus Stadler de la editorial Piper. A los dos me obliga el mayor agradecimiento.

I

«Cuestiones de conocimiento natural» reza el título de la primera parte del presente libro. Aquí me estoy refiriendo fundamentalmente a la biología y a la inconmensurable riqueza de las formas de la vida.

Cuanto más penetremos en alguno de los muchos dominios de la biología -no importa desde qué perspectiva-, tanto más inconcebible se manifiesta la riqueza de formas de las estructuras biológicas en cada nivel y tanto más maravilloso aparece su armónico funcionamiento conjunto.

El último capítulo de la primera parte está dedicado a Johannes Kepler, quien con mayor ahínco buscó la armonía en la obra creadora física de Dios y vio recompensada su búsqueda con el hallazgo de las tres leyes, que determinan los movimientos de los planetas de manera muy abstracta, pero también sumamente armónica, y que llevan su nombre. Entre los tres gigantes del espíritu -los coetáneos Galileo y Kepler y su sucesor Newton-, los cuales han creado conjuntamente (y con otros) nuestra ciencia de la naturaleza, quizá sea Kepler el más grande. Es con toda seguridad la personalidad más atractiva, más sincera y más modesta. Los tres eran investigadores apasionados y trabajadores infatigables; los tres llevaban a cabo un durísimo trabajo, a menudo decepcionante durante mucho tiempo, pero los recompensó la gran suerte del descubridor, que ve el mundo bajo una nueva luz: distinto, más hermoso, armónico e incluso mejor que cualquier ser humano antes que él, y sabe entonces que su duro trabajo ha sido bendecido por la suerte, por una casi inmerecida suerte, pues muy fácilmente hubiera podido suceder de otra manera.

Únicamente Kepler, entre estos tres grandes mencionados, no sólo se conformó con cumplir su cometido, sino que también lo consignó todo por escrito con solicitud y franqueza. Y también entendió como ningún otro que los pensadores griegos de la antigüedad -desde Tales a Aristóteles, Aristarco y Tolomeo- fueron originariamente los que habían legado sus osadas ideas al predecesor de Kepler, a Copérnico.

Su gran modestia ayudó a Kepler -más que a los otros dos-a ser consciente una y otra vez de sus errores, pudiendo aprender así de ellos; errores que sólo podían superarse con las mayores dificultades. Estos tres gigantes espirituales se encontraban profundamente presos, cada uno a su manera, de una superstición. («Superstición» es un término que debemos emplear con muchísima precaución: con el conocimiento de que sabemos bien poco y de que es seguro que nosotros mismos, sin ser conscientes de ello, estemos atrapados en diferentes formas de superstición.) Galileo se encontraba profundamente anclado en la creencia en un movimiento circular natural -precisamente la creencia que Kepler derrotó tras largos combates consigo mismo y en el seno de la astronomía. Newton escribió un voluminoso libro sobre la historia tradicional (fundamentalmente bíblica) de la humanidad, que él se encargó de corregir con arreglo a principios deducidos en su época muy ostensiblemente de la superstición. Y Kepler no fue sólo un astrónomo, sino también un astrólogo -razón por la que Galileo y muchos otros lo rechazaron.

Pero la astrología -la superstición de Kepler- fue combatida por él mismo en sus manifestaciones dogmáticas: fue un astrólogo autocrítico. Enseñaba que el destino anticipado en las estrellas no es inevitable, sino que se puede eludir por medio de nuestra voluntad ética. Se trata de una concesión importante a los críticos de la astrología. Quizá fuera él, entre los tres grandes citados, el menos dogmático en su superstición.

Kenley, 12 de julio 1994

K.R.P.

PRIMERA PARTE

CUESTIONES DE CONOCIMIENTO NATURAL

Capítulo 1

LA TEORÍA DE LA CIENCIA DESDE UN PUNTO DE VISTA TEÓRICO-EVOLUTIVO Y LÓGICO¹

La tesis fundamental que quisiera someter a su consideración en esta conferencia, se puede formular de la siguiente manera:

Tanto las ciencias naturales como las ciencias sociales parten siempre de *problemas*; de que algo despierta nuestra *admiración*, como decían los filósofos griegos. Las ciencias utilizan en principio para resolver esos problemas el mismo método que emplea el sano entendimiento humano: el método de *ensayo y error*. Expresado con más exactitud: es el método de proponer *tentativamente* soluciones de nuestro problema y después eliminar las falsas soluciones como erróneas. Este método presupone que trabajamos con una pluralidad de soluciones *a modo de prueba*. Una solución tras otra es puesta a prueba y eliminada.

Bien mirado, ese comportamiento parece ser el único lógicamente posible. Es el mismo comportamiento que ponen en práctica organismos inferiores, tan inferiores incluso como la ameba unicelular, cuando intenta solucionar un problema. En ese caso, hablamos de movimientos de prueba, por medio de los cuales el organismo intenta desembarazarse de un problema gravoso. Los organismos más elevados pueden *aprender*, por medio del ensayo y error, cómo se soluciona un determinado problema. Podemos decir que también ellos realizan movimientos de prueba -movimientos de prueba intelectuales; y el aprender consiste, en lo esencial, en que se va ensayando un movimiento de prueba tras otro, hasta que se encuentra uno que resuelve el problema. Podemos comparar la solución exitosa del animal con una *expectativa* y, por consiguiente, con una *hipótesis* o una *teoría*; pues el comportamiento del animal nos muestra que espera (aunque quizá sea de manera inconsciente o disposicional) que en un caso semejante los mismos movimientos de prueba volverán a resolver ese problema.

Podemos decir que el comportamiento de los animales, y también de las plantas, muestra que los organismos están preparados para las regularidades o legalidades. *Esperan* regularidades o legalidades en su entorno y la mayoría de estas expectativas son, supongo yo, genéticamente condicionadas, es decir, innatas.

Para el animal surge un *problema* cuando se frustra una expectativa. Esto le conduce a realizar movimientos de prueba y, por consiguiente, a ensayar posibilidades para sustituir la expectativa frustrada por una nueva.

Si un organismo superior es decepcionado demasiado a menudo en sus expectativas, se derrumba. No puede resolver el problema, se hunde.

Quisiera resumir lo que he dicho hasta ahora sobre el aprendizaje por medio del ensayo y error en un esquema que se desarrolla en tres fases. El esquema consta de las etapas siguientes:

- a) el problema;
- b) los intentos de solución;
- c) la eliminación.

¹ Conferencia radiofónica para la NDR, 7 de marzo de 1972.

La etapa a) de nuestro esquema es, pues, el *problema*. El problema surge cuando entra en escena una perturbación; una perturbación, bien de las expectativas innatas, bien de las expectativas que habían sido descubiertas o aprendidas por medio del ensayo y el error.

La etapa b) de nuestro esquema la constituyen los *intentos de solución*, esto es, los ensayos para *solucionar el problema*.

La etapa c) de nuestro esquema es *la eliminación* o la supresión de los intentos de solución que no han tenido éxito.

Es esencial para este esquema que sea plural: la primera etapa, el problema, puede presentarse en singular, pero no así el segundo nivel, al que he denominado *ex profeso* en plural «intentos de solución». Ya en el experimento con animales hablamos de movimientos de prueba; y el término «movimientos de prueba» es plural. Tendría poco sentido denominar movimiento de prueba a un movimiento aislado.

Etapa b), los intentos de solución son, por consiguiente, movimientos de prueba y, por eso, se citan en plural; se someten en la tercera etapa de nuestro esquema al *procedimiento de eliminación*.

Etapa c), *la eliminación es negativa*: la eliminación es en lo esencial una eliminación de *errores*. Si se elimina un intento de solución fracasado o equivocado, el problema se queda sin resolver y da motivo para nuevos intentos de solución.

Pero, ¿qué sucede si un intento de solución tiene finalmente éxito? Suceden dos cosas distintas: primera, el intento de solución que ha tenido éxito se *aprende*; esto sucede en animales en general, de forma que -si vuelve a surgir un problema semejante- se repiten brevemente y a grandes rasgos los anteriores movimientos de prueba, incluido el exitoso, en el orden de sucesión original; se recorren hasta el intento de solución con éxito.

El aprendizaje consiste en que los intentos de solución fracasados o eliminados van decayendo progresivamente en el nivel de sugerencias, de forma que, finalmente, el intento de solución exitoso entra en escena casi solo. En esto consiste el procedimiento de eliminación, que se basa esencialmente en una pluralidad de intentos de solución.

Podemos decir que el organismo ha aprendido de esta manera una nueva *expectativa*. Podemos describir su comportamiento de la siguiente manera: el organismo espera que el problema se solucione por medio de movimientos de prueba, y, ciertamente, en definitiva por medio del último, por medio del movimiento de prueba que ya no es eliminado.

El desarrollo de esa expectativa en el organismo se corresponde en el nivel teórico-científico, como veremos enseguida, con la formación de hipótesis o teorías. Pero antes de pasar a comentar la formación de teorías científicas, quisiera llamar la atención todavía sobre otra aplicación biológica de mi *esquema en tres etapas*:

- a) el problema;
- b) intentos de solución;
- c) eliminación,

que se puede concebir también como esquema de la teoría evolutiva darwinista. No sólo es aplicable a la evolución de los organismos individuales, sino también al *origen de las especies*. Una modificación de las condiciones medioambientales o también de la estructura interna del organismo produce, de acuerdo con nuestro esquema en tres etapas, un problema. Se trata de un problema de adaptación para la especie; es decir, la especie sólo puede continuar existiendo si se resuelve el problema por medio de un

cambio en su estructura genética. ¿Cómo sucede esto según la concepción del darwinismo? Nuestro aparato genético está construido de tal forma que siempre aparecen nuevas modificaciones o mutaciones en la estructura genética. El darwinismo admite que esas mutaciones actúan como *intentos de solución*, en el sentido del punto b) de nuestro esquema en tres etapas. La mayoría de las mutaciones son fatales: son mortales para los portadores de la mutación, para el organismo en el que hacen su aparición. Con ello, pues, son *eliminados*, de acuerdo con el punto c) de nuestro esquema. Por lo tanto, en nuestro esquema en tres etapas, tenemos que aludir de nuevo al pluralismo esencial de la segunda etapa, de los intentos de solución. Si no hubiera *muchísimas* mutaciones, no entrarían en absoluto en cuestión como *intentos de solución*. Tenemos que admitir que una *mutabilidad* bastante grande es esencial para el funcionamiento de nuestro aparato genético.

Ahora ya puedo centrar por fin mi atención sobre mi tema fundamental, la teoría de la ciencia o la lógica de la ciencia.

Mi primera tesis al respecto es que la ciencia es un fenómeno biológico. La ciencia ha surgido a partir del conocimiento precientífico, es un perfeccionamiento sumamente notable de la forma de conocimiento del entendimiento humano sano, que se puede concebir, por su parte, como un perfeccionamiento del conocimiento animal.

Mi segunda tesis es que nuestro esquema en tres etapas también es aplicable a la *ciencia*.

Ya he mencionado al principio que las ciencias, como ya vieron los filósofos griegos, parten del *problema* de la *admiración* por algo, que en sí mismo puede ser algo cotidiano, pero que para el pensador científico se convierte en admiración, en problema. Mi tesis es que todo desarrollo científico sólo puede entenderse admitiendo que su punto de partida es un *problema* o una *situación problemática*, es decir, el surgimiento de un problema en una situación determinada de nuestro saber total.

Este punto es de gran importancia. La teoría científica antigua enseñaba -y lo sigue enseñando- que el punto de partida de la ciencia es nuestra percepción sensorial o la observación sensible. Esto suena, a primera vista, razonable y convincente, pero es absolutamente falso. Se puede mostrar fácilmente por medio de la siguiente tesis: *sin problema, ninguna observación*. Si les pido: «¡Por favor, observen!», deberían preguntarme, con arreglo al uso del lenguaje: «Sí, ¿pero qué? ¿Qué debo observar?». Con otras palabras, me piden que les proporcione un *problema* que se pueda resolver por medio de su observación; y si no les proporciono ningún *problema*, sino sólo un *objeto*, esto está ya mucho mejor, pero no es satisfactorio en modo alguno. Si les digo, por ejemplo: «Por favor, observen ustedes su reloj», continuarán sin saber lo que realmente quiero que observen. Pero si les planteo un *problema* absolutamente trivial, la cosa cambia. Puede que no se interesen por el problema, pero al menos saben lo que deben comprobar por medio de su percepción u observación. (Como ejemplo, podrían plantearse el problema de si la luna está en cuarto menguante o en cuarto creciente; o en qué ciudad ha sido impreso el libro que leen en este momento.)

¿Cómo llegó la antigua teoría científica a la falsa idea de que en la ciencia partimos de percepciones sensoriales o de observaciones, en lugar de partir de problemas?

La antigua teoría científica dependía en este punto de la teoría del conocimiento del entendimiento humano sano. Éste nos dice, en efecto, que nuestro saber acerca del mundo exterior depende siempre de las impresiones de nuestros sentidos.

En general, soy un gran admirador del entendimiento humano sano; supongo, incluso, que, sólo con que seamos un poco críticos, el entendimiento humano sano es el consejero más valioso y más

seguro en todas las situaciones problemáticas posibles. Pero no siempre es seguro; y cuando se llega a cuestiones teórico-científicas o teórico-gnoseológicas, es de la mayor importancia enfrentarse a él de manera realmente crítica.

Desde luego, es cierto que nuestros órganos sensoriales nos informan sobre nuestro medio ambiente y que los necesitamos - imperiosamente para ese fin. Pero no podemos extraer de ello la conclusión de que nuestro conocimiento comienza con la percepción sensorial. Por el contrario, nuestros sentidos son, desde un punto de vista teórico evolutivo, herramientas que se han desarrollado para resolver determinados *problemas* biológicos. De forma que los *ojos* de los animales y de los humanos se han desarrollado aparentemente para que los seres vivos que cambian de lugar y que pueden moverse, se pongan en guardia a tiempo contra choques peligrosos con cuerpos duros, en los que se pueden lesionar. Desde un punto de vista teórico evolutivo, nuestros órganos sensoriales son el resultado de problemas e intentos de solución, exactamente igual que nuestros microscopios o nuestros catalejos. Y esto muestra que el problema, desde el punto de vista biológico, aparece *antes* que la observación o percepción sensorial: las observaciones o las percepciones sensoriales son medios importantes de nuestros *intentos de solución* y juegan su papel fundamental en la *eliminación*. Por consiguiente, mi esquema en tres etapas es aplicable a la lógica científica o metodología de la siguiente manera:

1. El punto de partida es siempre un *problema* o una situación problemática.
2. Luego siguen los *intentos de solución*. Éstos consisten siempre en teorías, y esas teorías son, puesto que son ensayos, muy a menudo erróneas: son y serán siempre hipótesis o suposiciones.
3. También en la ciencia aprendemos por medio de la *eliminación* de nuestras falsas teorías.

Nuestro esquema en tres etapas:

- a) *problema*;
- b) *intentos de solución*;
- c) *eliminación*,

es aplicable, por lo tanto, a una descripción de la ciencia.

Con ello llegamos a nuestra cuestión central:

¿Dónde reside lo específico de la ciencia humana? ¿Cuál es la diferencia decisiva entre una ameba y un gran científico como Newton o Einstein?

La respuesta a esta cuestión es: lo específico de la ciencia reside en la aplicación consciente del *método crítico*; en la etapa c) de nuestro esquema, la de la eliminación de nuestros errores, procedemos conscientemente de forma crítica.

Sólo el método crítico explica el crecimiento extraordinariamente rápido de la forma del saber científico, el extraordinario progreso científico.

Todo conocimiento precientífico, ya sea animal o humano, es *dogmático*; y con la invención del método científico, es decir, del método crítico, comienza la ciencia.

La invención del método crítico presupone en cualquier caso un *lenguaje* humano descriptivo y un lenguaje en el que se pueden desarrollar *argumentos* críticos. El método crítico presupone posiblemente incluso una escritura. Pues el método crítico consiste esencialmente en que nuestros intentos de solución, nuestras teorías y nuestras hipótesis se nos puedan *presentar* objetivamente, lingüísticamente formuladas, de forma que puedan convertirse en *objetos de una investigación crítica consciente*.

Es muy importante comprender la enorme diferencia que existe entre un pensamiento meramente subjetivo o pensado privadamente o tenido por cierto, una imagen disposicional psicológica, y el *mismo* pensamiento una vez formulado lingüísticamente (o quizá incluso por escrito), con lo que se puede someter a la consideración de la discusión pública.

Mi tesis es que se trata de un paso enormemente importante, un paso por así decir sobre un abismo, el que va desde mi pensamiento no pronunciado: «Hoy lloverá» al mismo pensamiento, pero expresado en una proposición «Hoy lloverá». En un primer momento, este paso, la expresión de un pensamiento, no parece ser en absoluto un paso tan grande. Pero la formulación lingüística significa que algo, que antes era parte de mi personalidad, de mis expectativas y quizá de mis temores, ahora se presenta objetivamente y, con ello, se hace accesible a la discusión crítica general. Y la distinción es inmensa también para mí mismo. La proposición expresada, por ejemplo, la predicción expresada, se separa de mi persona por medio de la formulación lingüística. Con ello, se hace independiente de mis estados de ánimo, esperanzas y temores. Se ha *objetivado*: la pueden afirmar *tentativamente* otros, pero también yo mismo, al igual que se puede negar *tentativamente*; las razones en pro y en contra se pueden sopesar y discutir; y puede llegarse a una formación de partido pro y contra la predicción.

Aquí llegamos a una importante distinción entre dos significados del término «saber» -saber en sentido subjetivo y objetivo. De ordinario se considera el saber como un estado subjetivo o espiritual. Se parte de la forma verbal «yo sé» y se explica el saber como una determinada especie de creencia que descansa en *razones suficientes*. La interpretación subjetiva del término «saber» ha influido demasiado la antigua teoría científica, pero es completamente inservible para una teoría de la ciencia, puesto que la ciencia consiste en proposiciones objetivas, formuladas lingüísticamente, en hipótesis y en problemas, y no en expectativas o convicciones subjetivas.

La ciencia es un producto del espíritu humano, pero ese producto es tan objetivo como una catedral. Si se dice que una proposición es un pensamiento lingüísticamente expresado, esto es ciertamente correcto, pero no descubre la objetividad de la proposición con el suficiente rigor. Esto está conectado con la ambigüedad del término «pensamiento». Como han subrayado especialmente los filósofos Bernhard Bolzano y, después de él, Gottlob Frege, hay que distinguir el proceso de pensamiento subjetivo del contenido objetivo o del contenido lógico² o informativo de un pensamiento. Cuando digo: «Los pensamientos de Mahoma son muy diferentes de los de Buda», entonces no hablo sobre los procesos mentales de los dos hombres, sino sobre el contenido lógico de las dos doctrinas o teorías.

Los procesos mentales pueden guardar una relación causal. Cuando digo: «La doctrina de Spinoza estaba influida por la de Descartes», estoy describiendo una relación causal entre dos hombres y constato algo sobre los procesos mentales de Spinoza.

Pero si digo: «Sin embargo, la doctrina de Spinoza está en contradicción con la de Descartes en algunos puntos importantes», entonces hablo sobre el contenido objetivo lógico de las dos doctrinas y no sobre procesos mentales. Me refiero sobre todo al contenido lógico de las proposiciones cuando subrayo el carácter de la objetividad del lenguaje humano. Y cuando antes decía que únicamente el pensamiento expresado puede convertirse en objeto de la crítica, lo que tenía en mente era, no que el proceso psicológico del pensamiento se puede discutir críticamente, sino sólo el contenido lógico de una proposición.

Quisiera recordarles ahora otra vez mi esquema en tres etapas:

- a) *problema*;
- b) *intentos de solución*;
- c) *eliminación*,

² N.T. Popper utiliza el término «Inhalt» para designar el contenido objetivo y «Gehalt» para designar el contenido lógico, en sentido fregeano.

asimismo, quiero que recuerden mi observación de que este esquema de la adquisición de nuevo conocimiento es aplicable desde la ameba hasta Einstein.

¿Dónde reside la diferencia? Esta cuestión es decisiva para la teoría de la ciencia.

La diferencia decisiva se encuentra en la etapa c), en la *eliminación* de los intentos de solución.

En el desarrollo precientífico del saber, la *eliminación* es algo que nos sobreviene: es el medio ambiente quien elimina nuestros intentos de solución; nosotros no tomamos parte activa de la eliminación, sino sólo pasiva; nosotros *sufrimos* la eliminación, y si destruye demasiado a menudo nuestros intentos de solución o si destruye un intento de solución que antes había tenido éxito, destruye con ello no sólo el intento de solución, sino a nosotros mismos, es decir, al soporte de los intentos de solución. Esto aparece claro en el caso de la selección darwiniana.

Lo decisivamente nuevo del método científico y de la actitud científica se encuentra, pues, en que nosotros participamos activamente en la ciencia y estamos interesados en la eliminación. Los intentos de solución están objetivados; ya no estamos identificados con nuestros intentos de solución. Independientemente de que seamos más o menos conscientes del esquema en tres etapas, o de que no lo seamos en absoluto, lo nuevo en la actitud científica consiste en que intentamos eliminar activamente nuestros intentos de solución. Sometemos nuestros intentos de solución a la crítica, y dicha crítica trabaja con todos los medios de que disponemos y que podemos elaborar. Por ejemplo: en lugar de esperar hasta que nuestro medio ambiente refute un intento de solución, una teoría, intentamos modificar el medio ambiente de manera que resulte *lo más desfavorable posible* para nuestro intento de solución. De esta forma, sometemos nuestras teorías a prueba; y, en verdad, intentamos someterlas a la prueba más difícil. Hacemos todo lo posible para eliminar nuestras teorías, pues nosotros mismos queremos encontrar las teorías que son falsas.

La pregunta de dónde reside la diferencia decisiva entre la ameba y Einstein, puede responderse de la siguiente manera: la ameba huye de la falsación: su expectativa forma parte de ella, y los portadores precientíficos de expectativas e hipótesis son a menudo aniquilados por la refutación de la hipótesis. Einstein, por el contrario, ha objetivado su hipótesis; la hipótesis es algo fuera de él; y el científico puede destruir su hipótesis por medio de su crítica, sin tener que perecer él mismo por ello. En la ciencia dejamos que nuestras hipótesis mueran por nosotros.

Con ello he llegado a mi hipótesis, a aquella teoría que fue desacreditada como paradójica por tantos adeptos de la teoría científica. Mi tesis fundamental es: lo que diferencia a la actitud científica y al método científico de una actitud precientífica, es el método de *los intentos de falsación*. Todo intento de solución, toda teoría, es puesta a prueba tan rigurosamente como nos es posible. Ahora bien, un examen riguroso es siempre un intento de descubrir las *debilidades* de lo examinado. Así, nuestra contrastación de la teoría es también un intento de descubrir sus debilidades. La contrastación de una teoría es, por consiguiente, un intento de refutar o falsar la teoría.

Esto no significa, naturalmente, que un investigador que consigue falsar su propia teoría tenga que alegrarse siempre de ello. La teoría fue formulada por él como un *intento de solución*, y esto significa que la teoría tendrá que resistir también duros exámenes. Muchos científicos que falsan un intento de solución que estaba lleno de esperanzas, se desilusionarán mucho personalmente.

El fin de falsar la teoría no es a menudo ningún fin personal del científico, y también bastante a menudo un verdadero científico intentará defender una teoría, en la que él pone grandes esperanzas, frente a un intento de falsación.

Esto es absolutamente laudable desde el punto de vista de la teoría científica; pues, ¿cómo podríamos sino distinguir una *verdadera* falsación de una *aparente*? Necesitamos en la ciencia una especie de formación de partidos en pro y en contra de cada teoría que se someta a un examen serio; pues necesitamos una *discusión* científica racional. Y la discusión no siempre conducirá a una clara decisión.

La importante y nueva actitud que hace de la ciencia lo que es, es en todo caso la actitud *crítica*, y ésta se logra, sobre todo, por medio de la formulación lingüística objetiva, pública, de sus teorías. Esto conduce después usualmente a la toma de partido y, con ello, a la discusión crítica. A menudo permanece la discusión en empate a lo largo de muchos años, como la famosa discusión entre Albert Einstein y Niels Bohr. En cualquier caso, no tenemos ninguna garantía de que toda discusión científica pueda decidirse. No hay ninguna garantía para el progreso científico.

Por consiguiente, mi tesis fundamental es que lo nuevo que diferencia la ciencia y el método científico de la preciencia y de la actitud precientífica, es la actitud conscientemente crítica ante los intentos de solución; por lo tanto, la participación activa en la eliminación, los intentos activos de eliminación, los intentos de criticar, es decir, de falsar.

Los intentos contrarios de salvar una teoría de la falsación, tienen también su función metodológica, como hemos visto. Pero mi tesis es que una actitud dogmática semejante es, en lo esencial, característica para el pensamiento *precientífico*, mientras que la actitud crítica, el intento consciente de falsación, conduce a la *ciencia* y domina el *método científico*.

A pesar de la función metodológica que tiene, sin duda, la formación de partidos científicos, es importante, en mi opinión, que cada investigador individual se aclare sobre el significado fundamental de los intentos de falsación y también de la falsación que a veces da buen resultado. Pues el método científico no es *acumulativo*, como enseñaba, por ejemplo, Bacon de Verulamio o sir James Jeans, sino esencialmente *revolucionario*. El progreso científico consiste esencialmente en que algunas teorías son superadas y sustituidas por otras teorías. Esas nuevas teorías tienen que ser capaces de resolver todos aquellos problemas que habían resuelto las antiguas teorías, al menos tan bien como ellas. Así resuelve la teoría de Einstein el problema del movimiento de los planetas y, en general, la macromecánica tan bien y *quizá incluso mejor* que la teoría newtoniana. Pero la teoría revolucionaria parte de nuevos supuestos y va en sus consecuencias bastante más allá que la antigua teoría, con la que se encuentra también en abierta contradicción. Esta contradicción permite idear experimentos que puedan decidir entre la antigua y la nueva teoría; pero sólo en el sentido de que al menos una de las dos teorías pueda falsarse: los experimentos pueden demostrar la superioridad de la teoría que los sobreviva, pero no su verdad; y la teoría que ha sobrevivido se puede superar de nuevo, por su parte, muy pronto.

Si un investigador ha comprendido esta situación, también se enfrentará críticamente él mismo a su teoría predilecta, por él creada. Preferirá contrastarla él mismo y, eventualmente, falsarla, que dejar esa tarea a sus críticos.

Un ejemplo, del que estoy orgulloso, es mi viejo amigo, el fisiólogo cerebral y Premio Nobel sir John Eccles. Coincidió por primera vez con John Eccles en la Universidad de Otago (Dunedin, Nueva Zelanda), donde pronuncié una serie de conferencias. Él llevaba años ocupándose experimentalmente con el problema de cómo el estímulo nervioso es trasladado de una célula nerviosa a otra por medio de la sinapsis, esto es, con la cuestión de la «transmisión sináptica». Una escuela que actuaba sobre todo en Cambridge en torno a sir Henry Dale, suponía que unas moléculas de una «sustancia transmisora» química superaban la sinapsis (que separa a las células nerviosas), trasladando así el estímulo de una célula a las otras. Sin embargo, los experimentos de Eccles habían mostrado que la duración temporal de la transmisión" era extraordinariamente corta -en su opinión, demasiado corta para la sustancia

transmisora-, y por esta razón desarrolló en todos sus detalles la teoría de una transmisión puramente eléctrica, tanto para la transmisión de la excitación nerviosa como para la transmisión de las inhibiciones. Pero voy a dejar hablar a Eccles mismo:³ «Hasta 1945 tuve las siguientes ideas convencionales sobre la investigación científica: primero, que las hipótesis resultan de la colección cuidadosa y metódica de datos experimentales. Ésta es la idea inductiva sobre la ciencia, que se remonta a Bacon y Mili. La mayoría de los científicos y filósofos siguen creyendo todavía que en eso consiste el método científico. Segundo, que la bondad de un científico se juzgará a partir de la veracidad de las hipótesis desarrolladas por él, las cuales deberían ampliarse, sin duda, con la acumulación de nuevos datos, pero que deberían servir -así se esperaba- como fundamentos firmes y seguros de sucesivos desarrollos teóricos. Un científico prefiere hablar sobre sus datos experimentales y considerar sus hipótesis sólo como herramientas de trabajo. Por último -y éste es el punto más importante-, es sumamente lamentable y un signo de fracaso, si un científico opta por una hipótesis que será refutada por nuevos datos, de forma que, finalmente se deba abandonar por completo.

[...] desechar, y esto me deprimió extraordinariamente. Había estado embarcado en una controversia sobre sinapsis y creía entonces que la transmisión sináptica entre las células nerviosas era en su mayor parte de naturaleza eléctrica. Admitía la existencia de un componente químico tardío más lento, pero creía que la rápida transmisión por medio de la sinapsis transcurría por un camino eléctrico. En este momento aprendí de Popper que no era nada injurioso científicamente reconocer como falsas las propias hipótesis. Ésta fue la novedad más hermosa que experimentaba en mucho tiempo. Popper mismo me convenció incluso para que formulara mis hipótesis sobre la transmisión sináptica excitatoria e inhibitoria producida eléctricamente, tan precisa y rigurosamente que retaran a la refutación -y ésta aconteció un par de años más tarde, en su mayor parte gracias a mis colegas y a mí mismo, cuando comenzamos en 1951 a hacer derivaciones intracelulares de motoneuronas. Gracias a la teoría popperiana pude aceptar gozoso la muerte de mi idea favorita, que había conservado durante casi 20 años, y estaba al mismo tiempo en situación de cooperar tanto como fuera posible a la «historia de la transmisión química» que por su parte, era la idea favorita de Dale y Loewi. Por fin había experimentado el gran y liberador poder de la teoría de Popper sobre los métodos científicos...

«Aquí se muestra una sucesión singular. Se demuestra que yo había estado dispuesto demasiado deprisa a desechar la hipótesis eléctrica de la transmisión sináptica. Los muchos tipos de sinapsis que habían sido objeto de mis trabajos, son seguramente de índole química, pero hoy se conocen muchas sinapsis eléctricas, y mi libro sobre la sinapsis (1964) incluye dos capítulos sobre transmisión eléctrica, ¡tanto inhibitoria como excitatoria!».

Es digno de mención que ni Eccles ni Dale tenían razón con sus rompedoras teorías para la investigación cerebral; pues ambos creían que sus teorías eran válidas para todas las sinapsis. La teoría de Dale era válida para las sinapsis de las que ellos se ocupaban entonces; pero era en general tan poco válida como la de Eccles. Sin embargo, los adeptos de Dale [... estaban demasiado seguros de su...]

En otro lugar, en su discurso de recepción del Premio Nobel, escribe Eccles:

«Ahora puedo incluso alegrarme de la falsación de una teoría favorita, pues semejante falsación es un éxito científico».

Este último punto es importante sobremanera: aprendemos siempre muchísimo por medio de una falsación. No sólo aprendemos *que* una teoría es falsa, sino que también aprendemos *por qué* es falsa. Y, sobre todo, obtenemos un *nuevo problema, más rigurosamente formulado*; y un nuevo problema es, como ya sabemos, el verdadero punto de partida de un nuevo desarrollo científico.

³ [...]

Quizá se hayan asombrado ustedes porque he mencionado tan a menudo mi esquema en tres etapas. En parte lo he hecho porque quería prepararles para otro esquema semejante, pero en cuatro fases; un esquema que es característico para la ciencia y el desarrollo del desarrollo científico. El esquema en cuatro fases se puede obtener a partir de nuestro esquema en tres etapas -esto es, problema, intentos de solución, eliminación-, si denominamos a la primera etapa como «el antiguo problema» y añadimos una cuarta como «los nuevos problemas». Si, a continuación, sustituimos los «intentos de solución» por «teorías a modo de prueba» y la «eliminación» por «intentos de eliminación por la discusión crítica», llegamos a ese esquema en cuatro etapas que es característico para la teoría científica.

Tiene el siguiente aspecto:

- a) el antiguo problema;
- b) formaciones de teorías a modo de prueba;
- c) intentos de eliminación por medio de discusión crítica, incluido examen experimental;
- d) los nuevos problemas, que surgen de la discusión crítica de nuestras teorías.

Mi esquema en cuatro etapas permite hacer una serie completa de observaciones teórico-científicas.

Sobre el problema. Los problemas precientíficos y los problemas científicos primarios son de naturaleza práctica, pero pronto son sustituidos, en virtud del ciclo de cuatro etapas, por problemas teóricos, al menos en parte. Es decir, la mayoría de los nuevos problemas surgen de la *crítica de las teorías*: son intrateóricos. Esto ya es válido para los problemas en la cosmogonía de Hesíodo y aún más para los problemas de los filósofos griegos presocráticos; y es válido para el mayor número de problemas de las ciencias naturales modernas: los problemas mismos son productos de las teorías y de las dificultades, que se descubren por la discusión crítica en las teorías. Dichos problemas teóricos son, fundamentalmente, preguntas por las *explicaciones*, por las teorías explicatorias: las respuestas a modo de prueba, que suministran las teorías, son en efecto *intentos de explicación*.

A los problemas prácticos se pueden sumar también los problemas que predicen algo. Pero desde el punto de vista *intelectual* de la ciencia pura, las predicciones pertenecen a la etapa c), es decir, a la discusión *crítica*, al *examen*. Son intelectualmente interesantes porque nos permiten contrastar, en la realidad y en la praxis, la pretensión de verdad de nuestras teorías, que representan intentos de explicación.

Podemos seguir leyendo en nuestro esquema de cuatro etapas que partimos en la ciencia de un *ciclo* de antiguos problemas y que concluimos con nuevos problemas que, por su parte, actúan de punto de partida de un nuevo *ciclo*. A causa del carácter cíclico o periódico de nuestro esquema, podemos comenzar en *cada una* de sus cuatro etapas. Podemos comenzar por las *teorías*, esto es, en la etapa b) de nuestro esquema. Por tanto, podemos decir que el científico parte de una *teoría más antigua* y que, por medio de su discusión crítica y eliminación, llega a los problemas que intenta entonces resolver por *nuevas teorías*. Se trata, precisamente a causa del carácter cíclico, de una interpretación absolutamente sólida.

También habla a su favor el que podamos caracterizar la *confección de teorías satisfactorias* como la *finalidad* de la ciencia. Pero, por otra parte, la cuestión acerca de bajo qué circunstancias puede considerarse una teoría como *satisfactoria*, conduce de nuevo directamente al *problema como punto de partida*. Pues evidentemente es la primera exigencia que planteamos a la teoría, que *resuelva* los problemas que están necesitados de explicación, aclarando las dificultades de que consta el problema.

Por último podemos escoger también como nuestro punto de partida la eliminación o supresión

de las teorías existentes hasta el momento. Pues se puede decir que la ciencia toma siempre su punto de partida del fracaso de una teoría; ese fracaso, la eliminación, conduce entonces al problema de reemplazar la teoría eliminada por una teoría mejor.

Yo personalmente prefiero el *problema* como punto de partida, pero tengo perfectamente claro que el *carácter cíclico* del esquema hace posible considerar cada una de las etapas como punto de partida de un nuevo desarrollo.

Esencial para el nuevo esquema en cuatro etapas, es su carácter dinámico: cada una de las etapas contiene, por así decir, una motivación lógica, interna, de continuar a la siguiente etapa. La ciencia, tal y como aparece en esta lógica científica, es en lo esencial una apariencia concebida en *crecimiento*; es esencialmente *dinámica*, nunca es algo *acabado*: No existe ningún punto en el que encuentre su fin definitivamente.

Todavía existe otro motivo por el que prefiero el *problema* como punto de partida. La *distancia* entre un *antiguo problema* y sus derivados, los *nuevos problemas*, me parece que caracteriza el progreso científico mucho más espectacularmente que, por ejemplo la distancia entre las antiguas teorías y la siguiente generación de nuevas teorías, que reemplazan a las antiguas.

Tomemos como ejemplo las teorías gravitatorias newtoniana y einsteiniana. La distancia entre las dos teorías es grande; pero es posible traducir la teoría newtoniana al lenguaje einsteiniano, el formalismo del denominado cálculo tensorial; y si se hace esto, como, por ejemplo, lo ha hecho el profesor Peter Havas, encontramos que la diferencia entre las dos teorías consiste únicamente en la velocidad de propagación finita de la gravitación, esto es, en la velocidad finita de la luz c . Esto significa que Havas ha conseguido formular la teoría einsteiniana de modo que por la sustitución de la velocidad de propagación *finita* c por una velocidad *infinita*, la teoría einsteiniana se convierte en la newtoniana.

Pero sería un gran error concluir a partir de esta reflexión que el mayor progreso de la teoría consiste en la velocidad de propagación finita de la gravitación.

Sostengo que se ve mucho más claramente el progreso y el carácter dinámico del desarrollo, si se compara *los problemas* que han descubierto los críticos de la teoría newtoniana, por ejemplo, Ernst Mach, con *aquellos problemas* que fueron descubiertos por los críticos de la teoría einsteiniana, sobre todo por Einstein mismo.

Por consiguiente, si se comparan los antiguos y los nuevos problemas, se ve la gran distancia, el gran progreso. En lo esencial, sólo ha quedado pendiente uno de los antiguos problemas, el denominado principio de Mach. Esto es, la pretensión de que debamos concebir la *inercia* de las masas pesadas como un efecto de las lejanas masas del universo. Einstein estaba muy decepcionado de que su teoría no suministrase una solución satisfactoria a este problema. Es cierto que su teoría de la gravitación explicaba la inercia como resultado de la gravitación; pero si hacemos que desaparezcan las masas en la teoría gravitatoria einsteiniana, la teoría se convierte en la teoría especial de la relatividad, y la inercia sigue aún sin masa que la origine.

Einstein mismo vio en este punto una de las carencias más delicadas de su teoría; y el problema de incorporar el principio de Mach a la teoría de la gravitación, ha preocupado desde hace medio siglo a todos los investigadores de este campo.

Por tales motivos me parece, pues, mejor comenzar nuestro esquema en cuatro etapas con el *problema*. Ahora bien, en cualquier caso el esquema muestra en qué consiste la novedad del desarrollo dinámico de la ciencia frente al desarrollo *precientífico*: en nuestra participación activa en el proceso de

eliminación por medio de la invención del lenguaje, de la escritura y de la discusión crítica. Mi tesis fundamental es que la ciencia se originó por medio de la invención de la discusión crítica.

Una importante consecuencia de mi tesis fundamental se refiere a la cuestión: ¿Cómo se diferencian las teorías científico-empíricas de otras teorías? Este problema no es, por su parte, un problema científico-empírico, sino un problema teórico-científico; se trata de un problema que pertenece a la lógica científica o a la filosofía de la ciencia. La respuesta a este problema, que se puede deducir de mi tesis fundamental, es la siguiente:

Una teoría científico-empírica se diferencia de otras teorías porque puede fracasar en experiencias posibles; es decir, que son posibles y se pueden describir ciertas experiencias que falsarían la teoría, si hiciéramos realmente semejantes experiencias.

He denominado «problema de demarcación» al problema de delimitar las teorías científico-empíricas del resto de las teorías, y «criterio de demarcación» a la solución que propongo.

Mi propuesta para resolver el problema de la demarcación consiste en el siguiente criterio de demarcación: Una teoría pertenece a la ciencia empírica cuando, y sólo cuando, entra en contradicción con experiencias posibles, esto es, es falsable en principio por medio de la experiencia.

He denominado a este criterio «criterio de falsabilidad».

El criterio de falsabilidad puede ilustrarse con muchas teorías. Así, es falsable, por ejemplo, la teoría de que la vacunación protege contra las viruelas: si alguien que ha sido correctamente vacunado, contrae a pesar de ello las viruelas, se ha falsado la teoría.

El ejemplo se puede utilizar también para mostrar que el criterio de falsabilidad alberga sus propios problemas. Si entre millones de seres humanos vacunados, sólo *uno* contrajera las viruelas, apenas podríamos considerar nuestra teoría como falsada. Antes bien admitiremos que algo no estuvo en orden con la vacunación o con la vacuna. Y, en principio, una salida semejante siempre es posible: si hemos de confrontarnos con una falsación, siempre podemos excusarnos de alguna manera; podemos introducir una hipótesis auxiliar y rechazar la falsación. Podemos «*inmunizar*» nuestras teorías contra todas las falsaciones posibles (para utilizar una expresión del profesor Hans Albert).

Por lo tanto, la aplicación del criterio de falsabilidad no es siempre fácil. Sin embargo, el criterio de falsabilidad posee su valor. Es aplicable a la teoría de la vacunación contra las viruelas, aun cuando la aplicación quizá no sea siempre tan sencilla: si el porcentaje de todos los seres humanos que fueron vacunados y sin embargo contrajeron viruelas, es aproximadamente el mismo que el porcentaje de los no vacunados que contrajeron viruelas (o quizá incluso mayor), todos los científicos renunciarán a la teoría de la protección de la vacuna.

Comparemos ahora ese caso con el caso de una teoría que, en mi opinión, no es falsable; por ejemplo, con la teoría de Freud del psicoanálisis. Esa teoría sólo se podría falsar, en principio de forma evidente, si pudiéramos describir un comportamiento de seres humanos que entre en contradicción con la teoría. Tales teorías falsables del comportamiento, existen; por ejemplo, la teoría de que un hombre que siempre ha dado muestras de *honradez* durante su larga vida, no puede volverse de pronto un ladrón en su ancianidad, si se encuentra en una circunstancia de seguridad financiera.

Esa teoría es seguramente falsable, y supongo que aquí y allá se presentan casos que se han falsado de hecho, así que la teoría, en la formulación ofrecida, es sencillamente *falsa*.

Pero en oposición a esta teoría, no parece existir ningún comportamiento imaginable de seres humanos que pudiera refutar el psicoanálisis. Si un hombre salva la vida a otro, con riesgo de su propia vida, o si, por el contrario, amenaza la vida de un viejo amigo, aunque lo consideremos un comportamiento humano raro, no entrará en contradicción con el psicoanálisis. El psicoanálisis puede explicar, en principio, todo comportamiento humano, incluso uno más raro que los mencionados. Por lo tanto no es empíricamente falsable, no es contrastable.

Con ello no quiero decir que Freud no haya visto muchas cosas correctamente. Pero sostengo que su teoría no tiene ningún carácter empírico-científico: no es comprobable.

Por contraposición, existen teorías en las que la comprobación es posible, como muestra nuestro ejemplo de la vacunación, pero son, sobre todo, teorías físicas, químicas o biológicas.

Desde la teoría de la gravitación de Einstein tenemos motivos para suponer que la mecánica newtoniana es falsa, aunque representa una aproximación extraordinaria. Pero, en cualquier caso, tanto la teoría de Newton como la de Einstein son falsables, aunque siempre es posible, naturalmente, excusarse de realizar una falsación por medio de una estrategia inmunológica. Mientras que el psicoanálisis de Freud no entra en contradicción con ningún comportamiento humano imaginable, el comportamiento de una mesa que se bambolea y gira en una sesión de espiritismo entra en contradicción con la teoría newtoniana. Si la taza que está llena de té sobre una mesa empieza de pronto a bailar, y gira y da vueltas, esto sería una falsación de la teoría de Newton; en particular, si a pesar de todos los giros y vueltas el té no se derrama. Se puede decir que la mecánica está en contradicción con una infinidad de formas de comportamiento imaginables de cuerpos físicos -muy al contrario que el psicoanálisis, que no entra en contradicción con ningún comportamiento humano posible.

La teoría gravitatoria de Einstein se vería afectada por casi todas las infracciones imaginables de la mecánica newtoniana, precisamente porque la mecánica newtoniana supone una muy buena aproximación a la einsteiniana. Pero Einstein va más allá, buscando expresamente casos que, cuando se observan, pudieran refutar su teoría, pero no la de Newton.

Así escribió Einstein, por ejemplo, que renunciaría a su teoría, como refutada, si no se encontrara el desplazamiento del rojo que él había calculado en el espectro del satélite Sirius y de otras diminutas estrellas fijas blancas.

Por lo demás, es interesante que el mismo Einstein tuviera una actitud tan sumamente crítica respecto a su propia teoría de la gravitación. Aunque ningún experimento de verificación (todos fueron sugeridos por él mismo) resultó desfavorable para su teoría, no consideraba a la misma como completamente satisfactoria, por motivos teóricos. Estaba completamente convencido de que su teoría, como todas las teorías científico-naturales, tenía el carácter de un *intento de solución provisional*, es decir, un carácter *hipotético*. Pero lo que él sostenía entraba más en cuestiones de detalle. Aportaba *razones* de por qué había que considerar a su propia teoría como defectuosa y como insatisfactoria desde el punto de vista de su propio programa de investigación; y sugería una serie de exigencias que debía cumplir una teoría satisfactoria.

Ahora bien, lo que pretendía para su teoría originaria de la gravitación era que representara una *aproximación mejor* a la teoría buscada que la proporcionada por la teoría de la gravitación newtoniana y, por ello, también *una mejor aproximación a la verdad*.

La idea de la *aproximación a la verdad* es, en mi opinión, una de las ideas más importantes de la teoría científica. Tiene conexión con el hecho de que la discusión crítica entre teorías competidoras es, como hemos visto, muy importante para la teoría de la ciencia. Pero una discusión crítica está regulada por ciertos valores. Necesita un principio regulativo o, en terminología kantiana, una idea regulativa.

Entre las ideas regulativas que dominan la discusión crítica de teorías competidoras hay que destacar tres de suma relevancia: primero, la idea de la *verdad*; segundo, la idea del *contenido* lógico y empírico de una teoría; y tercero, la idea del *contenido de verdad* de una teoría y de la *aproximación a la verdad*.

Puede observarse que la idea de la *verdad* domina la discusión crítica en el hecho de que discutimos críticamente las teorías con la esperanza de eliminar las *falsas* teorías. Esto muestra que estamos guiados por la idea de la búsqueda de teorías *verdaderas*.

La segunda idea regulativa, la idea del *contenido* de una teoría, nos conduce a buscar teorías con gran contenido informativo. Las tautologías o proposiciones aritméticas triviales, como 12 por 12 es 144, están vacías de contenido: no resuelven ningún problema científico-empírico. Los problemas difíciles sólo se pueden resolver con gran contenido lógico y empírico.

La magnitud del contenido de una teoría es lo que puede denominarse la *audacia* de la misma: cuanto más sostengamos una teoría, tanto mayor es el riesgo de que la teoría sea *falsa*. Así pues, ciertamente buscamos la verdad, pero estamos interesados únicamente en verdades audaces, arriesgadas. Nuevamente la teoría de la gravitación newtoniana o einsteiniana son buenos ejemplos de teorías audaces, que tienen un gran contenido lógico, lo mismo que la teoría cuántica de los átomos, o la teoría de la información genética cifrada, que resuelve en parte el problema de la transmisión hereditaria.

Tales teorías audaces tienen, como he dicho, un gran contenido; y ciertamente, un gran contenido lógico y empírico.

Estos dos conceptos de contenido se pueden explicar de la siguiente manera: El contenido lógico de una teoría es su *masa deductiva*, es decir, la cantidad o clase de todas las proposiciones que se pueden *deducir lógicamente* a partir de la teoría en cuestión. Cuanto mayor sea el número de conclusiones extraídas, mucho mejor.

Quizá sea todavía más interesante la idea *del contenido empírico* de una teoría. Para comprender esta idea, *partimos de que una ley empírica* o una teoría *empírica prohíbe* ciertos procesos observables. (La teoría «todos los cuervos son negros» prohíbe la existencia de cuervos blancos; y la observación de un cuervo blanco refuta la teoría.) En este sentido, hemos visto que el psicoanálisis freudiano no prohíbe, precisamente, *ningún* procedimiento observable. Su contenido lógico es seguramente grande; pero su contenido empírico es nulo.

Por lo tanto, puede designarse el contenido empírico de una teoría como la cantidad o la clase de las proposiciones empíricas prohibidas por la teoría; es decir, como la cantidad o la clase de las proposiciones empíricas que entran en contradicción con la teoría.

Para emplear una ilustración muy simple: la teoría de que *no existe ningún cuervo blanco* entra en contradicción con la proposición «aquí hay un cuervo blanco». *Prohíbe*, por así decir, la existencia de cuervos blancos. Ahora bien, la teoría de que *todos los cuervos son negros* tiene un contenido empírico todavía mayor. No sólo prohíbe cuervos blancos, sino también cuervos azules, verdes y rojos: la clase de las proposiciones prohibidas es mucho mayor.

Una proposición empírica o una proposición procedente de la observación, que entra en contradicción con una teoría, puede caracterizarse como *posibilidad de falsación* de la teoría en cuestión, o como un *falsador potencial* de la teoría: si se observa una posibilidad de falsación real, la teoría está *empíricamente falsada*.

La proposición «aquí hay un cuervo blanco» es, pues, una posibilidad de falsación o un falsador potencial, tanto de la teoría pobre de contenido de que no existe ningún cuervo blanco, como de la teoría más rica de contenido de que todos los cuervos son negros.

La proposición: «El 10 de febrero de 1972 se entregó en el jardín zoológico de Hamburgo un cuervo verde» es una posibilidad de falsación o un falsador potencial de la teoría de que todos los cuervos son negros, pero también de la teoría de que todos los cuervos son rojos o azules. Si se acepta como verdadera una proposición semejante, un falsador potencial semejante, en virtud de observaciones habrá que considerar falsadas de hecho todas aquellas teorías a cuyos falsadores pertenezca dicha proposición. Lo interesante es que, cuanto más dice la teoría, mayor es la cantidad de sus falsadores potenciales. Dice más cosas y puede explicar más problemas: su *potencial explicativo* o su *fuerza potencial explicativa* es mayor.

Podemos comparar de nuevo, desde este punto de vista, la teoría de la gravitación de Newton con la de Einstein. Encontramos que el contenido empírico y la fuerza explicativa potencial de la teoría de Einstein son mucho mayores que los de Newton. Pues sostiene mucho más. No sólo describe todas las clases de movimiento que describe la teoría newtoniana, en particular las órbitas planetarias, sino que también describe la influencia de la gravitación sobre la luz -un círculo de problemas sobre los que Newton no tiene nada que decir, ni en su teoría de la gravitación ni en su óptica. La teoría de Einstein es, pues, más arriesgada: se puede falsar, en principio, por medio de observaciones que dejan intacta la teoría newtoniana. Por ello el contenido empírico, la cantidad de falsificadores potenciales de la teoría de Einstein, es considerablemente mayor que el contenido empírico de la teoría de Newton. Y finalmente, la fuerza explicativa potencial de la teoría de Einstein es mucho mayor: en el caso de que supongamos asegurados por las observaciones semejantes efectos ópticos como, por ejemplo, el desplazamiento del rojo en el espectro del satélite Sirius previsto por Einstein, esos efectos ópticos se *explicarán* también por medio de la teoría de Einstein.

Pero, aun cuando las observaciones relevantes todavía no se hubieran llevado a cabo, podemos decir que la teoría einsteiniana es superior *potencialmente* a la teoría de Newton: tiene el mayor contenido empírico y el mayor potencial explicativo. Esto significa que es teóricamente más interesante. Pero la teoría de Einstein está, al mismo tiempo, mucho *más amenazada* que la de Newton. Está mucho más expuesta a las falsaciones, justamente porque la cantidad de sus falsadores potenciales es mayor.

Por ello, también es mucho más rigurosamente verificable que la teoría de Newton, que es por su parte muy rigurosamente verificable. Si se mantiene en pie ante esas verificaciones, si sale airosa de esas contrastaciones, no podemos decir, no obstante, que es verdadera, pues se puede falsar en verificaciones posteriores; pero podemos decir que no sólo es mayor su contenido empírico, sino que también su contenido de verdad es mayor que el de la teoría de Newton: es decir, que la cantidad de las proposiciones *verdaderas* que son deducibles de ella es mayor que la que es deducible de la de Newton. Y podemos, pues, continuar diciendo que la teoría de Einstein parece, a la luz de la discusión crítica, que también se sirve por entero de los resultados de la verificación experimental, como una mejor *aproximación a la verdad*.

La idea de la aproximación a la verdad presupone -lo mismo que la idea de la verdad como principio regulativo- una *concepción realista del mundo*. No presupone que la realidad es tal y como la describen nuestras teorías científicas; pero presupone que existe una realidad y que podemos aproximarnos más y más a una descripción adecuada de la realidad con nuestras teorías, que son nuestras ideas creadas por nosotros mismos y, por eso, siempre idealizaciones, si aplicamos el método de ensayo y error en cuatro etapas. Pero no basta con el método. También tenemos que tener suerte; pues las condiciones que encontramos sobre nuestro planeta y que posibilitan la vida y el desarrollo del lenguaje humano, de la conciencia humana y de la ciencia humana, son sumamente raras en el cosmos,

si el cosmos se parece aunque sea remotamente al que describe la ciencia. Pues, según la ciencia, el mundo está casi completamente vacío de materia y lleno fundamentalmente sólo de radiación caótica; y en los pocos lugares, en los que no está vacío, está lleno de materia *caótica*, demasiado caliente, por lo general, para la formación de moléculas o demasiado frío para el desarrollo de seres vivos, tal y como nosotros los conocemos. La cuestión es si existe, aparte de nosotros, vida en el universo o no -en todo caso, la vida es, desde un punto de vista cosmológico, un fenómeno sumamente raro, totalmente extraordinario. Y en el desarrollo de la vida el desarrollo de un método crítico y científico es también algo sumamente raro y, por ello, en el sentido del cálculo de probabilidades, es algo casi infinitamente improbable. Es decir, que nos tocó el premio gordo cuando surgieron la vida y la ciencia.

La *concepción realista del mundo*, junto con la idea de la aproximación a la *verdad*, me parece que son indispensables para una comprensión de la siempre *idealizadora* ciencia. Además, la concepción realista del mundo me parece que es la única humana: sólo ella explica que existen otros seres humanos que viven, sufren y mueren como nosotros.

La ciencia es un sistema producto de ideas humanas: hasta cierto punto tiene razón el idealismo. Pero esas ideas pueden fracasar en la probabilidad. Por ello tiene razón el realismo en última instancia. Quizá se tenga por un momento la impresión que me he ido de mi tema con esas observaciones sobre el realismo y con esas alusiones a la polémica del realismo. Pero no es ése el caso. La polémica realista está de rabiosa actualidad, por el contrario, en la mecánica cuántica. Es uno de los problemas más modernos y más abiertos de la actual teoría de la ciencia.

No me enfrento a este problema, como ya ha quedado claro, como imparcial. Tomo partido, enteramente, *por el realismo*. Pero existe en la mecánica cuántica una escuela idealista muy influyente. De hecho, sólo existen todas las sombras idealistas imaginables, y un célebre mecánico cuántico ha extraído incluso consecuencias solipsistas de la mecánica cuántica; sostiene que esas consecuencias solipsistas se siguen con necesidad forzosa de la mecánica cuántica. A eso sólo puedo responder, si esto es así, debe haber algo falso en la mecánica cuántica, todo lo digno de admiración que sea y todo lo extraordinaria que resulte como aproximación a la verdad. La mecánica cuántica ha resistido los embates de verificaciones sumamente rigurosas. Pero sólo podemos concluir a partir de esto su proximidad a la verdad, si somos realistas.

La lucha por el realismo y el objetivismo en la teoría de la ciencia durará todavía mucho más tiempo. Se trata de un problema abierto y actual. Al mismo tiempo se trata de un problema, por la teoría de la ciencia, como se ha señalado, que en cierta medida se ha superado a sí mismo. Espero haber aclarado suficientemente cuál es mi postura respecto a este problema fundamental.

Capítulo 2

LA REDUCCIÓN CIENTÍFICA Y EL ESTADO INCOMPLETO ESENCIAL DE LA CIENCIA⁴

I

La tesis de la que parto es la siguiente: en conexión con la llamada «reducción», existen para los biólogos tres cuestiones fundamentales:

1. ¿Podemos reducir la biología a la física, o sea, a la física o la química? O ¿podemos esperar estar en situación algún día de reducirla por completo a la física (o física y química)?
2. ¿Podemos reducir (o esperar reducir) a la biología toda experiencia consciente, que a menudo atribuimos a los animales, y en caso de que la cuestión 1 se responda afirmativamente, reducirla luego a la física y a la química?
3. ¿Podemos reducir a la experiencia animal la autoconciencia y la creatividad de los espíritus humanos o esperar reducirla y, con ello, en caso de que las cuestiones 1 y 2 se respondan afirmativamente, a la física y a la química?

Está claro que la respuesta a estas cuestiones dependerá en parte del significado del término «reducir». Pero por razones que he expuesto ya en otro lugar,⁵ estoy en contra del método del análisis de significado y contra el intento de querer resolver problemas serios por medio de definiciones. Lo que yo propongo en lugar de eso es lo que sigue.

En primer lugar, discutiré algunos ejemplos de reducciones tanto con éxito como sin éxito, en particular la reducción de la química a la física; y abordaré la cuestión de qué sobra en esta reducción.

En el desarrollo de la discusión sostendré tres tesis: *primera*, que los científicos deben ser reduccionistas en el sentido de que nada en la ciencia suponga un éxito mayor semejante que una reducción llevada a cabo con éxito (como por ejemplo la reducción de Newton -o, mucho mejor dicho, la explicación-⁶ de las leyes de Kepler y Galileo por medio de su teoría gravitatoria y su corrección de esas leyes (véanse mis explicaciones detalladas).⁷ Una reducción con éxito es quizá la forma más afortunada de todas las explicaciones científicas que uno se puede imaginar, puesto que efectúa lo que Meyerson⁸

⁴ Publicado por primera vez en inglés en *Studies in the Philosophy of Biology* (comps. F.J. Ayala y T. Dobzhansky), Londres, Macmillan, 1974, págs. 259-283. Debo agradecer enormemente a David Miller y Jeremy Shearmur sus observaciones críticas a una primera versión de esta conferencia, que tuvo lugar en 1972, en un Congreso de biólogos y filósofos, organizado por los profesores Francisco Ayala y Teodosio Dobzhansky.

⁵ Véase mi libro *Die offene Gesellschaft und ihre Feinde* [La sociedad abierta y sus enemigos], vol. II, 1959 (7ª ed., 1992), cap. II, apartado II, págs. 15-29, J.C.B. Mohr (Paul Siebeck), Tübinga, N.T.: existe traducción castellana, a cargo de Eduardo Loedel, en Paidós Studio/Básica; la referencia puede encontrarse aquí, en las págs. 205-215.

⁶ (Añadido en la corrección de pruebas): En el texto de la conferencia no he prestado atención -quizá por negligencia, quizá también porque albergo una cierta aversión contra las sutilezas terminológicas-, a la distinción posible, sin más, entre una explicación en general y una reducción en el sentido de una explicación por medio de una teoría establecida o «fundamental». Probablemente sería de gran interés hacer la distinción entre una explicación de algo conocido por medio de una teoría nueva (desconocida), por una parte, y una reducción a una teoría antigua (conocida), por la otra parte. En lugar de una referencia pormenorizada a esta distinción, he añadido las notas a pie de página y el epílogo al final del texto, con la esperanza de poder paliar con ello los posibles malentendidos.

⁷ Véase mi libro *Objektive Erkenntnis* [Conocimiento objetivo] (1973), Hoffmann und Campe, Hamburgo, (8ª ed., 1992), cap. 5. Existe trad. cast. de Carlos Solís Santos, en Tecnos, Madrid, 1974.

⁸ Meyerson, É. (1908), *Identité et Réalité*, F. Alean, París. Trad. inglesa: Meyerson, É. (1930), *Identity and Reality*, Alien and Unwin, Londres.

puso especialmente de relieve: la identificación de lo desconocido con lo conocido. Por contraposición a una reducción, sin embargo, una explicación aclara, con ayuda de una nueva teoría, lo conocido -el problema conocido-, por medio de algo desconocido: una nueva conjetura.⁹

Segunda, defenderé la tesis de que los científicos, como siempre, gustan de orientarse filosóficamente al holismo, acogen con satisfacción el reduccionismo como *método*: tienen que ser o bien reduccionistas ingenuos o bien reduccionistas más o menos críticos, aunque sean también más o menos reduccionistas pesimistas críticos, como argumentaré, porque apenas ha existido nunca en la ciencia una reducción significativa que tuviera éxito. Incluso los intentos de reducción más afortunados dejan un resto irresoluble sobrante.

Tercera, defenderé la tesis de que no parece existir ningún buen argumento a favor de un *reduccionismo* filosófico mientras que, al contrario, existen buenos argumentos contra el esencialismo, con el que parece estar estrechamente emparentado el reduccionismo filosófico. Desde luego, voy a proponer que, no obstante, no deberíamos desistir de los intentos de reducción por razones metodológicas. La razón de esto es que nosotros mismos podemos aprender enormemente a partir de intentos de reducción sin éxito e incompletos, y que los problemas que quedan abiertos por ese procedimiento pertenecen a las más valiosas posesiones espirituales de la ciencia: yo pienso que una insistencia mayor en aquello que a menudo se considera como fracaso científico (en otras palabras: los grandes problemas abiertos de la ciencia), puede ser muy provechosa.

II

Aparte de la reducción de Newton, la reducción de las fracciones racionales a pares ordenados de números naturales (es decir, a relaciones o proporciones entre ellas) es una de las pocas reducciones conocidas, al menos para mí, que obtuvieron un éxito casi completo. Fue una conquista de los griegos, aun cuando se debería decir que incluso esa reducción deja sobrante un *resto* del que sólo en el siglo XX alguien se volvió a ocupar (con la reducción exitosa del par ordenado a un par desordenado de pares desordenados, realizada por Wiener y Kuratovski;¹⁰ además, no se debería pasar por alto que se trata de una reducción a *cantidades* de pares equivalentes, en lugar de tan sólo a los pares mismos). Secundaba el programa pitagórico cosmológico de investigaciones de la aritmetización, que se vino abajo con la demostración de la existencia de números irracionales, por ejemplo de la raíz cuadrada de 2, 3, o 5.¹¹ Platón reemplazó después el programa cosmológico de la aritmetización por uno de geometrización, y este programa lo siguieron con éxito desde Euclides hasta Einstein. Sin embargo, se hizo necesaria una nueva aritmetización con la invención del cálculo infinitesimal por Newton y Leibniz (y con el problema de excluir los resultados paradójicos, que no son capaces de dejar fuera sus propios métodos intuitivos) - una nueva reducción a los números naturales. Y a pesar de los éxitos sumamente espectaculares del siglo XIX y de comienzos del XX, podemos decir ahora, como yo creo, que esa reducción no tuvo un éxito completo.

Para mencionar únicamente un resto irresuelto: una reducción a una serie de números naturales o a una cantidad, en el sentido de las modernas teorías cuantitativas, no es lo mismo (y ni siquiera

⁹ Véase mi libro *Vermutungen und Wiederlegungen [Conjeturas y refutaciones]*, 1994, J.C.B. Mohr (Paul Siebeck), Tübinga. (Los 10 primeros capítulos de este libro se publicarán en este año como parte I; la parte 2 aparecerá en breve). N.T.: El título completo de esta obra en su traducción castellana, a cargo de Néstor Míguez (Paidós, 1979) es: *El desarrollo del conocimiento científico. Conjeturas y refutaciones*.

¹⁰ Wiener, N. (1914), «A simplification of the logic of relations». *Proceedings of the Cambridge Philosophical Society*, 17, págs. 387-390; y Kuratovski, C. (1920), «Sur la notion de l'ordre dans la théorie des ensembles». *Fundamenta mathematica*, 2, págs. 154-166.

¹¹ Véanse mis libros *La sociedad abierta y sus enemigos*, vol. I, cap. 6, nota 9 [N.T.: en trad. cast. citada, págs. 493-499]; y *Conjeturas y refutaciones*, cap. 2 [N.T.: págs. 80-115 de la trad. cast. citada].

semejante) que una reducción a una cantidad de pares ordenados equivalentes de números naturales. Mientras se utilizó la idea de una cantidad ingenuamente y de manera puramente intuitiva (como la de Cantor), esto no fue quizá evidente. Pero las paradojas de cantidades infinitas (que discutieron Bolzano, Cantor y Russell) y la necesidad de axiomatizar las teorías cuantitativas, han mostrado, dicho con cautela, que la reducción proporcionada no era una aritmetización simple -una reducción a los números naturales-, sino una reducción a la teoría axiomática de la cantidad; y ésta se ha mostrado como una empresa sumamente complicada y, en alguna medida, arriesgada.

Para resumir este ejemplo: el programa de la aritmetización, es decir, de la reducción de la geometría y de los números irracionales a los números naturales, ha fracasado en parte. Pero el número de los problemas y la cantidad de saber inesperados que se han generado por este fracaso, es avasallador. Esto, así lo sostendré, puede generalizarse: en el mismo aspecto donde no hemos tenido éxito como reduccionistas, puede volverse sumamente valioso el número de los resultados interesantes e inesperados que hemos obtenido a lo largo del camino hacia el fracaso.

III

He aludido brevemente al fracaso de la reducción intentada de los números irracionales a los números y, al mismo tiempo, he indicado que los intentos de reducción forman parte de las actividades científicas, así como de las explicaciones, simplificaciones y comprensiones matemáticas.

En adelante quisiera discutir más pormenorizadamente los éxitos y fracasos de los intentos de reducción en la física, en particular los éxitos parciales de la reducción de la macro-física a la microfísica y de la química, tanto a la microfísica como a la macrofísica.

IV

En otro lugar,¹² he calificado como «explicación última» al intento de explicar o reducir las cosas apelando a algo que necesita o es susceptible de una explicación posterior, en particular una «esencia» o una «sustancia» (*ousía*).

Un ejemplo demostrativo de esto es la reducción cartesiana de toda la física de cuerpos inanimados a la idea de una *sustancia extensa*; una sustancia (materia) con una única cualidad esencial: la extensión espacial.

Este intento de reducir toda la física a una única cualidad aparentemente esencial de la materia, tuvo un éxito tan extraordinario que condujo a una imagen comprensible del universo físico. El universo físico cartesiano era un mecanismo móvil de relojería, de «torbellinos» (como los denominaba Descartes), conectados unos con otros como las ruedas en un mecanismo. En esa obra de relojería, cada «cuerpo», o sea, cada «parte de la materia», choca con la parte colindante y su vecino, a su vez lo empuja hacia la dirección contraria. Sólo podía encontrarse materia en el mundo físico, y todo el espacio estaba lleno de ella. Sí, incluso el espacio estaba todavía reducido a la materia, puesto que no existía ningún espacio vacío, sino sólo la extensión esencialmente espacial de la materia. Y sólo había una única clase de causación puramente física: *Toda causación era choque* o acción inmediata.

Esta forma de ver el mundo la percibió como satisfactoria incluso Newton, aunque se sintió obligado por su teoría de la gravitación a introducir una nueva clase de causación: la *fuerza de atracción* o acción remota.

¹² Véase *Conjeturas y refutaciones*, cap. 3 [N.T.: págs. 116-141 de la trad. cast. citada].

Fue el casi increíble éxito de explicación y de predicción de la teoría de Newton, lo que hizo fracasar el programa de reducción cartesiano. Como he aventurado en otro sitio,¹³ Newton mismo intentaba poner en práctica el programa de reducción cartesiano, explicando la atracción de la gravitación por medio de un «impulso» (presión de radiaciones combinada con un efecto de protección) procedente de un bombardeo de partículas cósmicas (el intento se trae a colación vinculado normalmente con el nombre de Le Sage). Pero yo suponía también que Newton era consciente de la objeción aniquiladora contra esta teoría. Por encima de todo, esa teoría pretendía reducir la fuerza de atracción y la acción remota a choque y efecto inmediato; pero también significaba que todos los cuerpos móviles se movían en un medio resistente, que actuaba como un freno para sus movimientos (piénsese en el violento choque de la lluvia contra el parabrisas de un coche, en comparación con el que se produce en el cristal de la ventana posterior del mismo), y la aplicación de la ley de inercia de Newton la invalidaría.

El intento de una reducción última de la fuerza de atracción al choque fracasa, por consiguiente, a pesar de su atractivo intuitivo y aunque Newton mismo rechazaba la concepción, en su opinión absurda de que la fuerza de atracción, como acción remota, pudiera ser una cualidad esencial de la materia.

V

Hasta aquí nuestro primer (y muy simple) ejemplo para mostrar una reducción científica muy prometedora y su fracaso, también para hacer ver cuánto puede aprenderse a partir del intento de una reducción y del descubrimiento de su fracaso.

(Supongo que ese fracaso fue la razón inmediata de que Newton describiera el espacio como *sensorium* de Dios. El espacio era, por así decirlo «consciente» de la distribución de todos los cuerpos en él; en cierto sentido era omnisciente. También era omnipresente, pues transmitía ese saber con velocidad infinita a todos los lugares afectados y en cualquier momento que se deseara. Hasta aquí era el espacio, que poseía al menos dos atributos característicos de la esencia divina, por sí solo parte de la esencia divina. Esto era, tal y como yo creo, otro intento de Newton de proporcionar una explicación última esencialista.)

Se puede considerar la reducción cartesiana como ilustración de mi advertencia de que debemos ensayar reducciones por motivos metodológicos. Pero quizá indica también la razón de por qué afirmo que no podemos ser de ninguna manera, como reduccionistas, excesivamente confiados, sino que debemos ser, por lo que respecta al éxito completo de nuestro intento de reducción, convenientemente pesimistas.

VI

Está claro, creo yo, que el intento cartesiano de reducirlo todo en el mundo físico a la extensión y al choque (lo que, si tengo razón con mi suposición histórica, también era el intento de Newton), se tornó fracaso en el momento en que se confrontó con el éxito de la teoría de la gravitación de Newton. Dicho éxito fue tan potente que los newtonianos, empezando por Roger Cotes, comenzaron a considerar la teoría newtoniana misma como una explicación última y, con ello, la *atracción gravitatoria* como una propiedad esencial de la materia, a pesar de las propias opiniones contrarias de Newton. Newton no había visto, con todo, ninguna razón de por qué la *extensión* (sus átomos) y la *inercia* no podían ser cualidades esenciales de la masa (véanse mis explicaciones).¹⁴ De acuerdo con esto, podemos decir que

¹³ *Conjeturas y refutaciones*, cap. 3, nota 21 [N.T.: pág. 127 de trad. cast.].

¹⁴ *Conjeturas y refutaciones*, cap. 3 [N.T.: págs. de trad. cast. citadas en nota 13].

Newton era claramente consciente de la distinción entre masa inerte y pesada, que más tarde fue puesta de manifiesto por Einstein, y del problema que se planteaba por su proporcionalidad (o identidad) -de un problema que, debido al oscurantismo del vestigio esencialista, se perdió casi completamente de vista en la época entre Newton y Eötvös o hasta Einstein.

La teoría de la relatividad especial de Einstein destruía la identidad esencial de materia inerte y pesada, y éste es el motivo por el que intentaba explicarla más o menos *ad hoc* por medio de su principio de equivalencia. Pero cuando se descubrió (primero por Cornelio Lanczos) que las ecuaciones gravitatorias de Einstein conducían por sí mismas al principio -antes admitido por separado- de que todo cuerpo que gravita se mueve en una línea espaciotemporal geodésica, se redujo de hecho el principio de inercia a las ecuaciones gravitatorias y, con ello, la masa inerte a la masa pesada. (Creo que Einstein, aunque estaba muy impresionado por la importancia estos resultados, no aceptaba por completo que esto resolvía satisfactoriamente el problema central de Mach -la explicación de la inercia-, conocido como el famoso, pero de ninguna manera claro, «principio de Mach»: el principio de que la inercia de todo cuerpo individual proviene del efecto combinado de todos los demás cuerpos del universo. Para decepción de Einstein, ese principio -al menos en algunas de sus interpretaciones- era incompatible con la teoría de la relatividad general, que se convierte, para un espacio sin todos los cuerpos, en teoría especial de la relatividad, en la que sigue siendo válida la ley de inercia frente a las ideas de Mach.)

En mi opinión, nos las tenemos que ver aquí con un ejemplo sumamente satisfactorio de una reducción exitosa: la reducción de un principio de inercia generalizado a un principio de gravitación igualmente generalizado. Pero sólo en contadas ocasiones ha sido considerada bajo este aspecto; ni siquiera una sola vez por Einstein, aunque percibiera en gran medida el significado de un resultado que podía servir, desde un punto de vista puramente matemático, como elegante, pero no especialmente importante. Pues la dependencia o independencia de un axioma dentro de un sistema de axiomas reviste, en general, sólo un interés formal. Por esta razón, ¿por qué debía ser de importancia si la ley del movimiento en una línea geodésica, se tenía que admitir como un axioma separado o se podía deducir de la teoría restante de la gravitación? La respuesta a esta cuestión es que la identidad de masa inerte y pesada se *explica* por medio de esa deducción, y la primera es reducida a la última.

En este sentido podría decirse que el gran problema de Newton de la acción remota (expresado en el lenguaje propio del esencialismo) no se resolvió tanto por la velocidad finita de la acción recíproca gravitatoria de Einstein, sino más bien por medio de la reducción de la materia inerte a la materia pesada.

VII

Newton y los newtonianos sabían naturalmente de la existencia de fuerzas magnéticas y eléctricas; y al menos hasta comienzos del siglo XX, se emprendieron intentos para reducir la teoría electro magnética a la mecánica newtoniana o a una forma modificada de ella.

El problema esencial en este desarrollo era la reducción de las fuerzas *prima facie* no centrales (fuerzas de Oersted) a las fuerzas centrales, las únicas que parecían armonizar en una teoría newtoniana modificada. Los nombres sobresalientes en ese desarrollo fueron Ampère y Weber.

También Maxwell comenzó primeramente con el intento de reducir el campo de fuerzas (o bien, líneas de fuerza) electromagnéticas de Faraday a un mecanismo newtoniano, o sea, a un modelo del éter luminoso. Pero abandonó el intento (aunque no el éter luminoso como soporte del campo electromagnético). Asimismo, Helmholtz se sintió atraído por un programa de reducción newtoniano y, en parte, cartesiano y cuando propuso a su discípulo Heinrich Hertz trabajar en este problema, le pareció haberlo

hecho en la esperanza de salvar el programa de investigaciones de la mecánica. Sin embargo, aceptó la confirmación de Hertz de las ecuaciones de Maxwell como contraprueba. Tras Hertz y JJ. Thomsom el programa de investigación diametralmente opuesto ganó en atractivo: era el programa de una reducción de la mecánica a la teoría electromagnética.

VIII

La teoría electromagnética de la materia, es decir, la reducción tanto de la mecánica como también de la química a una teoría electromagnética del atomismo, tuvo éxito de una manera espectacular desde, lo más tarde, 1912, el año del modelo planetario del átomo de Rutherford, hasta digamos 1932.

De hecho, la mecánica cuántica (o bien la «nueva teoría cuántica», como se la denominaba antes) fue, al menos hasta 1935, simplemente otra denominación para aquello que entonces se tenía por la forma definitiva de la reducción de la mecánica a la nueva *teoría electromagnética de la materia*.

Para tener presente la importancia que le fue concedida a esta reducción por los físicos de primer orden, poco antes de la mecánica cuántica, quisiera citar a Einstein, quien escribió:¹⁵ «Puesto que según nuestras concepciones actuales... las partículas elementales... [es decir, electrones y protones] no son otra cosa que concentraciones del campo electromagnético, nuestro concepto actual del mundo conoce dos... realidades, a saber, el éter gravitatorio y el campo electromagnético o -como también podrían denominarse- espacio y materia».

Reparen en el «no son otra cosa» que he subrayado, porque es característico para una reducción en gran escala. De hecho, Einstein intentó hacia el final de su vida reunir los campos gravitatorios y los campos electromagnéticos en una teoría de campos unitaria, incluso después de que su concepción de 1920 fue superada -o, mucho mejor dicho, se vino abajo-(en particular a causa del descubrimiento de las fuerzas nucleares).

Una concepción en lo esencial igualmente reduccionista fue aceptada en aquella época (1932) por casi todos los físicos de primera categoría: por Eddington y Dirac en Inglaterra y, salvo Einstein, por Bohr, de Broglie, Schrödinger, Heisenberg, Born y Pauli en el contenido europeo. Robert A. Millikan¹⁶ nos ha ofrecido una exposición especialmente impresionante de esta concepción en aquellos tiempos, en el Instituto California de Tecnología:

«De hecho, nunca se ha llegado a una simplificación más hermosa en la historia de las ciencias naturales que en la serie de descubrimientos que alcanzó su apoteosis en 1914 y, finalmente, ayudó a la teoría a conseguir la suposición prácticamente universal de que el mundo material consta de únicamente dos entidades fundamentales, a saber, electrones positivos y negativos, con exactamente la misma carga, pero muy diferente cantidad de masa, con lo que el electrón positivo (hoy denominado usualmente "protón") es 185 veces más pesado que el negativo (hoy denominado usualmente "electrón")».

Ese paso reduccionista fue puesto por escrito precisamente en el momento oportuno: en el mismo año Chadwick¹⁷ publicó su descubrimiento del neutrón, y Anderson¹⁸ fue el primero en descubrir el positrón. Algunos de los mayores físicos, por ejemplo Eddington,¹⁹ eran, sin embargo, todavía de la opinión, según la tesis de Yukawa de la existencia de mesones (como más tarde se los denominó), de que con la aparición de la mecánica cuántica, la teoría electromagnética de la materia había entrado en su estadio final, y de que toda la materia está compuesta de electrones y protones.

¹⁵ *Äther und Relativitätstheorie [Éter y teoría de la relatividad]* (1920).

¹⁶ [...]

¹⁷ Chadwick, J. (1932), «Possible existence of neutron», *Nature* 129, pág. 132

¹⁸ [...]

¹⁹ [...]

IX

De hecho, la reducción de la mecánica y de la química a la teoría electromagnética parecía casi perfecta. Lo que había aparecido a Descartes y Newton como la esencia de la materia que llena el espacio y como el choque cartesiano, se había reducido (como Leibniz había reclamado ya hacía mucho tiempo) a *fuerzas de repulsión* -a fuerzas que son ejercidas por electrones negativos sobre electrones negativos. La neutralidad eléctrica de la materia quedó explicada por el mismo número de protones positivos y de electrones negativos; y la ionización de la materia por la pérdida de electrones (o excedente de electrones) de la envoltura electrónica exterior del átomo.

La química quedó reducida a la física, al menos aparentemente, por medio de la teoría cuántica de Bohr del sistema periódico de los elementos -una teoría que, mediante la aplicación del principio de exclusión de Pauli, se perfeccionó de ingeniosa manera-, y la teoría de la combinación química y de la naturaleza de los enlaces químicos covalentes, quedaron reducidas gracias a Heitler y London²⁰ a una teoría de la valencia (homeopolar), que hacía uso precisamente del principio de Pauli.

Aunque era patente que la materia es más bien una estructura compleja que una sustancia irreductible, nunca antes se ha dado una unidad semejante en el universo de la física o una dimensión semejante en reducción.

Tampoco se ha alcanzado nunca más desde entonces.

Naturalmente, seguimos creyendo en la reducción del choque cartesiano a fuerzas electromagnéticas; y también la teoría de Bohr del sistema periódico de los elementos sigue existiendo, a pesar de los considerables cambios a causa de la introducción del isótopo. Pero todo lo restante en esa hermosa reducción del universo a un universo electromagnético con dos partículas como piedras angulares estables, se ha disuelto mientras tanto. En el transcurso de ese proceso de disolución hemos aprendido una enorme cantidad de nuevos hechos: ésta es una de mis tesis fundamentales; pero la simplicidad de la reducción está allí.

Ese proceso, que comenzó con el descubrimiento de neutrones y positrones, se ha continuado desde entonces con el descubrimiento de nuevas partículas elementales. Pero la teoría de las partículas no es por una vez la dificultad principal: el derrumbamiento propiamente dicho sobrevino por el descubrimiento de nuevas clases de fuerzas, en particular de las fuerzas del núcleo con un radio de acción extremadamente corto, las cuales no son reducibles a fuerzas electromagnéticas y gravitatorias. Las fuerzas gravitatorias no intranquilizaron en exceso a los físicos en aquellos días, puesto que acababan de ser explicadas por la relatividad general, y existía la esperanza de superar el dualismo de las fuerzas gravitatorias y de las fuerzas electromagnéticas en una teoría de campo unitaria. Ahora bien, en la física tenemos que tratar con al menos cuatro tipos de fuerzas muy diferentes y todavía irreductibles: gravitación, interacción débil, fuerzas electromagnéticas y fuerzas nucleares.

X

Con esto, se había reducido con éxito al electromagnetismo la mecánica cartesiana, considerada por Descartes y Newton al unísono como la base a la que se debía reducir todo lo demás y dicha reducción al electromagnetismo se sigue practicando hoy todavía. Pero ¿qué sucede con la reducción notoriamente más impresionante de la química a la física cuántica?

²⁰ Heitler, W. y London, F. (1927), «Wechselwirkung neutraler Atome und homöopolare Bindung nach der Quantenmechanik» [Interacción de átomos neutrales y enlace homeopolar según la mecánica cuántica], *Zeitschrift für Physik* 44, págs. 455-472.

Admitamos por mor del argumento que disponemos de una reducción completamente satisfactoria de las combinaciones químicas a la teoría cuántica (tanto de las covalentes, o enlaces de pares de electrones, como de los enlaces no covalentes de iones), a pesar de la característica advertencia de Pauling,²¹ el autor del libro *The Nature of the Chemical Bond*, de que él no podía «definir» (o, aún más precisamente, describir) en qué consiste la naturaleza de las combinaciones químicas. Admitamos además, por mor del argumento, que disponemos de una teoría completamente satisfactoria de las fuerzas nucleares y del sistema periódico de los elementos y sus isótopos, en particular de la estabilidad e inestabilidad de los núcleos pesados. ¿Disponemos con ello de una reducción completamente satisfactoria de la química a la mecánica cuántica?

Yo creo que no. Tiene que añadirse una idea enteramente nueva, una idea que no tenga que ver directamente nada con la teoría física: la idea de la evolución, de la historia de nuestro universo, de la cosmogonía.

La explicación para esto es que la tabla periódica de los elementos y la teoría (no formulada) de Bohr del sistema periódico explican los núcleos pesados como combinaciones a partir de los más ligeros -en última instancia, a partir de núcleos de hidrógeno (protones) y neutrones (que, por su parte, se *podían* considerar de nuevo como un tipo de combinación a partir de protones y electrones). Y esa teoría admite que los elementos más pesados tienen una historia- que las cualidades de sus núcleos son el resultado de un extraño proceso, en el que varios núcleos de hidrógeno son llevados, bajo condiciones que sólo raras veces se dan en el cosmos, a fusionarse con núcleos más pesados.

Tenemos un montón de pruebas documentales de que esto sucedió de hecho y todavía sigue sucediendo; que los elementos más pesados tienen una historia evolutiva y que el proceso de fusión, por medio del cual el hidrógeno pesado se transforma en helio, es la fuente principal de la energía de nuestro sol, así como de la bomba de hidrógeno. El helio y todos los elementos más pesados son, por consiguiente, el resultado de la evolución cosmológica. Su historia, en particular la historia de los elementos más pesados, es, según las actuales concepciones cosmológicas, una historia muy especial. Actualmente se considera a los elementos más pesados como los productos de explosiones de la estrella Supernova. Puesto que el helio representa, según algunas apreciaciones recientes, el 25 % de toda la materia (calculada según la masa) y el hidrógeno dos tercios o tres cuartos de toda la materia (calculada según la masa), parece ser que todos los elementos más pesados aparecen muy raras veces (constituyen juntos quizá un uno o dos por ciento de la masa total). Según esto, la tierra y probablemente también los otros planetas de nuestro sistema solar, constan fundamentalmente de materiales muy raros (y como yo quisiera suponer, muy valiosos).

En la actualidad, la teoría más difundida del origen del universo²² -la teoría del Big Bang caliente-, supone que la mayor parte del helio sería el producto del Big Bang mismo: que fue producido dentro de los primeros minutos de la existencia del universo en expansión. El precario estatus científico de esta especulación (que originariamente se remonta a Gamov) no necesita ser subrayado expresamente. Y puesto que tenemos que remitirnos a teorías de este tipo en el intento por reducir la química a la mecánica cuántica, apenas puede suponerse que dicha reducción haya tenido enteramente éxito. En realidad, hemos reducido la química (al menos una parte) menos a una teoría física y mucho más a la cosmología: la moderna cosmología relativista clásica comenzaba como una teoría física aplicada; pero como Bondi ha puesto de manifiesto, ésos tiempos parecen haber terminado ya, y no debemos perder

²¹ Pauling, L. (1959) (contribución de debate) en *The Origin of Life on the Earth* (Proceedings of the First International Symposium on The Origin of Life on the Earth, Moscú, 19 a 24 de Agosto de 1957, comps. A.I. Oparin y otros), ed. F. Clark and R.L.M. Synge, Pergamon Press, Londres, pág. 119.

²² (Añadido después de la conferencia): Esta teoría podría verse amenazada ahora por la nueva teoría del desplazamiento del rojo, que ha sido propuesta por J.C. Pecker, A.P. Roberts y J.P. Vigiier: «Non-velocity red-shifts and photon-photon interactions», *Nature* 237 (1972), págs. 227-229.

de vista el hecho de que algunas de nuestras ideas (por ejemplo, las ideas que llegaron con Dirac y con Jordán) casi únicamente se pueden describir como intentos de reducir la teoría física a la cosmogonía. Y aunque ambas, la cosmología y la cosmogonía, son partes enormemente fascinantes de la física, y aunque entre tanto son cada vez mejor contrastables, se encuentran al margen de la física científica y apenas son lo suficientemente maduras para servir como fundamento para una reducción de la química a la física. Ésta es una de las razones de por qué veo como incompleta y como algo problemática la llamada reducción de la química a la física; pero, naturalmente, acojo con satisfacción todos esos nuevos problemas.

XI

Pero la reducción de la química a la física deja tras de sí todavía un segundo resto. Según nuestra concepción actual, solamente el hidrógeno -y en particular su núcleo- constituye el material de construcción de todos los otros núcleos. Suponemos que los núcleos positivos, hasta el último, se repelen mucho unos a otros eléctricamente en distancias muy cortas, mientras que en distancias todavía más cortas (que sólo pueden tener lugar si las fuerzas de repulsión son superadas por velocidades enormes) se atraen mutuamente por las fuerzas de sus núcleos.

Ahora bien, esto significa que atribuimos propiedades relacionales al núcleo de hidrógeno, que son ineficaces bajo la cantidad avasalladora de las condiciones bajo las que existen núcleos de hidrógeno en nuestro universo. Es decir, que esas fuerzas nucleares son potencialidades que sólo se vuelven activas bajo condiciones extremadamente raras: bajo temperaturas sumamente elevadas y bajo una presión muy elevada. Pero esto significa que la teoría de la evolución de los sistemas periódicos se acerca a una teoría de cualidades esenciales, que se caracteriza por algo semejante a la *predestinación o armonía preestablecida*.²³ En cualquier caso, un sistema solar como el nuestro depende, según las teorías actuales, de la preexistencia de esas cualidades o, más bien, potencialidades.

Además, la teoría del origen de los elementos más pesados en explosiones-Supernova conduce a *una segunda clase de predestinación o armonía preestablecida*. Pues acaba en la afirmación de que las fuerzas gravitatorias (que aparentemente son las más débiles de todas y hasta ahora no conectadas con las fuerzas nucleares y las fuerzas electromagnéticas) pueden volverse tan poderosas en grandes acumulaciones de hidrógeno, que superan la enorme fuerza de repulsión eléctrica entre los núcleos y llegan, gracias a las fuerzas nucleares, a la fusión. En este caso, existe la armonía entre las potencialidades inherentes de las fuerzas nucleares y de la gravitación. No quiero afirmar que toda filosofía de la armonía preestablecida es necesariamente falsa. Pero sí que creo que una armonía preestablecida como reducción satisfactoria no viene al caso: más bien quisiera pensar que la apelación a ella es una confesión de que ha fracasado el método con el que se debía reducir lo uno a lo otro.

La reducción de la química a la física es, por consiguiente, todo menos completa, incluso cuando hacemos suposiciones que son favorables para la reducción de manera no realista. Más bien presupone esa reducción una teoría de la evolución cósmica, una cosmogonía y, aparte de eso, dos clases de armonía preestablecida, para posibilitar que sean activadas potencialidades latentes o propensiones relativas con escasa probabilidad, que están instaladas en el átomo de hidrógeno.

Deberíamos reconocer, creo yo, que aquí tenemos que tratar evidentemente con representaciones de *emergencia* y con *cualidades emergentes*.²⁴ Si hacemos esto, vemos que esa enormemente interesante

²³ He empleado aquí el término «armonía preestablecida» para poner de relieve que nuestra explicación no se refiere a las cualidades manifiestamente físicas del átomo de hidrógeno. Antes bien, se ha postulado una cualidad del núcleo de hidrógeno, hasta ahora desconocida e inesperada.

²⁴ Empleo aquí el término «emergente» para insinuar, según todas las ciencias, un paso evolutivo imprevisible.

reducción ha dejado tras de sí una imagen muy extraordinaria del universo -en cualquier caso, extraordinaria para los reduccionistas; y éste era el punto que me había propuesto tratar en este apartado.

XII

Para resumir lo dicho hasta ahora: he intentado aclarar el problema de la reducción de la mano de ejemplos y, con ello, he intentado mostrar que algunas de las reducciones más impresionantes en la historia de las ciencias naturales están muy alejadas de haber tenido un éxito completo, sino que más bien han dejado tras de sí un resto sobrante. Ciertamente, podría afirmarse (véase nota 5) que la teoría de Newton fue una reducción completa, exitosa de las teorías de Kepler y Galileo. Pero incluso si suponemos que entendemos mucho más de física de lo que es el caso, y que tenemos una teoría unitaria de campos que proporciona con gran aproximación la relatividad general, la teoría cuántica y las cuatro clases de fuerzas como casos excepcionales (quizá se afirma esto implícitamente en la teoría de campo unitaria de Mendel Sachs), incluso entonces podemos decir que la química no ha sido reducida enteramente a la física. Pues la física, sobre la que se efectúa la llamada reducción de la química, es una física que presupone tanto la evolución, la cosmología y la cosmogonía, como la existencia de cualidades emergentes.

Por otra parte, en nuestro intento de reducción no completamente exitoso, particularmente de la química a la física, hemos aprendido una cantidad de cosas increíble. Nuevos problemas han conducido a nuevas teorías hipotéticas y algunas de ellas, como por ejemplo la fusión del núcleo, no sólo a experimentos confirmatorios, sino también a una nueva tecnología. Desde el punto de vista metodológico nuestros intentos de reducción fueron enormemente exitosos, aunque se pueda decir que los intentos de reducción como tales usualmente han fracasado.

XIII

La historia que he contado y la enseñanza que he sacado de ella, apenas sorprenderá a un biólogo. También en la biología ha sido el reduccionismo (en forma de fisicalismo o materialismo) extraordinariamente exitoso, cuando no completamente exitoso. Pero incluso allí donde no triunfó, ha conducido a nuevos problemas y nuevas soluciones.

Quizá podría formular mi parecer de la siguiente manera: el reduccionismo ha fracasado como filosofía. Pero desde un punto de vista metodológico, los intentos reduccionistas han conducido en detalle a un desconcertante éxito tras otro, y para la ciencia incluso los fracasos fueron extraordinariamente fructíferos.

Quizá sea comprensible que algunos de aquellos que han alcanzado ese éxito científico, no estaban convencidos sin más del fracaso del reduccionismo como filosofía. Quizá les dé que pensar mi análisis del éxito y fracaso del intento de reducir la química completamente a la física cuántica, y les dé motivo para enfrentarse de nuevo con el problema.

XIV

Se podría considerar los puntos esenciales que he referido hasta el momento como una elaboración de una pequeña nota de Jacques Monod, que se encuentra en el prólogo a su libro *Azar y necesidad*: «No se puede predecir ni resolver toda la química con la ayuda de la teoría cuántica (o reducir a la teoría cuántica), que sin duda constituye, no obstante, la base universal».²⁵

²⁵ Monod, J., *Zufall und Notwendigkeit*, (1970), Piper, Munich, 1971, pág. XII. N.T.: Existe traducción castellana de Francisco Ferrer Lerín (revisada por Antonio Cortés Tejedor), del francés original, en Tusquets, Barcelona 1981; esta cita se encuentra en la pág. 10.

En el mismo libro Monod formula una hipótesis (naturalmente ninguna aserción) respecto al origen de la vida que es harto sorprendente y que podemos considerar desde el punto de vista aquí alcanzado. La hipótesis de Monod consiste en que la vida se originó a partir de la materia inanimada por una combinación de circunstancias azarosas, en extremo poco probables, y que esto debió ser no sólo un suceso escasamente probable, sino incluso un suceso con probabilidad cero -un suceso *único*.

Esta hipótesis es contrastable experimentalmente (como Monod ha puesto de manifiesto recientemente en una discusión con Eccles). Si tuviéramos éxito con la producción de vida bajo ciertas condiciones bien definidas, se refutaría la hipótesis de la unicidad de la vida. Por consiguiente, se convertiría la hipótesis en una hipótesis contrastada científicamente, aunque no fuera ésta su primera intención.

¿Por medio de qué es la hipótesis de Monod plausible en general? En primer lugar, contamos con el hecho de la unicidad (aproximada) del código genético, lo que podría ser, sin embargo, el resultado de la selección natural, como Monod expone. Lo que hace del origen de la vida y del código genético un enigma desconcertante es que el código genético no tiene ninguna función biológica, a no ser que se traduzca, es decir, a no ser que conduzca a la síntesis de las proteínas, cuya estructura queda determinada por el código. Pero, como Monod desarrolla, la maquinaria, por medio de la cual la célula (al menos la célula no primitiva -la única célula que nosotros conocemos) traduce el código, comporta «al menos cincuenta componentes macromoleculares, *que están ellos mismos codificados en el ADN*»,²⁶ y, por ello, (¡al menos hoy!) su propia existencia. Según esto, el código sólo se puede traducir si se emplean en ello ciertos productos de su traducción. Por tanto, nos encontramos aquí con un evidente círculo desconcertante: un *circulus vitiosus* que es válido para cada intento de desarrollar un modelo o una teoría de la génesis del código genético.

Así pues, podríamos vernos confrontados con la posibilidad de que el origen de la vida (como el origen del universo) se convierta en un obstáculo impenetrable para la ciencia y que quede como resto en todos los intentos de reducir la biología a la física y química. Pues incluso si la hipótesis de Monod de la unicidad del origen de la vida es refutable -precisamente por los intentos de reducción-, acabaría, en el caso de que sea verdadera, en una reducción exitosa. Monod, que es reduccionista por motivos metodológicos, llega con esta hipótesis a una posición, a la que nosotros estamos abocados, en mi opinión, por nuestra anterior discusión acerca de la reducción de la química a la física: a la posición de un reduccionista crítico que continúa con sus ensayos reduccionistas, aun cuando no tenga ninguna esperanza en un éxito último. Nuestra esperanza descansa en esos intentos continuados de reducción y no en el reemplazo de los métodos reduccionistas por los «holistas», como Monod subraya en otro pasaje de su libro -nuestra esperanza de saber más de antiguos problemas y de descubrir nuevos problemas, que puedan ayudarnos a conseguir, por su parte, nuevas soluciones, nuevos descubrimientos.

No quiero discutir aquí el holismo en detalle, pero sí me gustaría decir un par de palabras al respecto. Podemos afirmar, por la aplicación de métodos holistas experimentales (como los trasplantes de células en embriones), que éstos son reduccionistas en sentido metodológico, aun cuando esos métodos estén inspirados por ideas holistas. Por otra parte, las teorías holistas son ya necesarias, trivialmente, en la descripción de átomos y moléculas, por no hablar de organismos y poblaciones de genes. No hay ningún límite²⁷ para la pluralidad de suposiciones posiblemente fructíferas, sean o no holistas. Por lo que respecta a mi tesis fundamental, se discute exclusivamente el carácter de los métodos experimentales en la biología, esto es, si no todos tienen carácter más o menos reduccionista. Por lo demás, David Miller me recuerda que se da una situación semejante respecto a las teorías deterministas e indeterministas. Aunque creo que debemos ser indeterministas metafísicos, deberíamos buscar, sin embargo, leyes deterministas o causales -excepto allí donde los problemas por solucionar son en sí mismos de carácter probabilista.

²⁶ *Ibid.* pág. 143. N.T.: véase pág. 155 de la ed. cast. citada.

²⁷ Subrayaré esto especialmente en el punto (2) del epílogo a este trabajo.

XV

Quisiera señalar que incluso en el caso de que la hipótesis de Monod de la unicidad del origen de la vida fuera refutada porque consiguiéramos producir vida a partir de materia inanimada, bajo determinadas condiciones experimentales, esto no acabaría en una reducción completa. No quisiera concluir *a priori* que una reducción es posible; pero durante mucho tiempo hemos producido vida a partir de la vida, sin entender lo que hacíamos; e incluso antes de tener la más mínima idea de biología molecular o de código genético. Seguramente sea posible que consigamos producir vida a partir de materia inanimada, sin que dispongamos de un entendimiento físico-químico completo de lo que realmente hacemos -por ejemplo, cómo hemos conseguido romper el *circulus vitiosus* que se encuentra en la base de la traducción del código.

En cualquier caso, podemos decir que la irrupción completamente inesperada de la biología molecular ha convertido el problema del origen de la vida en un enigma todavía mayor de lo que era antes: nos hemos creado problemas nuevos y más profundos.

XVI

Como he intentado mostrar, el intento de reducir la química a la física requiere la introducción de una teoría de la evolución en la física, es decir, un recurso a la historia de nuestro cosmos. Todavía más indispensable es, tal y como parece, una teoría de la evolución en la biología. Igualmente indispensable es la idea de una finalidad o de una teleología o (para emplear un término de Monod) una teleonomía, o sea, la idea muy semejante de la solución del problema -una idea que es completamente ajena al dominio del tema de las ciencias no biológicas (aun cuando el papel que se juega en esas ciencias por los *máxima* y *mínima* y por el cálculo de variaciones, no lo podemos ver ni de forma remotamente análoga).

Naturalmente, el gran logro de Darwin fue haber mostrado que existe una posibilidad de explicar la teleología en conceptos no teleológicos o usualmente causales. El darwinismo es la mejor explicación que tenemos. Por el momento, no existe ninguna hipótesis seria que le haga la competencia.²⁸

XVII

Los problemas y las soluciones de los problemas parece que se han originado junto con la vida. Aun cuando parece actuar algo así como una selección natural antes del origen de la vida -por ejemplo, una selección de los elementos más estables gracias a la destrucción radioactiva de los menos estables-, no podemos decir que el sobrevivir sea un «problema» en cualquier sentido del término para los núcleos de los átomos. Y la estrecha analogía entre cristales y microorganismos y sus partes moleculares se derrumba precisamente aquí. Los cristales no conocen ningún problema del crecimiento o de la reproducción o del sobrevivir. Sin embargo, la vida se encuentra confrontada con el problema de la supervivencia desde el principio: sí, podemos, si queremos, describir la vida sencillamente como resolver problemas y los organismos vivientes como los únicos complejos en el universo que resuelven problemas. (Los ordenadores son *instrumentos* para resolver problemas, pero no resuelven problemas en ese sentido.)

Con esto no quiere decirse que tengamos que atribuir a toda vida una *conciencia* de los problemas a resolver: incluso en el terreno humano nos ocupamos continuamente de resolver problemas -por ejemplo, el problema de mantener el equilibrio-, sin que seamos conscientes de ello.

²⁸ Véase *Objektive Erkenntnis* [Conocimiento objetivo], cap. 7.

XVIII

Está casi fuera de duda que los animales tienen conciencia e incluso que transitoriamente pueden ser conscientes de un problema. Pero es posible que el origen de la conciencia en el reino animal sea un misterio tan grande como el origen de la vida misma.

Sobre este asunto no quisiera decir nada más que no me parece que el pansiquismo o el hilozoísmo sean de la menor ayuda, con la tesis de que la materia en general (aunque sea en un grado ínfimo) está dotada de conciencia. No constituye otra cosa, en la medida en que puede tomarse en serio, que una teoría más de la predestinación o de la armonía preestablecida. (Naturalmente, formaba parte de la teoría de Leibniz de la armonía preestablecida en su forma originaria.) Pues la conciencia no tiene ninguna función en la materia inanimada; y si atribuimos (con Leibniz, Diderot, Buffon, Haeckel y muchos otros) conciencia a las partículas inanimadas (mónadas, átomos, partículas), lo hacemos con la vana esperanza de beneficiarnos de ello para explicar la existencia de formas de conciencia en los animales, que tienen en éstos una función tan importante.

Pues no puede ponerse en duda que la conciencia tiene una función en los animales y que se tiene que considerar casi como un órgano corporal. Por muy difícil que sea de aceptar: tenemos que suponer que se trata de un producto de la evolución, de la selección natural.

Aunque con esto podría formularse el programa para una reducción, no es en sí mismo una reducción, y la situación se presenta más o menos carente de esperanza para los reduccionistas; lo que explica por qué los reduccionistas han negado en general, o bien la apropiación de la hipótesis del pansiquismo, o bien -en el pasado más remoto- la existencia de conciencia (como la conciencia de un dolor de muelas).

A pesar de que la filosofía behaviorista esté bastante en boga actualmente, en mi opinión no puede tomarse más en serio una teoría de la existencia de la conciencia que una teoría de la no existencia de materia. Ambas teorías «resuelven» el problema de la relación entre alma y cuerpo. En ambos casos, la solución consiste en una simplificación radical: la negación del alma o del cuerpo. Pero en mi opinión, esta solución es demasiado simple.²⁹ En el apartado XXI, donde critico el paralelismo psicofísico, añadiré algo más sobre esta segunda «cuestión esencial», y en particular sobre el pansiquismo.

XIX

Sobre las tres cuestiones fundamentales de la reducción que anunciaba al comienzo de esta conferencia, he discutido ya brevemente dos. Llego ahora a la tercera, la cuestión de la reducción de la autoconciencia humana y de la creatividad del espíritu humano.

Como sir John Eccles lo ha subrayado repetidas veces, esta tercera cuestión consiste en el problema de la «conexión entre espíritu y cerebro»; y Jacques Monod denomina al problema del sistema central nervioso humano el «segundo frente», cuyas dificultades son enteramente comparables con el «primer frente», el problema del origen de la vida.

Sin duda, no carece de peligro detenerse en este segundo frente, sobre todo para un profano en biología; sin embargo, quisiera decir que me parecen más prometedores los intentos de una reducción parcial en este terreno que en la segunda cuestión. Lo mismo que en el campo de la primera cuestión, me parece que con los métodos reduccionistas pueden descubrirse, y quizá incluso solucionarse, muchos

²⁹ Véase *ibid.*, cap. 8.

más nuevos problemas que en el campo de la segunda cuestión -un campo que, comparativamente, se presenta estéril. Quizá no necesito subrayar que me parece sumamente improbable una reducción completamente exitosa en las tres regiones, si no completamente imposible.

Con esto podría decirse que he cumplido mi promesa de discutir, como anunciaba al comienzo de esta conferencia, los problemas denominados esenciales de la reducción o, en cualquier caso, de mencionarlos. Con todo, me gustaría insistir un poco más en el tercero -el problema de la relación cuerpo-alma o de la relación alma-cuerpo- antes de pasar a mi tesis de que toda ciencia es incompleta y no puede llegar a completarse.

XX

Considero como altamente improbable resolver el problema de la emergencia de la conciencia en los animales (cuestión b), el problema de entender la conciencia y, eventualmente, reducirla a la fisiología; y pienso de manera semejante respecto al otro problema de la emergencia de la autoconciencia específicamente humana (cuestión c), es decir, del problema cuerpo-alma. No obstante, creo que podemos arrojar al menos un poco de luz sobre el problema del yo humano.

En muchos aspectos soy un dualista cartesiano,³⁰ aun cuando prefiero caracterizarme como un pluralista; y, naturalmente, no creo en las dos «sustancias» de Descartes. La materia no es, como hemos visto, ninguna «sustancia» última con la cualidad esencial de la extensión, sino que consta de complejas estructuras, de cuya composición sabemos muchísimo, incluida una explicación (parcial) de su «extensión»: «Ocupa espacio» (o «es extensa») a causa del choque eléctrico entre las partículas.

Mi primera tesis es que la autoconciencia humana es extremadamente compleja en toda su aparentemente irreductible unidad y que quizá se puede aclarar en parte.

En una serie de conferencias que pronuncié en mayo de 1969 en la Universidad de Emory (lo mismo que algunos años antes en conferencias en la School of Economics de Londres), defendí la concepción de que la conciencia humana más elevada, o sea, la conciencia de un yo, no se encuentra en los animales. He continuado defendiendo la concepción de que la suposición de Descartes, según la cual el alma humana está localizada en la glándula pineal, no necesita ser tan absurda como a menudo se la ha hecho aparecer, y que, en vistas de los resultados que Sperry ha conseguido con hemisferios del cerebro partidos,³¹ la localización podría buscarse en el centro del lenguaje, en el hemisferio izquierdo del cerebro. Como Eccles ha expuesto recientemente,³² esa suposición viene apoyada en cierta manera por los experimentos posteriores de Sperry (que en aquella época no me eran conocidos): después se puede describir el cerebro derecho como el de un animal muy inteligente, mientras que sólo el cerebro izquierdo parece ser específicamente humano y consciente de sí mismo.

Apoyé mi suposición en el papel que le atribuyo al desarrollo de un lenguaje específicamente humano.

Todo lenguaje animal -si no todo comportamiento animal-tiene, como Karl Bühler ha puesto de manifiesto, una función *expresiva* y una *comunicativa* (o señalizadora). El lenguaje humano, por su parte, tiene junto a éstas algunas otras funciones que son características y que le convierten en un «lenguaje» en sentido estricto e importante del término. Bühler llamó la atención sobre la función básica *descriptiva* del lenguaje humano, y yo he subrayado más tarde³³ que tiene otras funciones (como la prescriptiva,

³⁰ *Conjeturas y refutaciones*, cap. 3; véase lugar citado en nota 12.

³¹ Sperry, R.W. (1964), «The great cerebral commissure», *Scientific American* 210, págs 42-52; y Eccles, J.C. (1970), *Facing Reality*, Springer, Berlín, Heidelberg, Nueva York, págs 73-79.

³² Eccles, J.C. (1972), *Unconscious actions emanating from the human cerebral cortex* (no publicado).

³³ *Conjeturas y refutaciones*, cap. 3

consultiva, etc.), de las cuales la más importante y característica para el nombre es la *argumentativa*. (El profesor Alf Ross³⁴ resalta que pueden añadirse todavía muchas otras funciones, por ejemplo, la de ordenar, pedir o prometer.)

No creo (y no lo he creído nunca) que alguna de estas funciones sean reducibles a algunas otras, al menos las dos funciones más elevadas (descripción y argumento) a las dos inferiores (expresión y comunicación). Además, éstas son siempre y en todos los casos interesadas, lo que quizá sea la razón que explique por qué muchos filósofos las consideran erróneamente cualidades que son características para el lenguaje humano.

Mi tesis es que con las funciones elevadas del lenguaje humano se origina un nuevo mundo: el mundo de los productos del espíritu humano. He denominado a ese mundo «mundo 3» (siguiendo un consejo de sir John Eccles: originariamente lo denominaba «tercer mundo»). Así, denomino «mundo 1» al mundo de la materia física, campos de fuerza, etc.; «mundo 2» al mundo de las experiencias conscientes y quizá también inconscientes; y «mundo 3» en particular al mundo del lenguaje hablado (escrito o impreso), como contar historias, inventar mitos, teorías, problemas teóricos, errores y argumentos. (Los mundos de los productos artísticos y de las instituciones sociales son, o bien clasificados bajo el «mundo 3», o bien denominados «mundo 4» y «mundo 5»: se trata de una mera cuestión de gusto.)

Introduzco los conceptos «mundo 1», «mundo 2», «mundo 3», para subrayar la (limitada) *autonomía* de ese campo. La mayoría de los materialistas, fiscalistas o reduccionistas afirman que de esos tres mundos, sólo el mundo 1 existe realmente y que, por ello, es autónomo. Reemplazan el mundo 2 por comportamiento y el mundo 3 especialmente por comportamiento verbal. (Esta es, como he señalado más arriba, una de aquellas soluciones demasiado simples del problema alma-cuerpo, a saber, negar la existencia del espíritu humano y de la autoconciencia humana -de aquellas cosas que yo cuento entre las más sorprendentes y dignas de atención en el universo; la otra salida simple es el inmaterialismo de Berkeley y Mach: la tesis de que sólo existen sensaciones y de que la materia no es otra cosa que una «construcción» a partir de sensaciones.)

XXI

Respecto a la relación entre cuerpo (o sea, cerebro) y alma, podemos distinguir en lo esencial cuatro posiciones:

1. La negación de la existencia del mundo 1 de los estados corporales; es decir, un inmaterialismo, como el que defendieron tanto Berkeley como Mach (*El análisis de las sensaciones*).
2. La negación de la existencia del mundo 2 de los estados o acontecimientos espirituales, una concepción de ciertos materialistas, fiscalistas y filósofos behavioristas, es decir, filósofos que defienden la «identidad» de cerebro y espíritu.
3. La afirmación de un paralelismo universal entre estados espirituales y estados cerebrales, una posición que se denomina «paralelismo psicofísico». Fue introducida primeramente en el cartesianismo por Geulincx, Spinoza, Malebranche y Leibniz, fundamentalmente para evitar ciertas dificultades de la concepción cartesiana. (De manera semejante al epifenomenalismo, roba a la conciencia toda función biológica.)
4. La afirmación de que los estados espirituales y corporales se pueden influir recíprocamente. Ésta era la concepción de Descartes, la cual, como acabamos de admitir, es superada por 3.

³⁴ Ross, A. (1972), «The rise and fall of the doctrine of performatives», en *Contemporary Philosophy in Scandinavia* (comps. R.E. Olsen y A.M. Paul), Johns Hopkins Press, Baltimore, págs. 197-212.

Mi propia posición es que *debe* existir un paralelismo-espíritu-cerebro *hasta un cierto punto*. Ciertos reflejos, como el parpadear cuando vemos un objeto que se aproxima de repente, son aparentemente de más o menos carácter paralelista: la reacción muscular (en la que evidentemente participa el sistema nervioso central) se repite regularmente si se repite la impresión visual. En caso de que nuestra atención se dirija a ella, podemos volvernos también conscientes de esa reacción, y lo mismo vale para algunos otros reflejos (aunque naturalmente no para todos),

Pero sin embargo creo que la tesis de un paralelismo psicofísico *completo* -posición 3- es equivocada, y probablemente ya para los casos en los que se trata de meros reflejos. *En este sentido, quisiera defender una forma de interacción psicofísica*. Ésta contiene (como ya vio Descartes) *la tesis de que el mundo corporal 1 no está cerrado causalmente, sino que está abierto para el mundo 2* (originado mucho después) de los estados y sucesos espirituales -para los físicos se trata de una tesis no especialmente atractiva, pero una tesis que, como yo creo, está apoyada por el hecho de que el mundo 3 (incluidos sus campos autónomos) actúa sobre el mundo 1 *a través* del mundo 2.

Estoy dispuesto por completo a aceptar la opinión de que siempre que ocurre algo en el mundo 2, ocurre algo conectado con ello en el mundo 1 (en el cerebro). Pero para hablar de un paralelismo completo o universal, deberíamos estar en la situación de afirmar que «el mismo» estado espiritual o «el mismo» suceso espiritual va unido siempre a un estado fisiológico que se corresponde exactamente con él y *viceversa*.

Como ya he dicho, estoy enteramente dispuesto a admitir que se esconde algo correcto en esa afirmación y que, por ejemplo, la estimulación eléctrica de ciertas regiones del cerebro provoca regularmente ciertos movimientos o sensaciones características. Pero me pregunto si la afirmación, entendida como regla general para todos los estados espirituales, tiene en general un contenido; si no está quizá vacía. Pues podemos tener ciertamente un paralelismo entre mundo 2 y procesos cerebrales o entre mundo 2 -formas y procesos cerebrales, pero apenas podemos hablar de un paralelismo entre un proceso del mundo 2 sumamente complejo, único y no analizable y un proceso cerebral. Y existen muchos sucesos-mundo 2 en nuestra vida que son únicos. Incluso si dejamos de lado por una vez el problema de la innovación creativa, oír dos veces una melodía y reconocer que es la misma no es simplemente una repetición del mismo acontecimiento del mundo 2, puesto que la segunda audición de la melodía está conectada con un acto de *reconocimiento* de la melodía que ha faltado en la primera vez que la escuchamos. Lo que se repite es el objeto del mundo 1 (en ese caso la melodía), pero no el suceso del mundo 2. Sólo si nos determináramos a aceptar una teoría-mundo 2 que ve, lo mismo que la psicología de la asociación, los acontecimientos-mundo 2 compuestos de elementos casi atómicos, podríamos encontrar una clara diferenciación entre la parte repetida de la vivencia-mundo 2 -el *escuchar* la misma melodía- y la parte no repetida -el *reconocer* que se trata de la misma melodía- (donde la experiencia del reconocer puede regresar por su parte en otros contextos). Debería estar, pues, claro que una psicología atomística o analítica no nos lleva muy lejos.

El mundo 2 es extremadamente complejo. Mientras que sólo reparamos en campos como la percepción sensorial (es decir, la percepción de objetos-mundo 1), creemos quizá poder analizar el mundo 2 con métodos atomísticos o moleculares, por ejemplo con métodos-Gestalt (métodos que, en mi opinión, son completamente improductivos en comparación con los métodos biológicos o funcionales de Egon Brunswick o Richard Gregory). La aplicación de semejantes métodos se muestra como totalmente inadecuada, si pensamos en nuestros intentos únicos de inventar y comprender un objeto mundo-2, por ejemplo, un problema o una teoría.

La forma en que interactúa nuestro pensar y entender con el intento de una formulación lingüística, y se influye por ello; la forma en que primeramente tenemos un sentimiento vago por un problema o una teoría, que se vuelve más claro cuando intentamos formularlo, y todavía más claro si lo

ponemos por escrito, y en la que tomamos en apariencia críticamente nuestros intentos de resolverlo; la forma en que un problema se transforma y, sin embargo, puede permanecer en cierto sentido el viejo problema; la forma en que una ilación de las ideas está, por una parte, en sí misma relacionada, y por otra parte, claramente articulada: todo esto me parece estar fuera del campo de aplicación de los métodos analíticos o atomísticos, incluidos los interesantes métodos moleculares de la psicología de la Gestalt. En todos estos intentos entra en juego una historia única de acontecimientos únicos-mundo 2, y consecuentemente pierde todo contenido el discurso de procesos (estrictamente) paralelos fisiológicamente.

Más allá de esto, tenemos razón para admitir que si una determinada zona del cerebro está destruida, otra zona, a menudo con un insignificante y quizá incluso ningún menoscabo del mundo 2, se hace cargo de su función -otro argumento contra el paralelismo, un argumento que se apoya antes en experimentos del mundo 1 que en cualquier forzada reflexión vaga sobre las experiencias complejas del mundo 2.

Todo esto suena muy antirreduccionista; y como filósofo que contempla nuestro mundo, con nosotros dentro de él, creo en el hecho, no en la posibilidad de una última reducción. Pero como metodólogo, esto no me conduce de ninguna manera a un programa de investigación antirreduccionista. Me lleva únicamente a la predicción de que nuestro saber y los problemas irresueltos de nuestro universo se extenderán con el desarrollo de nuestros intentos de reducción.

XXII

Volvamos ahora al problema de la autoconciencia específicamente humana; la concepción que había defendido era que la autoconciencia surge de la interacción (feedback, si se prefiere) entre el mundo 2 y los mundos 1 y 3. Mis argumentos para el papel que el mundo 3 juega en ello, son los siguientes: la autoconciencia humana se funda, entre otras cosas, en un número de *teorías* sumamente abstractas. Los animales e incluso las plantas tienen sin duda un sentido del tiempo y esperanzas temporales. Pero se necesita (sin pretender acercarse demasiado a Benjamín Lee Whorf) de una *teoría* casi explícita del tiempo, para verse a sí mismo como alguien que posee un pasado, un presente y un futuro; alguien que tiene una idea personal y es consciente de su identidad personal en esa historia (que está conectada con la identidad de su cuerpo). Según esta *teoría*, durante el tiempo del sueño, aunque perdamos la continuidad de la conciencia, nosotros -nuestros cuerpos- son esencialmente los mismos; y sobre la base de esta teoría, sucede que podemos traer a la memoria acontecimientos pasados (en lugar de ser meramente influidos en nuestras expectativas y reacciones por ellos, en lo que veo la forma primitiva que tiene la memoria en los animales).

Sin duda, algunos animales tienen personalidad; tienen algo perfectamente análogo al orgullo y la ambición, y aprenden a reaccionar a un nombre. La autoconciencia humana, por el contrario, está anclada en el lenguaje y (tanto implícita como explícitamente) en teorías formuladas. Un niño aprende a formular su nombre para denominarse a sí mismo y sólo después una palabra como «ego» o «yo», y aprende su uso con la conciencia de la continuidad de su cuerpo y su yo; lo acumula junto con el saber de que la conciencia no siempre está ininterrumpida. La gran complejidad y la no autonomía del alma humana o del yo humano, se vuelven particularmente claros si nos acordamos de que existen casos en los que los seres humanos han olvidado quiénes son; han olvidado una parte o el conjunto de su historia pasada, pero al menos han conservado, o quizá recuperado, una parte de su yo. En cierto sentido su memoria no se ha perdido, pues *recuerdan cómo* se anda, come e incluso habla. Pero *no se acuerdan de que* proceden de Bristol, o de cuál es su nombre y su dirección. En la medida en que no encuentran el camino de casa (lo que los animales hacen normalmente), su autoconciencia se encuentra afectada incluso por encima del nivel normal de la memoria animal. Pero en caso de que no hayan perdido la capacidad del lenguaje, ha quedado conservada una conciencia humana, que va más allá de la conciencia animal.

No soy ningún gran amigo del psicoanálisis, pero sus hallazgos parecen apoyar la concepción de la complejidad del yo humano opuesta al recurso cartesiano a una sustancia pensante. Lo que me interesa es contrastar que la autoconciencia humana contiene al menos una conciencia de la continuidad (sumamente teórica) temporal e histórica del propio cuerpo; una conciencia de la conexión entre la propia memoria consciente y el cuerpo único que pertenece a uno mismo; y la conciencia de la interrupción normal y periódica de la propia conciencia por medio del sueño (que presupone a su vez una teoría del tiempo y de la periodización temporal. Por encima de esto, la conciencia incluye pertenecer espacial y socialmente a un determinado lugar y círculo de personas. Sin duda, mucho de esto tiene una base instintiva y también es propio de los animales. Mi tesis es que, aun cuando se eleva al terreno de la conciencia humana no hablada, el lenguaje humano, esto es, la interacción entre los mundos 2 y 3, desempeña un papel importante.

Está claro que la unidad del yo humano se basa considerablemente en la memoria y que los rendimientos de la misma no sólo se pueden atribuir a los animales, sino también a las plantas (y en cierto sentido quizá incluso a las estructuras no orgánicas, como los imanes). Por ello, es sumamente importante ver que el recurso a la memoria como tal no es suficiente para aclarar la unidad del yo humano. No es tan necesaria la memoria «usual» (sucesos pasados), sino una memoria de teorías que ponga en conexión la conciencia de tener un cuerpo con el mundo 3-teorías sobre el cuerpo (es decir, con la física); una memoria que tiene el carácter de un «comprender» por el mundo 3-teorías. Contiene las disposiciones que nos posibilitan traer a la memoria en caso de necesidad teorías explícitas del mundo 3, con la conciencia de que poseemos tales disposiciones y podemos utilizarlas en caso de necesidad para articular cualquier teoría. (Esto explicaría, naturalmente, en cierta medida, la diferencia entre la conciencia animal y la autoconciencia humana, con su dependencia del lenguaje humano.)

XXIII

Esos hechos me parecen fundamentar la imposibilidad de cualquier reducción del mundo 2 humano, del mundo de la conciencia humana, al mundo humano 1, es decir, en lo esencial, a la psicología cerebral. Pues el mundo 3 es al menos en parte independiente de los otros dos mundos. Si la parte independiente del mundo 3 puede estar en relación de interacción con el mundo 2, entonces el mundo 2 no puede, así me lo parece, ser reducible al mundo 1.

Mis ejemplos estándar para la autonomía parcial del mundo 3 proceden de la aritmética.

Propongo considerar la serie infinita de los números naturales como una invención, un producto del espíritu humano y como una parte del lenguaje humano desarrollado. (Parece haber lenguajes primitivos en los que sólo se puede contar «uno, dos, muchos», y otros en los que sólo se puede contar hasta cinco.) Pero tan pronto como es inventado un método de los números finitos, surgen diferencias y problemas autónomamente: los números pares e impares no son *inventados*, sino *descubiertos* en la serie de los números naturales, e igualmente los números primos y los muchos problemas resueltos e irresueltos que tienen que ver con ellos.

Esos problemas y los teoremas que los resuelven (como el teorema de Euclides de que no existe un número primo mayor que sea mayor que cualquier otro) surgen autónomamente -como parte de la estructura interna de la serie de los números naturales creada por los seres humanos e independiente de aquello que pensemos o no pensemos. Pero podemos entender, comprender o descubrir esos problemas y resolver algunos de ellos. Con esto, nuestro pensamiento, que pertenece al mundo 2, depende en parte de los problemas autónomos y de la verdad objetiva de teoremas, los cuales pertenecen al mundo 3: el mundo 2 no sólo crea al mundo 3, sino que es creado en parte por el mundo 3 en una especie de proceso de retroacoplamiento.

Mi argumento es, pues, el siguiente: el mundo 3 y en particular sus partes independientes no son reducibles evidentemente al mundo corporal 1. No obstante, puesto que el mundo 2 depende en parte del mundo 3, tampoco es reducible al mundo 1.

Fisicalistas o filósofos reduccionistas, como yo los he denominado,³⁵ están obligados, de esta manera, a negar la existencia de los mundos 2 y 3. Pero con ello se vuelve incomprendible toda la tecnología humana (en particular la existencia de ordenadores), que hace uso en tan gran medida de los teoremas-mundo 3; y tenemos que admitir que tan enormes cambios en el mundo 1, como son producidos por los constructores de aeropuertos o rascacielos, están producidos en última instancia, sin la invención del mundo 3-teorías o del mundo 2-planos, por el mundo corporal 1 mismo, en el que se basan. Están predestinados; son parte de una armonía preestablecida que ya está colocada en última instancia, en los núcleos de hidrógeno.

Esos resultados me parecen absurdos; y el behaviorismo filosófico o el fisicalismo (o sea, la filosofía de la identidad-cuerpo-alma) me parece que se puede reducir a esta absurdidad. Me parece que se aleja demasiado del sano entendimiento humano.

XXIV

El reduccionismo filosófico es, como creo, una equivocación. Proviene del deseo de reducirlo todo a una última explicación por medio de esencias y sustancias, es decir, a una explicación que no sea capaz ni necesite de una explicación posterior. Tan pronto como le concedemos a la teoría una última explicación, nos damos cuenta de que podemos seguir preguntando «¿por qué?». Las preguntas en busca de un porqué no conducen nunca a una última respuesta. Los niños inteligentes parecen saber esto, aun cuando terminan por condescender ante los adultos, quienes de hecho es imposible que tengan suficiente tiempo para responder en principio a una serie infinita de preguntas.

XXV

Aun cuando los mundos 1, 2 y 3 son en parte autónomos, pertenecen al mismo universo: tienen una interacción recíproca entre sí. No obstante, se puede mostrar fácilmente que el conocimiento del universo, en la medida en que él mismo es una parte del universo (como de hecho es el caso), necesariamente no se puede completar.

Imaginemos un hombre que dibuja un mapa exacto de la habitación donde trabaja. Dejémosle intentar que incluya en su dibujo del mapa el mapa que en ese momento dibuja. Está claro que esta tarea, que incluye un infinito número de mapas cada vez más pequeños dentro de cada mapa, no se puede completar: cada vez que añade un nuevo trazo en el mapa, crea un nuevo objeto que hay que dibujar, pero que todavía no está dibujado. El mapa, que debe contener un mapa de sí mismo, no se puede completar.

La historia del mapa muestra la incompletitud y apertura de un universo que contiene objetos de conocimiento-mundo 3. Además, se puede utilizar también como argumento de que nuestro universo es indeterminista. Pues, mientras cada uno de los diferentes «últimos» trazos que son incluidos de hecho en el mapa, determina por añadidura un trazo dependiente de aquéllos dentro de la serie infinita de mapas dibujados, la determinación del trazo es sólo válida mientras que no consideremos la falibilidad de todo el conocimiento humano (una falibilidad que juega un papel considerable en los problemas, teorías y errores del mundo 3). Tomemos esto en consideración: cada uno de esos «últimos» trazos, que

³⁵ *Objektive Erkenntnis [Conocimiento objetivo]*, cap. 8.

son incluidos en el mapa, representan un *problema* para el dibujante, el problema de incluir un trazo más que represente *exactamente* el último trazo. A causa de la falibilidad que le es propia a todo el conocimiento humano, es imposible que el problema del dibujante se pueda resolver con absoluta exactitud; y cuanto más pequeños son los trazos que realiza el dibujante, tanto mayor será la relativa inexactitud, la cual, en principio, es impredecible e indeterminada, y aumenta continuamente. De esta manera, la historia del mapa muestra cómo la falibilidad que es propia al conocimiento humano objetivo, también colabora al indeterminismo esencial de nuestro universo, sin contar con que la apertura y la dificultad de reconocimiento de un universo, muestra que éste contiene al conocimiento humano como parte de sí mismo.

Este ejemplo nos puede ayudar a adquirir el conocimiento de por qué una ciencia explicativa nunca puede ser completa; pues para completarla, tendríamos que dar una explicación de ella misma.

Un resultado todavía más sólido está concluso en el famoso principio de Gödel, que muestra que es imposible completar un sistema de la aritmética formalizada (aunque tengo que admitir que con la atracción del principio de Gödel y otros principios matemáticos incompletos en este contexto, ofrezco una protección muy considerable contra una posición débil de modo comparativo). Puesto que todas las ciencias físicas emplean la aritmética (y puesto que para un reduccionista sólo tiene realidad la ciencia que es formulada en símbolos físicos), se vuelven incompletas todas las ciencias físicas a través del principio de incompletitud de Gödel; lo que debería convencer a un reduccionista de que la ciencia en su conjunto es incompleta. Para un no reduccionista, que no cree en la reductibilidad de la ciencia en su conjunto a la ciencia físicamente formulada, la ciencia es, sin más, incompleta.

No sólo el reduccionismo filosófico es un error, sino que también parece equivocada la suposición de que el método de reducción puede alcanzar una reducción completa. Vivimos, como parece, en un mundo de la evolución emergente, en un mundo de problemas, cuyas soluciones, en la medida en que triunfan, producen problemas nuevos y más profundos. Vivimos, por consiguiente en un universo de novedad en desarrollo -de una novedad que, por lo general, no es reductible por completo a ningún estado precedente.

Sin embargo, el método del intento de reducción es sumamente fructífero, no sólo porque aprendemos muchísimo de sus éxitos parciales a causa de reducciones que han tenido un éxito parcial, sino también porque aprendemos de nuestros fracasos parciales, a partir de los nuevos problemas que nuestros fracasos sacan a la luz. Los problemas abiertos son casi tan interesantes como sus soluciones; éstas serían igualmente interesantes si cada solución inaugurara por su parte un nuevo mundo de problemas abiertos.

EPÍLOGO

Salvo pequeñas revisiones y una o dos referencias a este epílogo, he dejado la conferencia en su forma originaria. No obstante, antes de que se discutiera en el Congreso, la critiqué yo mismo, señalando como importantes las cuatro siguientes observaciones:

1. En la conferencia no se citan los intentos de una reducción de la termodinámica a la mecánica. Éste es un ejemplo importante de una reducción y, bajo el punto de vista de mi tesis, interesante sobre todo porque, a pesar de la importancia de los resultados de los intentos de solución, no llega a ninguna reducción completa, sin resto.
2. Una segunda y más importante observación se refiere a un punto que considero en el texto de la conferencia como más o menos evidente (sólo lo he mencionado brevemente en el apartado XIV; véase el texto en la nota 24). Se trata de lo siguiente: antes de poder hacer el intento de una

reducción, necesitamos un saber lo más grande y exacto posible de aquello que queremos intentar reducir. Antes de poder intentar hacer una reducción (es decir, en el nivel de las «totalidades») tenemos que movernos en el terreno de aquello que se debe reducir. Ya he llamado la atención al respecto en anteriores escritos.³⁶

3. Una tercera observación (que no se mencionó en el Congreso) está conectada con la distinción (a la que he aludido al comienzo de la conferencia: véase el texto, nota 3) entre una *reducción*, que explica una teoría por medio de una teoría existente, y una *explicación por medio de una nueva teoría*. Aun cuando no quiero discutir acerca de los términos, no sería adecuado denominar en el presente «reducción» a una explicación por medio de una nueva teoría. En el caso de que se acepte esa terminología, podría afirmarse que la explicación de la teoría de las ondas de la propagación de la luz por medio de la teoría de Maxwell del electromagnetismo, supone un ejemplo para una reducción completa exitosa (quizá el único ejemplo para una reducción completa exitosa). Sin embargo, se podría recomendar no describir esto como una reducción de una teoría a otra o de una parte de la física a otra, sino, en lugar de esto, como una nueva teoría radical que une con éxito dos campos de la física.
4. Sin querer abogar por lo que podría llamarse un programa de investigación antirreduccionista para la biología, lo que sigue me parece que es una valoración razonable de la situación: el programa newtoniano mecanicista de la física fracasó en el intento de incluir la electricidad y el magnetismo en su campo de validez o, dicho más exactamente, en la introducción de Faraday de las fuerzas no centrales. (El intento de Maxwell de reducir esas fuerzas no centrales a la teoría newtoniana por la construcción de un modelo mecánico del éter, se mostró como sumamente fructífero, en la medida en que lo propuso para sus ecuaciones de campo, pero, sin embargo, no tuvo éxito y se debió abandonar.)

El conocimiento de Einstein de que las teorías de Newton y Maxwell son incompatibles, condujo a la teoría de la relatividad especial. Por lo tanto, los físicos preferían antes una nueva teoría radical que aceptar una reducción. Un destino semejante sorprendió a la física, cuando tanto la mecánica como la teoría electromagnética (retrocediendo hacia Lorentz y Einstein) se dirigieron de forma unitaria a problemas nuevos y considerablemente estáticos de la microestructura de la materia. Esto condujo a la mecánica cuántica. No podemos excluir la posibilidad de que la inclusión de problemas biológicos conduzca a una expansión y revisión posterior de la física.

5. Añadido en 1978: Muchos de los problemas discutidos en esta conferencia fueron discutidos también pormenorizadamente en Popper/Eccles *The Self and Its Brain*, 1977 [*El yo y su cerebro*], Piper, Munich, 1982).

³⁶ Véase *Conocimiento objetivo*, cap. 7.