

Davidson, Donald. (1995). «¿Puede haber una ciencia de la racionalidad?». Trad. de Alicia Nudler y Susana Romaniuk. En *La racionalidad: su poder y sus límites*. (1996). Oscar Nudler (comp.). México: Paidós. pp. 273-293.

## ¿PUEDE HABER UNA CIENCIA DE LA RACIONALIDAD?\*

*Donald Davidson*

Muchos filósofos han dudado sobre si la psicología puede ser convertida en una ciencia seria. Wittgenstein (1963, II, XIV) escribe:

La confusión y esterilidad de la psicología no se puede explicar diciendo que es una "ciencia joven"; no se puede comparar su estado, por ejemplo, con el de la física en sus comienzos [...] En efecto, en psicología existen métodos experimentales y confusión conceptual [...]

La existencia de métodos experimentales nos hace creer que disponemos de los medios para resolver los problemas que nos preocupan; cuando en realidad problema y métodos pasan de largo sin encontrarse.

Entiendo esto como referido no ya a la psicología tal como era en la época en que Wittgenstein escribió, sino como un juicio *sub specie aeternitatis*. Gilbert Ryle parece haber sido de la misma opinión. Cuando se trata de explicar el comportamiento humano, Ryle (1949, págs. 324-325) piensa que es pretensioso esperar hacerlo mejor que el sentido común:

De modo que cuando escuchamos la promesa de ofrecer una nueva explicación científica de lo que decimos y hacemos, esperamos oír o ver alguna contrapartida de aquellos impactos (como en el caso de la física), algunas fuerzas o entidades que jamás imaginamos y que no podemos tener la esperanza de observar en su mundo subterráneo. Pero si no estamos dispuestos a dejarnos impresionar, encontraremos inaceptable la promesa del descubrimiento futuro de las causas ocultas de nuestras propias acciones y reacciones. Sabemos muy bien cuál fue la causa de que el granjero regresara del mercado con los cerdos sin vender: encontró que los precios eran más bajos de lo que esperaba. Sabemos muy bien por qué fulano frunció el ceño y golpeó la puerta al salir: fue insultado.

\*El *copyright* de la versión inglesa y de otras eventuales traducciones, reimpressiones, etcétera, pertenece al autor.

Mientras Wittgenstein y Ryle desdeñan la idea de una ciencia seria que aspire a explicar el comportamiento humano, Quine es ambivalente. ¿Piensa Quine que los conceptos de significado, comunicación, interpretación, creencia, etcétera pueden llegar a articularse en una ciencia seria del comportamiento? Dada la atención que Quine ha prestado a la comprensión del lenguaje, y su punto de vista de que la filosofía es continua con la ciencia, se podría pensar que su respuesta sería afirmativa. Como mostraré en seguida, hay alguna razón para pensar que, en efecto, ésa es su respuesta. Pero también hay razones para pensar que no.

J. B. Watson, el iniciador del conductismo moderno, pensaba que conceptos como los de creencia y deseo eran “herencias de un pasado salvaje”, “concepciones medievales”, del mismo tipo que “la magia y el vudú”. B. F. Skinner, un viejo amigo de Quine, lo expresó más moderadamente: “La objeción no es que esas cosas (se refiere a conceptos tales como intención, creencia y deseo) son mentales sino que no ofrecen ninguna explicación real y obstaculizan un análisis más efectivo”. Skinner habla repetidamente de “una alternativa a las formulaciones mentalísticas”, y agrega: “No estaría metido en esto si no pensara que los modos mentalísticos de pensar acerca del comportamiento humano obstaculizan vías mucho más efectivas”.

Quine (1975, pág. 92) parece coincidir con Skinner y Watson, como lo sugiere su abierto apoyo al conductismo. “En términos generales —escribe—, las actitudes proposicionales van por mal camino. Son los modismos que más obstinadamente resisten los patrones científicos.” Buena parte del capítulo de *Palabra y objeto* titulado “Huida de la intención” está dirigido contra aquellos que piensan que podemos hablar libremente de las proposiciones y las actitudes proposicionales sin buscar una base en el comportamiento. Esto es coherente con proveer tal base, esto es, legitimar esos mismos conceptos. Pero otras afirmaciones ponen en duda esa posibilidad. Después de aceptar la afirmación de Brentano de que las expresiones intencionales (las que usamos para manifestar actitudes proposicionales) no son reducibles a conceptos no-intencionales, Quine (1960, pág. 221) señala: “Uno puede aceptar la tesis de Brentano ya sea como mostrando la indispensabilidad de las expresiones intencionales y la importancia de una ciencia autónoma de la intención, o como mostrando la falta de base de las expresiones intencionales y la vaciedad de una ciencia de la intención. Mi actitud, a diferencia de la de Brentano, es la segunda”.

Quizás esto resolvería el asunto, pero no estoy seguro. En efecto, después de todo, ¿cuál es el *status* del intento de Quine de dar una interpretación conductista de la corrección de una traducción? Quine no intenta reconstruir los conceptos de significado, analiticidad, etcétera, tal como los han pensado los filósofos. Pero lo que sí provee intenta dar cuenta no sólo de los hablantes, sino de lo que dicen. Lo intenta al decir cuándo una traducción de las palabras del hablante es aceptable sobre fundamentos conductistas. Mi pregunta sigue en pie: ¿es esta empresa meramente lo mejor que podemos hacer, sin ser siquiera el comienzo de una ciencia, o es la dirección que debemos tomar si queremos ser científicos respecto del comportamiento verbal? En particular, ¿son los sustitutos conductualmente correctos para el significado y la analiticidad (por ejemplo, el estímulo significado y el estímulo analiticidad) irreducibles a términos fisiológicos o físicos, o podrían dar lugar a ciencias más precisas, con el correr del tiempo y el incremento del conocimiento? Quine a menudo se expresa como si, en efecto, pudieran.

En *Mente y disposiciones verbales* Quine distingue tres niveles de “explicaciones pretendidas” de los fenómenos lingüísticos: el mental, el conductual y el fisiológico. El mental lo descarta como “escasamente merecedor del nombre explicación”. ¿Pero significaría esto que pasar al nivel conductual debe cambiar el tema? En absoluto: “Reconozcamos que vale la pena dedicarse al estudio semántico del lenguaje con todos los escrúpulos del científico natural. Debemos estudiar el lenguaje como un sistema de disposiciones al comportamiento verbal...”. Antes en el mismo ensayo destaca el “hecho conspicuo de que el lenguaje es una empresa social que está orientada a los objetos intersubjetivamente observables en el mundo externo”, y sugiere que esto abre la puerta para lograr “una explicación propiamente fisicalista del lenguaje” (págs. 87-91 y 84).

El primer paso, pues, desde lo mental hacia las disposiciones al comportamiento, no cambia el dominio, esto es, el análisis semántico del lenguaje; sólo lo pone en vías de ser más científico. Las disposiciones para Quine son estados físicos - estados fisiológicos cuando la disposición es lo que usualmente llamaríamos mental, como, por ejemplo, la credulidad; y físicos en el caso de las disposiciones de objetos físicos, como la solubilidad. Y aunque no cree que se sepa actualmente cómo dar una interpretación fisiológica de ninguna disposición conductual, Quine parece estar seguro de que debe haber alguna. (Como para los presentes propósitos no

tiene sentido distinguir la fisiología de un dominio especial de la física, hablaré de aquí en adelante como si la física fuera toda la ciencia natural.) Sobre este punto, Quine escribe (1975, pág. 92):

Una disposición es, según mi punto de vista, simplemente un rasgo físico, una configuración o mecanismo [...] Las disposiciones al comportamiento son, pues, estados, rasgos o mecanismos fisiológicos. Al mencionarlas disposicionalmente las estamos individualizando mediante síntomas, pruebas conductuales. Habitualmente no estamos en condiciones de detallarlas en términos fisiológicos. [Sin embargo] La explicación más profunda, la fisiológica, analizaría esas disposiciones en términos explícitos de impulsos nerviosos y otros procesos orgánicos identificados anatómicamente y químicamente.

El razonamiento parece ser el siguiente: si un objeto tiene una disposición, este hecho debe depender de sus propiedades físicas. Así, todo lo que pueda ser explicado apelando a la disposición debe ser explicable en términos físicos, ya sea que sepamos o no cómo dar la descripción física relevante. La solubilidad ilustra este punto: en una época sabíamos que había alguna propiedad física desconocida de un objeto que lo hacía soluble; ahora sabemos cuál es esa propiedad. Quine también parece sostener que una interpretación apropiada del concepto de evidencia puede ser dada en última instancia en términos físicos. En *Palabra y objeto* (1960, pág. 17) dice: "Toda teoría realista de la evidencia debe ser inseparable de la psicología del estímulo y la respuesta, aplicada a los enunciados", y en *Las raíces de la referencia* (1973, págs. 3, 36, 12 y 33 y sgtes.): "Nuestro epistemólogo liberado termina como un psicólogo empírico". El proceso de aprendizaje, piensa, es accesible a la ciencia empírica. "Explorándolo la ciencia puede, en efecto, explorar la relación evidencial." Dado que "la atribución de una disposición conductual, aprendida o no, es una hipótesis fisiológica, aunque fragmentaria", podemos concluir que "las entidades mentales son inobjetables si se las concibe como hipotéticos mecanismos físicos y se las postula con una visión estrictamente dirigida a la sistematización de fenómenos físicos".

Varias ideas emergen en estos pasajes. El tema de la irreducibilidad del vocabulario mentalista, cuando se combina con la tesis de que podría haber una explicación seria—esto es, fisiológica o física—de la relación evidencial y otros conceptos mentales, es coherente únicamente con abandonar nuestra actual forma de hablar sobre actitudes proposicionales a favor de un

vocabulario limitado al de la fisiología o la física. La pretensión de que "las disposiciones al comportamiento verbal" son configuraciones físicas sugiere que, lejos de ser irreducibles al vocabulario físico, hay una reducción sensible en el horizonte.

Quizás alguna vez Quine estuvo inseguro acerca de la relación entre los vocabularios mental y físico, pero en escritos más recientes ha adoptado el punto de vista de que el discurso acerca de creencias, deseos, acciones y significados no es reducible a algo más científico, aunque su utilidad para las descripciones y explicaciones cotidianas no puede negarse. La relación entre lo mental y lo físico que Quine ahora parece aceptar es lo que he llamado "monismo anómalo", la posición que dice que no hay correlaciones estrictamente legaliformes entre fenómenos clasificados como mentales y fenómenos clasificados como físicos, aunque las entidades mentales sean idénticas, tomadas una a una, a las entidades físicas.<sup>1</sup> En otras palabras, hay una ontología simple, pero más de un modo de describir y explicar los ítemes en esa ontología.

Hay varias razones para la irreducibilidad de lo mental a lo físico. Una razón, apreciada por Quine, es el elemento normativo en la interpretación introducido por la necesidad de apelar a la caridad al hacer corresponder los enunciados de otros con los propios. Esta correspondencia nos fuerza a sopesar las plausibilidades relativas de diferentes desviaciones respecto de la coherencia y la verdad (según nuestro propio entender). Nada en la física corresponde al modo como este rasgo de lo mental modela sus categorías.

Otra razón, quizá más fácil de entender, radica en el carácter irreduciblemente causal de los conceptos mentales. Permítaseme dar primero un ejemplo no-mental. El estado de estar tostado por el sol es necesariamente un estado causado por la acción del sol. Ninguna física completa haría uso del concepto de estar tostado por el sol, no sólo porque parte de la explicación está ya incorporada dentro de la caracterización del estado, sino también porque dos estados de la piel podrían ser idénticos en cualquier modo intrínseco, y uno podría ser un caso de estar tostado por el sol mientras que el otro no.

Lo mismo sucede con las actitudes proposicionales, la semántica de las

1. Introduce la frase y la idea en "Mental Events" (1970), reimpresso en Davidson (1980).

palabras habladas y el comportamiento como normalmente lo entendemos. La razón, tanto en el caso de las actitudes como en el de la semántica, es la misma: lo que nuestras palabras significan, y aquello acerca de lo cual pensamos, está en parte determinado por la historia de su adquisición. Las condiciones de verdad de mi enunciado "La luna es convexa", o de mi creencia de que la luna es convexa, dependen en parte de la historia causal de mis relaciones con la luna. Pero podría suceder que dos personas estuvieran en estados físicos relevantemente similares (definiéndolos sólo en términos de lo que está dentro de la piel) y sin embargo que una estuviera hablando o pensando acerca de nuestra luna, y la otra no.

Cuando se trata de explicar el comportamiento como se lo concibe normalmente, este rasgo de las actitudes proposicionales es una ventaja porque el comportamiento, concebido como acciones, es también un concepto irreduciblemente causal. Esto es porque las acciones son típicamente descritas no sólo como movimientos, sino como movimientos que pueden explicarse por las razones que tiene un agente —sus creencias y deseos—. Así, si yo pago mi cuenta firmando un cheque, necesariamente se da el caso de que firmé el cheque porque quería pagar mi cuenta y creía que firmando el cheque pagaba mi cuenta. Las acciones son individuadas siguiendo las mismas líneas que las actitudes proposicionales. Es por eso que éstas sirven bien para explicar las acciones. Pero este modo de individuar y escoger acciones no es útil para crear una ciencia del comportamiento que pueda en principio convertirse en una provincia identificable de la fisiología o la física.

Ha habido numerosos intentos de extraer de las actitudes proposicionales un contenido puramente subjetivo (o "estrecho") no sujeto a las dificultades para la ciencia introducidas por el externalismo. Si esto pudiera hacerse, eliminaría un gran obstáculo para que la psicología fuera una ciencia, dejando sólo como problema el aspecto normativo de lo mental. La razón por la que pensadores como Jerry Fodor y Noam Chomsky quieren encontrar un elemento o aspecto puramente interno de las actitudes proposicionales es obvia: sólo si las propiedades mentales sobrevienen a las propiedades físicas del agente puede haber alguna esperanza de identificar las propiedades mentales con las propiedades físicas, o de encontrar conexiones legaliformes entre ambas. Si las propiedades mentales sobrevienen no sólo sobre las propiedades físicas del agente sino también sobre las propiedades físicas del mundo exterior al agente, no puede haber esperanza de descu-

brir leyes que predigan y expliquen el comportamiento solamente sobre la base de rasgos intrínsecos de los agentes. Tanto Fodor como Chomsky han dejado claro que piensan que una variedad interna de lo intencional es esencial para hacer de la psicología un estudio serio. Por razones conexas, Fodor ha rechazado también la mayoría de las formas del holismo, al menos en lo que concierne al lenguaje. Da varias razones, pero lo que parece motivar su rechazo es la convicción de que, a menos que los significados de las expresiones puedan ser ligados de modos legaliformes a configuraciones neurales específicas, no hay esperanza para una explicación seria de los fenómenos lingüísticos. Tales lazos, por supuesto, excluirían el externalismo.

Lo que creo cierto es que el holismo, el externalismo y el rasgo normativo de lo mental se mantienen en pie o se derrumban juntos: si éstos son rasgos de lo mental y obstaculizan una ciencia seria de la psicología, entonces Ryle, Wittgenstein y Quine, en su ánimo más pesimista, tendrían razón. No podría haber una ciencia seria, o ciencias serias, de lo mental. Creo que los elementos normativos, holísticos y externalistas que se encuentran en los conceptos psicológicos no pueden eliminarse sin cambiar radicalmente el objeto de estudio. No deseo discutir estos puntos en este trabajo, habiéndolo hecho en extenso en otro lugar.<sup>2</sup> Mi interés aquí es más bien preguntar qué se sigue si estoy en lo cierto. Claramente, no se sigue sin más argumento que no pueda haber una psicología científica. Que se infiera esto depende de lo que se entienda por "ciencia", y de si los rasgos que, según sostengo, caracterizan a lo mental son un obstáculo para ella. Lo que sí se sigue es que la psicología no puede ser *reducida* a la física, ni a ninguna otra de las ciencias naturales. Pero, a menos que simplemente legislemos que sea ciencia lo que puede reducirse a una ciencia natural, el fracaso en la reducción no podría ser usado en sí mismo para demostrar que lo que no puede ser así reducido no merece ser llamado ciencia.

Dado que mi propio enfoque de la descripción, análisis (en un sentido tosco) y explicación del pensamiento, el lenguaje y la acción tiene, por un lado, lo que considero que son algunas de las características de una ciencia,

2. Véanse, por ejemplo, "Mental Events" y "Three Varieties of Knowledge" en A. Phillips Griffiths (1991).

y que, por otro lado, ha sido atacado tanto por Fodor como por Chomsky como radicalmente "acientífico", me propongo examinar mi teoría, si ésta es la palabra, para ver si puede ser defendida como ciencia y cómo. Debo señalar de entrada que pienso que el resultado no es neto.

Una manera de pensar acerca del momento en que la psicología se constituyó en una ciencia empírica es ligarlo con el trabajo de Gustav Theodor Fechner, cuya vida se extendió a lo largo de casi todo el siglo XIX (1801-1887). Fechner empezó como físico, pero luego fue derivando, a través de la química, la fisiología y la medicina, hacia la metafísica (y más allá, hacia el misticismo). Fechner estaba interesado en la relación entre mente y cuerpo, o materia y espíritu, y se acercó a este problema buscando leyes cuantitativas que conectaran lo mental y lo físico. Weber ya había sugerido que el cambio más pequeño en la intensidad de una magnitud física requerido para producir una diferencia perceptible en la sensación no es una diferencia física fija, sino que es proporcional a la magnitud del estímulo. Fechner generalizó la ley: la intensidad experimentada de un estímulo físico es igual a una constante multiplicada por el logaritmo del estímulo físico. Dicho de modo más llano, a medida que un estímulo físico aumenta (por ejemplo, la intensidad o el tono de un sonido), incrementos iguales en la magnitud del estímulo físico resultan en incrementos más y más pequeños en la sensación experimentada. La constante varía según el sentido involucrado. Esta ley puede, por supuesto, ser sometida a prueba, y es aproximadamente correcta. La escala de decibeles de sonido es un ejemplo informal: intervalos iguales en la escala de decibeles son (más o menos) iguales subjetivamente, pero el cociente entre dos cantidades de poder acústico es igual a 10 veces el logaritmo común de ese cociente.

Fechner tuvo una idea acertada. Si los métodos científicos han de aplicarse a lo mental, lo serán proponiendo una teoría sólida y preguntando cómo puede ser puesta a prueba e interpretada empíricamente. Las teorías describen estructuras abstractas; sus interpretaciones empíricas preguntan si esas estructuras pueden ser descubiertas en el mundo real. La teoría de Fechner es relativamente fácil de interpretar en algunos casos, lo que quizá no es sorprendente, dada la base neurológica de la discriminación sensorial. Lo que ahora sabemos acerca de las neuronas, las redes neuronales y el —así llamado— procesamiento de la información que tiene lugar en los órganos de los sentidos y el cerebro, sugiere que podríamos esperar encon-

trar leyes cuantitativas que relacionen la discriminación sensorial y las magnitudes físicas de los estímulos. Pero hay escalas estrechamente relacionadas de sensaciones percibidas que son definitivamente sorprendentes, al menos para mí. Un buen ejemplo es la percepción de las relaciones entre intervalos en el tono de los sonidos. Los griegos sabían que si uno divide una cuerda vibrante en dos mitades, cada mitad suena una octava más alto que la cuerda completa, y dos tercios de la cuerda producen una quinta más alta que la cuerda completa. (Se cree que Pitágoras fue quien descubrió esto.) Pero lo que es sorprendente es que si uno hace sonar dos notas apartadas entre sí por alguna distancia arbitraria y le pide a una persona que haga sonar una tercera nota en el punto medio percibido, no sólo diferentes oyentes llegan aproximadamente al mismo tono, sino que los tonos así determinados están relacionados de tal manera que producen una escala de intervalos; esto es, que pueden asignarse números a los diferentes tonos de manera de conservar las relaciones entre los intervalos, no sobre la base de una magnitud física, como el largo de la cuerda o las vibraciones por segundo, sino enteramente sobre la base de lo que es subjetivamente percibido. La teoría que describe este hecho tiene todo el derecho a ser llamada teoría psicológica, porque no trata sino de relaciones entre fenómenos psicológicos.

De algún modo, ya he dado buenos ejemplos de teoría científica en el campo de la psicología: uno, en la forma de una ley general que relaciona la intensidad percibida de estímulos sensoriales con aspectos físicamente medibles de los estímulos; el otro, en la forma de la medición fundamental de intervalos de tonos percibidos. Pero, por supuesto, estos ejemplos no nos dicen nada sobre aquello que preocupa a quienes se preguntan si, o de qué manera, la psicología puede ser científica. Lo que a ellos les interesa es la descripción, predicción y explicación de acciones intencionales, y las actitudes asociadas tales como intención, creencia, deseo y significado lingüístico. Aquí consideraré una teoría particular que he propuesto. La describiré en líneas generales y luego examinaré en qué aspectos tiene los rasgos de una teoría científica.

La teoría que tengo en mente relaciona los conceptos de creencia, deseo y significado lingüístico. Dado que la teoría toma la creencia en una forma cuantificada —a veces llamada probabilidad subjetiva— y el deseo como medido en una escala de intervalo, como la temperatura Fahrenheit o la escala de tono subjetiva que he mencionado antes, inclu-

ye una versión de la, a veces, llamada teoría de la decisión. Es, pues, adecuada para la explicación de intenciones y acciones intencionales. A diferencia de la teoría tradicional de la decisión en la forma explícita dada primeramente por Frank Ramsey, o en la versión algo diferente inventada por Richard Jeffrey,<sup>3</sup> la teoría que tengo en mente incluye integralmente una teoría del significado. Puede por lo tanto denominarse una teoría unificada del lenguaje y la acción, o, dicho brevemente, La Teoría Unificada.

La Teoría Unificada describe o define una estructura abstracta. Esta estructura tiene ciertas propiedades interesantes y deseables que es posible probar. Así, uno puede probar, con respecto a la parte que toma prestada de la teoría de la decisión, tanto un teorema de la representación como un teorema de la unicidad. El primero dice, en efecto, que pueden asignarse números a las creencias y los deseos que preserven las restricciones cualitativas impuestas por la teoría; el segundo dice que los números asignados para medir probabilidades constituyen una escala de razones y los números que miden deseos constituyen una escala de intervalos.<sup>4</sup> Esto es adecuado para permitir —al menos “en teoría”— predicciones de acciones intencionales. La parte de la teoría que trata del significado lingüístico es, en efecto, una modificación de una teoría de la verdad tipo Tarski, y por lo tanto es demostrablemente capaz de proveer las condiciones de verdad para todas las emisiones de enunciados en un lenguaje del cual trata. La parte final de la teoría une la teoría de la decisión con la teoría de la verdad a través de un aparato formal que no intentaré describir aquí.<sup>5</sup> La posibilidad de ensamblar las dos teorías depende de dos cosas. La primera es que la teoría de la decisión muestre cómo extraer de las simples preferencias tanto utilidades cardinales como probabilidades subjetivas. La segunda es que las probabilidades subjetivas, cuando son tomadas como aplicadas a enunciados, sean suficientes para permitir una teoría del significado. Hay, así,

3. Sigo la versión de Jeffrey (1983). La teoría original de F. M. Ramsey ha sido reimpressa en D. H. Mellor (1990).

4. Esas observaciones acerca de las escalas relevantes se aplican estrictamente a la teoría de Ramsey; la teoría de Jeffrey es marginalmente diferente.

5. Para algunos detalles, véase Davidson (1990).

una ruta, técnicamente bastante bizantina, pero intuitivamente clara en cada uno de sus pasos, que va desde las elecciones simples hasta la interpretación detallada de palabras, deseos y creencias.

La posibilidad de tal teoría descansa en estructuras dictadas por nuestro concepto de racionalidad. Tanto la teoría de la decisión tal como yo la he usado, en la versión desarrollada por Richard Jeffrey, como las teorías de la verdad, por ejemplo, dependen en parte de la lógica. La teoría de la decisión de Jeffrey y las definiciones de verdad de Tarski dan por supuesta una lógica subyacente: estas teorías serían verdaderas sólo en el caso de lógicos perfectos. Más allá de ello, existe el supuesto de una distribución racional de probabilidades entre las proposiciones, y de una proporcionalidad de grados de creencia de acuerdo con las probabilidades condicionales. En otras palabras, se sostiene la verdad de las proposiciones en la medida en que su soporte evidencial lo hace racional. Así, toda la estructura de la teoría depende de los estándares y las normas de racionalidad.

Estas consideraciones siembran considerables dudas sobre las pretensiones científicas de la Teoría Unificada. Pero antes de ocuparme de las dudas, permítaseme extenderme un poco más acerca de la estructura general. Como toda teoría científica, la Teoría Unificada presenta una clara y precisa estructura formal con méritos demostrables. Hay sólo unos pocos conceptos indefinidos, y éstos son extensionales. El concepto primitivo básico es la relación triádica entre un agente y dos enunciados, que se entabla cuando el agente preferiría débilmente que un enunciado sea verdadero en lugar de otro. Esta relación es extensional en el sentido técnico de que un enunciado que afirme que se da una relación entre tres objetos apropiados (un agente y dos de los enunciados de dicho agente) conserva su valor de verdad (verdadero o falso) independientemente de cómo se describan esos tres objetos. Pero si el patrón observado de esas relaciones se ajusta a los términos de la teoría, es posible inferir los grados de creencia que el agente asigna a sus enunciados, cuánto quisiera el agente que esos enunciados fueran verdaderos y cuáles son las condiciones de verdad (esto es, los significados) de esos enunciados. En otras palabras, la teoría, si fuera verdadera de un agente, serviría para interpretar las creencias, los valores y las palabras de ese agente.

Esta aserción, aun protegida como está por la cláusula “si fuera verdadera”, requiere mucha defensa. Una cuestión es, por ejemplo, si la creencia, la evaluación y el significado son suficientes para dar apoyo a esa

interpretación de base amplia sin añadir, digamos, intención o percepción como variables ulteriormente relacionadas pero independientes, para no mencionar las emociones. Es también dudoso si una teoría de la verdad es adecuada para la interpretación del habla, aun cuando asumiéramos que una teoría de la verdad pudiera cubrir todas las expresiones idiomáticas de un lenguaje natural. Pero, aunque estos asuntos son importantes, me propongo dejarlos de lado por ahora para poder seguir adelante con el problema de si una teoría como la Teoría Unificada puede ser considerada científica. Mi conclusión hasta el momento es que, desde un punto de vista puramente formal, es una teoría poderosa, y, en la medida en que se corresponde con muchas de nuestras intuiciones concernientes a la naturaleza de la racionalidad, es una teoría atractiva.

Es cuando atendemos a la interpretación empírica de la teoría que surgen las cuestiones y los problemas básicos. Aquí quisiera distinguir entre la historia oficial acerca de cómo puede ser interpretada la teoría, y una versión no oficial. Oficialmente, es esencial poder mostrar cómo la teoría puede ser interpretada sin apelar a una evidencia que suponga la individuación de los contenidos de ninguna actitud proposicional. Tal forma de evidencia la constituyen, como se ha mencionado, los protocolos que especifican la preferencia de un agente de que sea verdadero un enunciado en lugar de otro. Dada suficiente evidencia de ese tipo, se puede construir una imagen de las creencias, los deseos y los significados del agente (esto es, las condiciones de verdad de sus emisiones). Una cantidad finita de tal evidencia sólo puede, por supuesto, confirmar la teoría, no verificarla. Eso es lo que podemos esperar. En breve bosquejo, la historia oficial toma esta ruta:

La versión de Jeffrey de la teoría de la decisión, aplicada a enunciados, nos dice que un agente racional no puede preferir un enunciado y su negación, a una tautología; ni una tautología, a un enunciado y su negación. Este hecho vuelve posible para un intérprete identificar, sin tener conocimiento de los significados de los enunciados del agente, todas las conectivas puramente proposicionales, tales como la negación, la conjunción y el bicondicional. Este conocimiento mínimo basta para determinar las probabilidades subjetivas de todos los enunciados del agente —cuán probable es, según el agente, que esos enunciados sean verdaderos— y luego para fijar los valores de verdad relativos de esos enunciados (desde el punto de vista del agente, por supuesto). Las probabilidades subjetivas pueden entonces usarse para interpretar los enunciados. En el caso de lo que Quine llama

enunciados de observación, los cambios en las probabilidades proporcionan las claves obvias para la interpretación de primer orden, cuando está dirigida a eventos y objetos fácilmente percibidos de forma simultánea por el intérprete y la persona que está siendo interpretada. Las probabilidades e implicaciones condicionales entre enunciados, al registrar lo que el hablante toma como evidencias para sus creencias, proveen al intérprete de lo que se necesita para interpretar términos y enunciados más teóricos. Ésta es la historia oficial. Su mérito radica no en su plausibilidad como explicación de cómo procedemos realmente para entender a los otros, sino en el hecho de que equivale a una prueba informal de lo adecuado de la teoría en cuanto a ofrecer el sustento necesario para la interpretación de las actitudes proposicionales básicas. (Esto debería compararse con la historia oficial de cómo la teoría de la decisión de Ramsey proporciona resultados suficientemente únicos para explicar el comportamiento de elección sobre la base de preferencias simples.)

Extraoficialmente, se puede admitir que como intérpretes que vivimos y trabajamos nunca tenemos suficiente evidencia del tipo requerido para seguir la ruta oficial, y siempre tenemos gran cantidad de otros tipos de evidencia. Hacemos suposiciones interminables acerca de la gente con que nos encontramos, acerca de lo que quieren, lo que están queriendo decir con lo que dicen, lo que creen acerca del entorno que compartimos con ellos, y por qué actúan como actúan. Nuestras habilidades como intérpretes entran en juego principalmente cuando una u otra de estas suposiciones resulta ser falsa, y para entonces tenemos mucho más que la pobre evidencia de la que depende la Teoría Unificada. Pero así es como debe ser. El objetivo de la teoría no era describir cómo realmente interpretamos, sino especular sobre qué hay en el pensamiento y el lenguaje que los hace interpretables. Si podemos contar una historia como la historia oficial acerca de cómo eso es posible, podemos concluir que las restricciones que la teoría plantea a las actitudes pueden articular algunos de sus rasgos filosóficamente significativos.

He descrito en su forma más transparente el arte de aplicar la teoría formal a un individuo real, considerando al intérprete y al hablante equipados con un conjunto maduro de conceptos y las aptitudes lingüísticas para expresarlos. Lo que falta al comienzo es un lenguaje compartido, y un conocimiento previo de las actitudes del otro. Puesto que la teoría y la historia oficial de cómo puede aplicarse están ya alejadas de la práctica

real, debemos suponer que la teoría arrojará sólo una luz muy oblicua sobre la adquisición de un primer lenguaje, y más aún sobre los orígenes del habla. Lo máximo que puede decirse es que si estamos de acuerdo en que el patrón de actitudes es como la teoría lo describe, se puede considerar tal vez que una criatura adecuadamente dotada por la naturaleza podría adquirirlo en compañía de otros que ya poseyeran lenguaje y pensamiento. La teoría puede también dar lugar a una hipótesis interesante acerca de los orígenes del lenguaje. Mencionaré esto al final.

Debo enfatizar que tratar de dar una aplicación empírica de la teoría formal pertenece en gran medida al dominio de la interpretación. La aplicación pretendida es a individuos, hablando puntualmente en un momento dado, puesto que podemos esperar que muchos de los valores y las creencias de cualquier persona cambien rápidamente a medida que cambia el mundo. Los ingredientes aparentemente cuantitativos, las medidas de grados de deseos y grados de creencias, no pertenecen a la teoría misma; como cualquier teoría de medición fundamental, los números simplemente hacen uso de la teoría sin ser parte de ella. Podríamos, si quisiéramos, usar la teoría simplemente como un recurso para registrar las relaciones de las actitudes entre sí, y las relaciones entre las actitudes y el mundo, su semántica. Pero en el caso de las creencias y las actitudes evaluativas, es conveniente representar estas relaciones entre los números, como los teoremas de representación de las teorías de la decisión prueban que es posible hacer.

Emerge aquí un rasgo especial de la Teoría Unificada que bien puede despertar sospechas: ¿qué es lo que desempeña el papel de los números cuando se trata de asignar *contenidos* a las palabras y actitudes de un agente? Lo que se necesita es una provisión potencialmente infinita de entidades con un patrón o estructura suficientemente complejos como para suministrar un modelo para las actitudes. Dada tal provisión, podemos entonces hacer un registro de los roles de las actitudes y las condiciones de verdad de los enunciados. Todo aquel que tiene un lenguaje dispone de tal conjunto de entidades, a saber, el conjunto (infinito) de sus propios enunciados; y éstos son todo lo que tenemos para interpretar a los demás. Es obvio que empleamos nuestros propios enunciados siempre que atribuimos un particular deseo o creencia, intención o significado a otra persona. Esto no quiere decir que mis enunciados sean los *objetos* de sus actitudes; yo solamente uso mis enunciados para rastrear lo que usted piensa y quiere decir, o para

decirme o decirle a otro, lo que usted piensa y quiere decir. Las actitudes no tienen objetos en ningún sentido psicológico o epistémico. Las actitudes son simplemente estados, y no requieren objetos ante la mente más que lo que las varas requieren números para tener cierta longitud.

Ahora bien, retornando a la cuestión inicial, nos preguntamos hasta qué punto o de qué modos es científica una teoría como la Teoría Unificada. Pienso que es claro que esa teoría no es reducible a una ciencia como la física o la neurobiología; sus conceptos básicos no pueden definirse en el vocabulario de ninguna ciencia física, y no hay reglas precisas que correlacionen, firme y confiablemente, hechos o estados descritos en el vocabulario psicológico, con hechos y estados descritos en el vocabulario de una ciencia física. Pero definir ciencia como aquello que puede reducirse a la física no sería interesante. ¿Hay otras dificultades? Se ha pensado que hay tres rasgos de la Teoría Unificada (y otras teorías similares) que la excluyen del dominio de la ciencia seria. Ellos son: sus supuestos de holismo y de externalismo, y sus propiedades normativas.

Toda la teoría está construida sobre las normas de la racionalidad; son estas normas las que han sugerido la teoría y le han dado la estructura que posee. Éstas están en gran medida incorporadas en las partes formales, axiomatizables, de la teoría de la decisión y de la teoría de la verdad, y son tan precisas y claras como cualquier teoría formal de la física. Pero las normas o consideraciones de racionalidad también entran en la aplicación de la teoría a agentes reales, en la etapa en que un intérprete asigna sus propios enunciados a la captación de los contenidos de los pensamientos y emisiones de algún otro. El proceso necesariamente implica decidir qué patrón de asignaciones hace al otro inteligible (no *inteligente*, por supuesto), y aquí se trata de usar los propios estándares de racionalidad para calibrar los pensamientos del otro. De algún modo, es como hacer corresponder una curva a un conjunto de puntos, como se hace en las mejores ciencias. Pero hay un elemento adicional en el caso psicológico: en física hay una mente que trabaja tratando de darle el mayor sentido posible a un objeto de estudio que es tratado como carente de cerebro; en el caso psicológico, hay un cerebro en cada extremo. Las normas son usadas como estándares de normas.

La Teoría Unificada es holística de punta a punta. Está diseñada para asignar contenidos a creencias, emisiones y valores simultáneamente, porque esas actitudes básicas son tan interdependientes que no sería posible



determinarlas de a una por vez, ni siquiera de a dos. Su tratamiento de cada uno de esos dominios también es holístico: los enunciados son interpretados en términos de sus relaciones con otros enunciados, las creencias en términos de sus relaciones con otras creencias, y así sucesivamente. Este holismo es característico de todo esquema de medición: los ítemes deben su medida a sus relaciones con otros ítemes. Un significado no podría asignarse a un único enunciado aislado más que lo que puede asignarse un peso o una posición a un único objeto aislado. El holismo de lo mental no puede, entonces, considerarse en sí mismo como un obstáculo para las pretensiones de científicidad de una teoría de lo mental. Todo lo contrario: la posibilidad de la teoría descansa en el holismo.

Las condiciones de verdad de las emisiones de un hablante determinan, y por lo tanto dependen, en parte, de las relaciones lógicas del enunciado emitido con otros enunciados. En el caso de enunciados de observación, las condiciones de verdad pueden depender también de la historia causal de las situaciones en que el enunciado fue aprendido y usado; ésta es una de las formas que toma el externalismo. Puesto que un externalismo perceptual de este tipo introduce un elemento irreduciblemente causal en la interpretación de la teoría; ésta no puede esperar emular a la física, que se ha esforzado con éxito por eliminar todos los conceptos causales de sus leyes. El externalismo pone límites a cuán completa puede ser una explicación psicológica, desde el momento en que introduce en el corazón del asunto elementos que ninguna teoría psicológica puede pretender explicar. Por otro lado, este rasgo por sí mismo hace a la teoría psicológica no menos científica que la vulcanología, la biología, la meteorología o la teoría de la evolución.

Tanto Fodor como Chomsky han criticado la Teoría Unificada y el método propuesto para su interpretación, que he llamado interpretación *radical* (radical porque no presupone conocimiento previo de las actitudes proposicionales del agente). Algunas de sus críticas parecen no dar en el blanco.<sup>6</sup> Tanto Fodor como Chomsky señalan que la interpretación radical da una explicación com-

6. La discusión siguiente abrevia un tratamiento más detallado de la crítica a Fodor en mi "Interpreting Radical Interpretation", en J. E. Tomberlin (1994), págs. 121-128. También impreso en R. Stoecker ed. (1993). Ambas fuentes incluyen el artículo de Jerry Fodor y Ernest LePore, "Is Radical Interpretation Possible?", al que estoy respondiendo. Los pasajes citados de Chomsky provienen de J. Earman, ed. (1991), págs. 108-9.

pletamente equivocada no sólo de cómo la lingüística estudia nuevas lenguas sino también de cómo los niños adquieren una primera lengua. En esto han sido comprensiblemente engañados por la vieja tendencia de los filósofos a discutir la cuestión teórica de cómo un lingüista o un niño podrían aprender una lengua desconocida o primera como si fuera una cuestión práctica acerca de cómo lo hacen realmente. He explicado a menudo que la interpretación radical no intenta ofrecer indicaciones útiles a los lingüistas reales, ni criticar sus métodos. Mucho menos pretende iluminar la —al menos para mí— misteriosa cuestión de la adquisición de una primera lengua.

Fodor y Chomsky critican el hecho de que la interpretación radical hace uso de mucha menos información de la que dispone el lingüista informado y metodológicamente sofisticado. Su irritación es alimentada por su convicción de que yo sostengo que la evidencia sobre la cual afirmo que una interpretación radical puede basarse es toda la evidencia que está legítimamente disponible. Chomsky en particular cree que yo ignoro sus descubrimientos acerca de cuánto de la sintaxis de los lenguajes naturales parece estar genéticamente programada. He argumentado, como ya mencioné, que es una condición para que una teoría del significado sea correcta ser de tal manera que si un intérprete supiera que ella es verdadera de un hablante, el intérprete podría entender lo que el hablante dijo. Por supuesto, negué que los intérpretes generalmente tengan, o al menos sepan que tienen, una teoría semejante; la teoría es, más bien, lo que el filósofo quisiera si se dispusiera a describir ciertos aspectos de las habilidades interpretativas del intérprete. Y entonces agregué en un ensayo que provocó particularmente a Chomsky, "No agrega nada a esta tesis decir que si una teoría describe correctamente la competencia de un intérprete, algún mecanismo en el intérprete debe corresponder a la teoría" (cf. Chomsky, "A Nice Derangement of Epitaphs" en R. Grandy y R. Warner, eds., 1986; reimpresso en LePore, ed., 1986). Chomsky cita esta observación, y comenta que "desde el punto de vista de las ciencias naturales, [este] comentario es completamente desatinado". Su discusión consiguiente deja en claro que lo que le disgusta es que piensa que estoy negando que saber cuál es ese mecanismo puede ser de algún interés. Pero por supuesto éste no es mi punto de vista, ni es lo que dije. Lo que dije fue, y tuvo la intención de ser, una tautología: si una píldora nos hace dormir, no agrega nada decir que la píldora tuvo poder dormitivo. Sería ampliamente interesante saber más acerca de la naturaleza de nuestras habilidades lingüísticas y los mecanismos subyacentes a ellas. ¿Quién lo negaría? Si yo tuviera dudas,

ellas sólo se relacionarían con las conclusiones filosóficas que Chomsky y algunos de sus seguidores han extraído de sus resultados en esta área.

Chomsky me ha acusado a mí, y especialmente a Quine, de suponer que todo lo que sabemos acerca del lenguaje debe estar basado en evidencia comportamental. Quine ha hablado por sí mismo en este asunto; yo ciertamente rechazaría la acusación. Si queremos saber todo acerca del lenguaje, su adquisición y sus usos, no hay límites a priori acerca de qué evidencia podría ser relevante. Pero comparto con Quine la convicción de que nuestra comprensión de lo que los hablantes quieren decir con lo que dicen está en parte basada, directa o indirectamente, en lo que podemos aprender o captar a partir de lo que percibimos que hacen. Por mucha gramática con que vengamos equipados desde la cuna, debemos aprender qué significan las palabras de cualquier lenguaje particular —no nacemos sabiendo inglés, o hebreo, o mandarín—; debemos tomar nuestra lengua materna de quienes ya la hablan. (El conductismo del que hablo no es, dicho sea de paso, reduccionista por naturaleza: no espero que ningún predicado intencional básico pueda definirse usando términos no-intencionales. La cuestión aquí concierne simplemente a la evidencia.)

Las críticas que Fodor y Chomsky han hecho a ciertos filósofos me parece que se basan, en gran medida (aunque no totalmente), en que han leído en esos filósofos puntos de vista que éstos no sostienen. He tratado de señalar algunos ejemplos. Pero hay también una falla en apreciar una diferencia en objetivos e intereses fundamentales. Chomsky aparentemente me ve como tratando de entender y explicar los mismos fenómenos que él, y por lo tanto como proponiendo hipótesis rivales a las suyas. Esto parece completamente incorrecto. Yo quiero saber qué es lo que hay en el pensamiento proposicional, nuestras creencias, deseos, intenciones y habla, que los hace inteligibles a otros. Ésta es una pregunta acerca de la naturaleza del pensamiento y el significado que no puede responderse descubriendo mecanismos neurales, estudiando la evolución del cerebro o encontrando evidencias que expliquen la increíble facilidad y rapidez con la que adquirimos nuestro primer lenguaje. Incluso si todos nacióéramos hablando inglés o polaco, podríamos preguntarnos cómo es que entendemos a los demás y qué es lo que determina los contenidos cognitivos de nuestros enunciados. Poco importa si llamamos científicos a algunos de esos proyectos y les negamos ese término a otros.

Lo que sí importa, sin embargo, es qué limitaciones tiene el estudio de las actitudes que he estado discutiendo, del mismo modo como, en

otro contexto, importan las limitaciones del trabajo de Chomsky o de Fodor —aunque las limitaciones son, por supuesto, diferentes—. ¿Cuáles son las deficiencias más obvias de la Teoría Unificada del lenguaje y el habla? Lo primero y quizá más notorio es el hecho de que la teoría formal (como opuesta a los rasgos de su aplicación empírica) no dice nada en absoluto acerca de inconsistencias. No sólo postula una lógica perfecta y un patrón consistente y racional de creencias y deseos, sino que también presupone racionalidad en el tratamiento de lo que tomamos por evidencia. Inconsistencias y defectos en el poder de razonamiento deben ser acomodados a través de la inyección de grandes dosis de lo que ha sido llamado caridad en el ajuste de la teoría a agentes reales.

Tal vez toda irracionalidad obvia aparece como inconsistencia, pero claramente no toda inconsistencia es lo que normalmente llamamos irracionalidad. La teoría formal no deja lugar para la irracionalidad, y por lo tanto no tiene poder para explicarla. Todas las explicaciones de la irracionalidad que nos preocupamos por ofrecer deben trabajar en contra de la Teoría Unificada, no con ella.

La Teoría Unificada, tal como la he descrito, es estática; no dice nada acerca de las formas de racionalidad relacionadas con la incorporación de nueva información a un sistema de pensamiento ya existente. Sin embargo, ésa es un área en la que hay esperanza. Se ha hecho mucho trabajo, por ejemplo los de Richard Jeffrey e Isaac Levi, para hacer dinámica la teoría de la decisión.

Por último, y quizá más significativamente, la interpretación de la teoría formal no descansa enteramente sobre la evidencia intersubjetiva ordinaria. Al medir magnitudes físicas, podemos usar los números para hacer un registro de las propiedades de eventos y objetos públicamente observados. Las propiedades relevantes de los números también pueden ser acordadas por todos aquellos a quienes les concierna. Sin embargo, las cosas son diferentes cuando una mente trata de entender a otra. La gente es tan públicamente observable como cualquiera otra cosa en la naturaleza, pero las entidades que usamos para construir una imagen de los pensamientos de otra persona deben ser nuestros propios enunciados, como nosotros los entendemos, u otras entidades de la misma procedencia y estructura. Los significados de nuestros enunciados son ciertamente dependientes de nuestras relaciones con el mundo —acerca del cual tratan esos enunciados— y de nuestras interacciones lingüísticas con otros. Pero no hay escapatoria del

hecho de que no podemos comprobar las credenciales objetivas de la medida que usamos del mismo modo como podemos comprobar nuestro entendimiento de los números. No podemos comprobar la corrección objetiva de nuestras propias normas por comparación con otros, porque hacer eso sería hacer otra vez un uso básico de nuestras propias normas.

Si los rasgos de una teoría psicológica como la que he estado ensayando, especialmente el último, muestran que una teoría psicológica es tan diferente de una teoría de las ciencias naturales que no merece ser llamada ciencia es algo que no sé, ni me preocupa. De lo que estoy seguro es que tal teoría, aunque pueda ser tan genuina como cualquiera otra, no está compitiendo con ninguna ciencia natural.

TRADUCTORAS: Alicia Nudler y Susana Romaniuk.

#### REFERENCIAS BIBLIOGRÁFICAS

- Davidson, D.: (1980) *Actions and Events*, Oxford, Oxford University Press.
- Davidson, D.: (1990) "The Structure and Content of Truth", *The Journal of Philosophy*, vol. 10, págs. 279-328.
- Earman, J. (comp.): (1991) *Inference, Explanation and Other Frustrations: Essays in the Philosophy of Science*, Berkeley/Los Ángeles, The University of California Press.
- Grandy, R. y Warner, R. (comps.): (1986) *Philosophical Grounds of Rationality*, Oxford, Oxford University Press.
- Jeffrey, R.: (1983) *The Logic of Decision*, 2a. ed., Chicago, University of Chicago Press.
- LePore, E. (comp.): (1986) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford, Blackwell's.
- Mellor, D. H. (comp.): (1990) *Ramsey Philosophical Papers*, Cambridge, Cambridge University Press.
- Phillips Griffiths, A. (comp.): (1991) *A. J. Ayer: Memorial Essays. Royal Institute of Philosophy Supplement: 30*, Cambridge, Cambridge University Press.
- Quine, W.V.: (1960) *Word and Object*, Cambridge, Mass., MIT Press.
- Quine, W.V.: (1973) *The Roots of Reference*, La Salle, Illinois, Open Court.

- Quine, W.V.: (1975) "Mind and Verbal Dispositions", en S. Guttenplan (comp.), *Mind and Language*, Oxford, Oxford University Press.
- Ryle, G. (1949), *The Concept of Mind*, Nueva York, Barnes y Noble.
- Stoecker, R. (comp.): (1993) *Reflecting Davidson: Donald Davidson Responding to an International Forum of Philosophers*, Berlín, de Gruyter.
- Tomberlin, J. E. (comp.): (1994) *Philosophical Perspectives*, vol. 8, *Logic and Language*, Atascadero, Calif., Ridgeview Press.
- Wittgenstein, L.: (1963) *Philosophical Investigations*, Oxford, Blackwell.