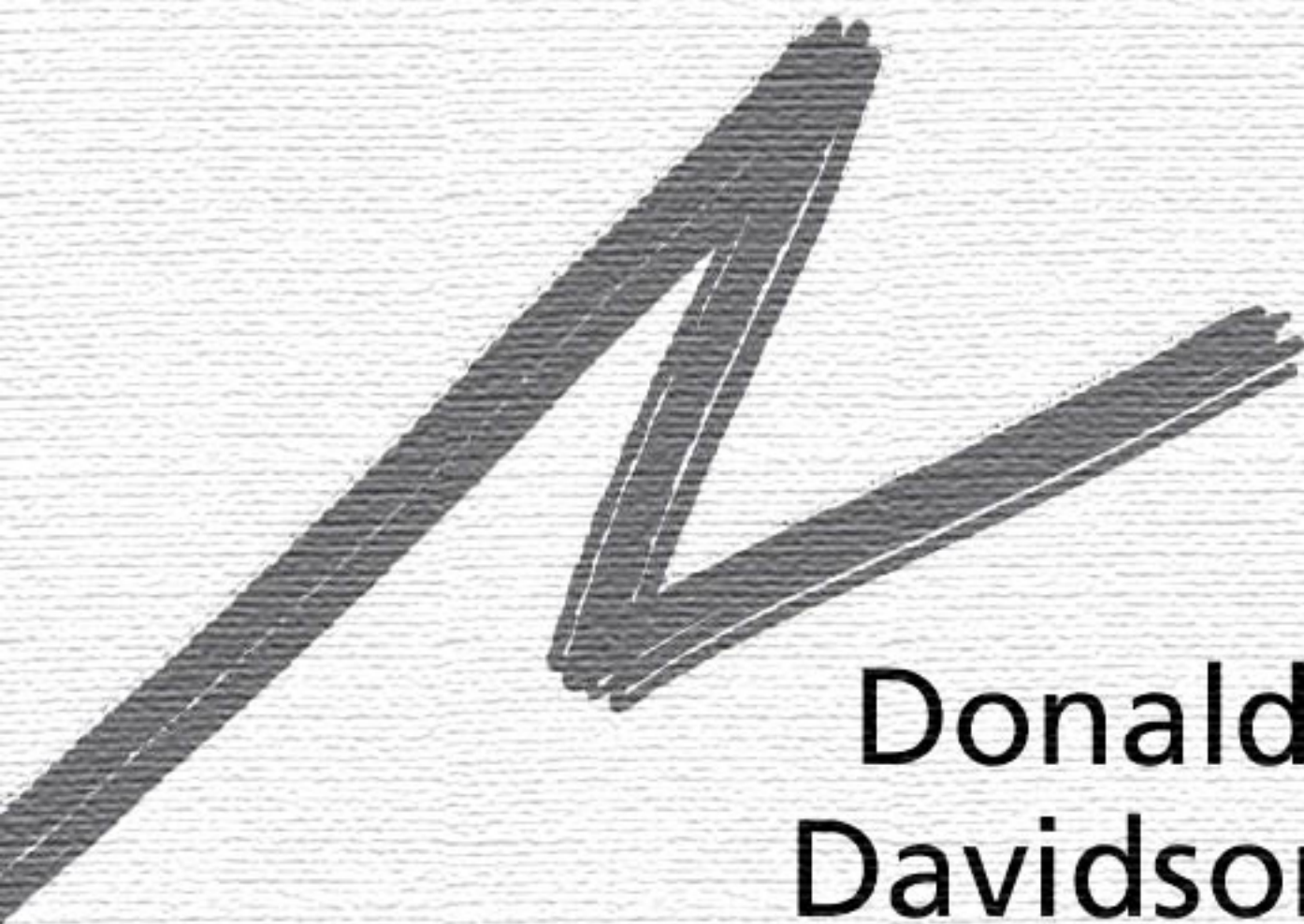


# Estructura y contenido de la verdad



Donald  
Davidson

# Estructura y contenido de la verdad

Donald Davidson

Traducido por M. J. Frápoli  
En Juan Antonio Nicolás y María José Frápoli (ed.), *Teorías  
de la verdad en el siglo XX*. Tecnos, Madrid, 1997

Edición original:  
“The structure and content of truth”,  
*Journal of Philosophy*, 87/6 (1990), pp. 279–328

Los números entre corchetes corresponden  
a la paginación de la edición impresa

letra e

## Estructura y contenido de la verdad<sup>1</sup>

Nada en el mundo, ni objeto ni evento, sería verdadero o falso si no hubiera criaturas pensantes. John Dewey, en cuyo honor y memoria se dieron las conferencias que constituyen este ensayo, sacó dos conclusiones: que el acceso a la verdad no puede ser una prerrogativa especial de la filosofía, y que la verdad debe tener conexiones esenciales con los intereses humanos. Era despreciativo con la tradición filosófica que consideraba la verdad como correspondencia entre el pensamiento y una realidad inaccesible a la investigación experimental y a la práctica ordinaria. Creía que esta imagen de la verdad fue designada para servir a la tesis de que los filósofos poseen una técnica privilegiada para lograr una forma de conocimiento diferente de, y superior a, la ciencia. Dewey<sup>2</sup> escribió que

---

<sup>1</sup> Presentado como tres conferencias sobre «El Concepto de Verdad», dadas en la Universidad de Columbia en noviembre 1989; la primera, «La Estructura de la Verdad», el 9 de noviembre; la segunda «Verdad y Conocimiento», el 16 de noviembre; y la tercera, «Los Contenidos de la Verdad», el 20 de noviembre. Estas conferencias, que fueron posibles gracias a la Fundación John Dewey, constituyen la serie sexta de las Conferencias John Dewey, que se establecieron en 1967 en honor al difunto John Dewey, que fue profesor de Filosofía en Columbia desde 1905 a 1930. Estoy agradecido a Akeel Bilgrami, Ernest LePore, Isaac Levi, y W. V. Quine por sus provechosas sugerencias y por su amistoso estímulo.

<sup>2</sup> *Experience and Nature*, Nueva York, Dover, 1958.

[...] la profusión de testimonios de la devoción suprema a la verdad por parte de la filosofía es una cuestión que despierta sospecha. Porque usualmente ha sido un preliminar a la afirmación de que hay un órgano peculiar de acceso a la verdad más alta y última. No hay nada de esto... La verdad es una colección de verdades; y estas verdades constituyentes están en ma-[147]nos de los mejores métodos de investigación y comprobación disponibles como cuestiones–de–hecho, métodos que son, cuando se los reúne bajo un único nombre, la ciencia. En cuanto a la verdad, pues, la filosofía no tiene un status preeminente [...] [*ibid.*, p. 410].

El objetivo de Dewey fue traer la verdad, y con ella las pretensiones de los filósofos, a la tierra. Podríamos pensar con justicia que Dewey confundió la cuestión de qué clase de concepto es la verdad con la cuestión de qué tipos de verdades hay. Pero está claro que los dos temas están relacionados, puesto que lo que cae bajo el concepto obviamente depende de lo que el concepto es. Y la idea de asegurar que el dominio de la verdad puede convincentemente traerse dentro del alcance de las capacidades humanas, recortando el concepto a medida, difícilmente es exclusivo de Dewey; Dewey se vio a sí mismo compartiendo los puntos de vista de C. S. Peirce y William James en esta cuestión, y de un modo u otro el tema básico reaparece hoy en los escritos de Hilary Putnam, Michael Dummett, Richard Rorty, y muchos otros.

Aquellos que desean desacreditar o desinflar el concepto de verdad a menudo empiezan rechazando cualquier insinuación de teoría de la correspondencia, pero Dewey<sup>3</sup> no vio ningún peligro en la idea de la

---

<sup>3</sup> *Essays in Experimental Logic*, Nueva York, Dover, 1953.

correspondencia siempre y cuando se la entendiera apropiadamente. Dijo [que] «La verdad significa, como una cuestión obvia, acuerdo, correspondencia, de idea y hecho», pero inmediatamente siguió «pero ¿qué es lo que acuerdo [o] correspondencia significan?» (*ibid.*, p. 304). Contestó «es verdadera la idea que trabaja para llevarnos a lo que se propone» (*ibid.*), y cita a James<sup>4</sup> con aprobación:

[...] cualquier idea que nos lleve prósperamente de cualquier parte de la experiencia a cualquier otra, ligando las cosas satisfactoriamente, trabajando [148] de manera segura, simplificando, ahorrando esfuerzo, es verdadera sólo por eso, verdadera de aquí en adelante [*ibid.*, p. 58].

Probablemente pocos filósofos serán ahora tentados por estas alegres y majestuosas formulaciones. Pero el problema que estaban tratando los pragmatistas —el problema de cómo se relaciona la verdad con los deseos humanos, las creencias, las intenciones y el uso del lenguaje— me parece el adecuado para concentrarse en él cuando pensamos acerca de la verdad. También me parece que no se está ahora más cerca de una solución a este problema de lo que se estaba en los días de Dewey.

---

<sup>4</sup> *Pragmatism*, Nueva York, Longmans & Green, 1907. En otra parte (*Logic: The Theory of Inquiry*, Nueva York, Holt, 1938), Dewey dice:

«La mejor definición de *verdad* desde el punto de vista lógico que conozco es la de Peirce: “La opinión que está destinada a ser aceptada al final por todos aquellos que investigan es lo que queremos decir por verdad”» (p. 58).

Pero habitualmente Dewey estaba más cerca de James: las ideas, las teorías, son verdaderas si son «instrumentales para una reorganización activa del entorno dado, para una eliminación de algún problema y perplejidad específica [...]. La hipótesis que funciona es la *verdadera*» (*Reconstruction in Philosophy*, Nueva York, Holt, 1920, p. 156).

Ver esto como el problema más importante acerca de la verdad —o en cualquier caso como un problema en absoluto— es asumir que el concepto de verdad *está* relacionado por caminos importantes con las actitudes humanas; algo de lo que no es inusual dudar. No es inusual, de hecho, dudar si el concepto de verdad tiene alguna importancia filosófica seria en absoluto.

Rorty recoge la intención de Dewey de eliminar la verdad de un ámbito tan exaltado al que sólo los filósofos se pueden atener cuando introduce sus *Consequences of Pragmatism*<sup>5</sup> con las palabras:

Los ensayos de este libro son intentos de sacar consecuencias de una teoría pragmatista acerca de la verdad. Esta teoría dice que la verdad no es la clase de cosa de la que uno esperaría tener una teoría filosóficamente interesante [...] no hay ningún trabajo interesante que hacer en esta área [*ibid.*, pp. XIII-XIV].

Pero me parece que Rorty se pierde la mitad de la miga de la actitud de Dewey hacia el concepto de verdad: Dewey dice que las *verdades* no son, en general, una provincia especial de la filosofía; pero insiste también en que la *verdad* es lo que funciona. Esto no es lo mismo que la tesis de que no hay nada interesante que decir acerca de la verdad. Dewey encontró muchas cosas interesantes que decir acerca de lo que funciona. Rorty<sup>6</sup> ha comparado mis puntos de vista sobre la naturaleza

---

<sup>5</sup> Minneapolis, Minnesota UP, 1982.

<sup>6</sup> «Pragmatism, Davidson and Truth», en LePore, ed., *Truth and Interpretation*, Nueva York, Blackweil, 1986, pp. 333–355 (ed. cast., «Pragmatismo, Davidson, y verdad», en *Objetividad, relativismo y verdad*, Paidós, Barcelona, 1996, pp. 173–205).

de [149] la verdad con los de Dewey. Encuentro agradable y penetrante mucho de lo que él tiene que decir sobre este tema, y creo que tiene razón en que, en líneas generales, comparto la actitud de Dewey hacia la verdad. En un sentido, sin embargo, un sentido al que acabo de referirme, Rorty puede haberse equivocado con nosotros dos; tal como yo lo leo, Dewey pensó que una vez que la verdad fue traída a la tierra, hubo cosas filosóficamente importantes e instructivas que decir acerca de sus conexiones con las actitudes humanas, conexiones constitutivas en parte del concepto de verdad. Éste es también mi punto de vista, aunque no creo que Dewey haya [trazado] las conexiones adecuadamente.

Rorty correctamente nota el papel fundamental que yo asigno al trabajo de Alfred Tarski, que proporciona una manera de discutir la comprensión del lenguaje, y él ve claramente que para mí esto está relacionado con el rechazo de una concepción representacional del lenguaje y de la idea de que la verdad consiste en un reflejo preciso de los hechos. Estas son cuestiones a las que volveré ahora. En este artículo, discutiré primero la noción, a menudo asociada con el enfoque de Tarski, de que el discurso de la verdad es esencialmente redundante, y que no tiene propiedades importantes más allá de aquellas especificadas en las definiciones de la verdad de Tarski. La primera sección termina con una defensa de la afirmación de que puede legítimamente considerarse que las definiciones de Tarski ofrecen verdades sustantivas acerca de un lenguaje, pero que en este caso debe de haber

---

Véase también su «Representation, Social Practise, and Truth», *Philosophical Studies*, XXX (1988): 215–228.

más en el concepto de lo que Tarski especificó. En la segunda sección del artículo vuelvo a varios intentos de decir qué más está involucrado: discuto las teorías de la correspondencia, teorías de la coherencia, y teorías que de una forma u otra hacen de la verdad un concepto epistémico. Yo rechazo todos estos tipos de teorías. En la sección tercera, propongo un enfoque que difiere del resto, uno que hace del concepto de verdad una parte esencial del esquema que todos necesariamente empleamos para entender, criticar, explicar y predecir el pensamiento y la acción.

## I. LA ESTRUCTURA DE LA VERDAD

La teoría de la redundancia es la que mejor encaja con expresiones como ‘es verdadero que’ o ‘es un hecho que’ cuando se prefijan a una oración. Tales expresiones pueden considerarse como conectivas oracionales veritativo–funcionales que, cuando se añaden a una [150] oración verdadera, dan como resultado una oración verdadera, y cuando se añaden a una oración falsa, dan como resultado una oración falsa. Estas conectivas funcionarían entonces exactamente como la doble negación (cuando la negación se concibe de manera clásica). Al menos en lo que concierne al contenido cognitivo y a las condiciones de verdad, tales añadidos son redundantes.

Frank Ramsey<sup>7</sup> parece haber pensado que todos los usos del concepto de verdad son como éste. Dice: «Es verdadero que César fue

---

<sup>7</sup> Facts and Propositions» (1927), reimpresso en *The Foundations of Mathematics*, Nueva York: Humanities, 1931, pp. 138–155.



asesinado' no significa más que César fue asesinado» (*ibid.*, p. 143). Entonces considera casos como 'Todo lo que él dice es verdadero' en los que la referencia a la verdad no se elimina tan fácilmente, y sugiere que, si nos restringimos a proposiciones de la forma  $aRb$ , podríamos tratar 'Todo lo que él dice es verdadero' como 'Para todo  $a, R, b$ , si él dice  $aRb$ , entonces  $aRb$ '. Ramsey añade que, si se incluyen todas las formas de proposición, las cosas se hacen más complicadas, «pero no esencialmente diferentes» (*ibid.*). Aunque Ramsey no siempre distingue claramente entre proposiciones y oraciones, o el uso de oraciones y su mención, uno tiene la impresión de que, si Ramsey hubiera llevado a cabo el análisis «más complicado», podría haber terminado con algo muy parecido a las definiciones de verdad de Tarski. En cualquier caso, Ramsey pensó que había dicho suficiente para mostrar que «no hay ningún problema de la verdad separado sino simplemente un enredo lingüístico» (*ibid.*, p. 142)<sup>8</sup>.

Ramsey se equivocaba si pensaba que el análisis del uso de 'verdadero' como conectiva veritativo–funcional podría aplicarse directamente a oraciones como 'Todo lo que él dice es verdadero', porque en el primer caso la expresión de verdad se considera una conectiva, mientras que en el último caso debe tratarse como un predicado y, si seguimos a Tarski, debe pertenecer a un lenguaje diferente del lenguaje de las oraciones de las cuales se predica. Sería posible tratar expresiones como 'es verdadero que' como predicados de proposiciones mejor que como conectivas oracionales, pero de nuevo la redundancia sería

---

<sup>8</sup> P. F. Strawson dice más o menos lo mismo en su famoso debate con J. L. Austin, en «Truth», *Proceedings of the Aristotelian Society, Sup. Vol. XXIV* (1950): 129–156.

mucho menos manifiesta de lo que Ramsey afirmó.

Muchos filósofos han considerado sin embargo el trabajo de Tarski esencialmente como una cuestión de aclaración de la intuición [151] de Ramsey. W. V. Quine<sup>9</sup>, por ejemplo, escribe: «Decir que el enunciado ‘Bruto mató a César’ es verdadero... es simplemente en efecto decir que Bruto mató a César», y nos dice, en una nota a pie de página, que hay que mirar a Tarski para el «desarrollo clásico» (*ibid.*, p. 24). Putnam mantiene que Rorty y Quine comparten este punto de vista acerca de la verdad. De acuerdo con Putnam<sup>10</sup>, Rorty y Quine creen que «llamar a una oración ‘verdadera’ no es adscribir una propiedad, la verdad, a una oración; es simplemente otra manera de afirmar la oración» (*ibid.*, p. 62). (Añade que a esto se le llama la «concepción desentrecomilladora» —«en la jerga de los filósofos del lenguaje davidsonianos»— (*ibid.*). Quizás es así, pero entonces yo no soy davidsoniano, porque yo no estoy tentado a referirme a las definiciones de verdad de Tarski como «desentrecomilladoras».) En cualquier caso, Putnam no acepta esta tesis; la está atacando como «puramente formal» y «vacía».

No tengo claro si Putnam piensa que el trabajo de Tarski sobre la verdad no es más que una mejora técnica sobre lo que básicamente es una teoría de la redundancia, pero otros ciertamente han tomado esta línea. Stephen Leeds<sup>11</sup> ha sugerido que la «utilidad» o importancia del

---

<sup>9</sup> *Word and Object*, Cambridge: MIT, 1960.

<sup>10</sup> «A Comparison of Something with Something Else», *New Literary History*, XVII(1985):61-79.

<sup>11</sup> «Theories of Reference and Truth», *Erkenntnis*, XI11 (1978): 111-130.

concepto de verdad podría consistir simplemente en esto, que nos da una manera de decir cosas como «la mayoría de nuestras creencias son verdaderas», donde queremos hablar de, o quizás aseverar, un conjunto de oraciones infinito o en cualquier caso no catalogable. Ramsey no explicó cómo hacer esto; Tarski sí. Paul Horwich<sup>12</sup>, como Leeds, considera que Tarski es un teórico de la redundancia; Horwich está persuadido de que, a pesar de nuestra intuición de que la verdad es concepto central e importante, «la noción de verdad fue completamente captada por Tarski» (*ibid.*, p. 192). A esta idea, que Tarski hizo todo lo que puede hacerse por el concepto de verdad, la llama Horwich la teoría deflacionista de la verdad.

Aunque no está de acuerdo con Horwich en que la verdad tal como Tarski la definió especifique las condiciones adecuadas de verdad para un tratamiento de lo que saben los usuarios del lenguaje, [152] Scott Soames<sup>13</sup> coincide en calificar de deflacionista el tratamiento de la verdad de Tarski, y como Horwich, cree que, cuando se trata de explicar el concepto de verdad, no deberíamos de pedir nada más, aparte de la aplicación de la verdad a proposiciones, etc.

Hartry Field<sup>14</sup>, en un útil artículo, explora el caso a favor de un concepto deflacionista de verdad, y muestra lo difícil que sería ir más allá de él. Explica lo que él quiere decir al llamar a una teoría de la

---

<sup>12</sup> «Three Forms of Realism», *Synthese*, LI(1982): 181–201.

<sup>13</sup> «What is a Theory of Truth?», *The Journal of Philosophy*, LXXXI, 8 (1984): 411–429.

<sup>14</sup> «The Deflationary Conception of Truth», en C. Wright y G. McDonald, eds., *Facts, Science and Morality*, Nueva York: Blackwell, 1987, pp. 55–117.

verdad deflacionista aproximadamente de la manera en que lo hace Horwich: la verdad es desentrecomilladora y nada más; pero está menos seguro que Horwich de que Tarski se vea (o deba verse) como un desentrecomillador, aunque cree que el trabajo de Tarski puede ser apropiado para el desentrecomillador. Michael Williams<sup>15</sup> ha caracterizado recientemente los puntos de vista de los desentrecomilladores de esta manera: ellos

[...] piensan que cuando hemos apuntado a ciertos rasgos formales del predicado de verdad (notablemente su rasgo ‘desentrecomillador’) y explicado por qué es útil tener un predicado como éste (por ejemplo como un mecanismo para afirmar conjunciones infinitas), hemos dicho prácticamente todo lo que hay que decir acerca de la verdad [*ibid.*, p. 424].

Él explícitamente acepta una actitud deflacionista hacia la verdad<sup>16</sup>.

¿Cómo son de plausibles estas distintas teorías deflacionistas de la verdad? Si restringimos la teoría de la redundancia a las ocurrencias de ‘verdadero’ como parte de una conectiva oracional veritativo–funcional (como en ‘es verdadero que la nieve es blanca’), entonces está claro que tales usos juegan sólo un pequeño papel en nuestro discurso de la verdad; ésta no puede ser toda la historia. ¿Pueden las teorías desentre-

---

<sup>15</sup> «Epistemological Realism and the Basis of Skepticism», *Mind*, XCVII (1988): 415–439.

<sup>16</sup> Véase «Do we (Epistemologists) need a Theory of Truth?», *Philosophical Topics*, XIV (1986): 223–242.

comilladoras hacerlo mejor? Las definiciones de verdad de Tarski son desentrecomilladoras en este sentido: dada la definición (y la teoría de conjuntos y la sintaxis formal), y dada una oración de la forma «‘la nieve es blanca’ es verdadera», podemos probar que la oración ‘la nieve es blanca’ es equivalente. Así, la oración en la que ‘la nieve es blanca’ sólo está mencionada es probablemente equivalente a la [153] oración ‘la nieve es blanca’ misma; la original «‘la nieve es blanca’» ha sido despojada de sus comillas; eliminar las comillas cancela, por así decirlo, el predicado de verdad. E incluso cuando no podemos eliminar las comillas porque no hay comillas que eliminar (como en ‘todo lo que él dijo era verdadero’ o ‘una regla válida de inferencia garantiza que de premisas verdaderas sólo se siguen conclusiones verdaderas’), Tarski nos ha mostrado cómo liberarnos del predicado de verdad, puesto que ha sido definido explícitamente<sup>17</sup>. Esto deja claro que las definiciones de verdad de Tarski no son estrictamente desentrecomilladoras, puesto que no dependen de despojar de las comillas a las oraciones individuales para eliminar los predicados de verdad. Menos aún dependen de usar oraciones reales que se dicen verdaderas para efectuar la eliminación; esto es obvio cuando la definición de verdad para un lenguaje se da en otro. No se puede encontrar un equivalente castellano de la oración inglesa «‘Schnee ist weiss’ es verdadera (en alemán)» quitando

---

<sup>17</sup> Este punto, a menudo atribuido a Leeds, fue hecho por Tarski en «The Semantic Conception of Truth», *Philosophy and Philosophical Research*, IV (1944), p. 359. Tarski nota también que el mero desentrecomillado no puede eliminar la palabra ‘verdadero’ de oraciones como ‘la primera oración escrita por Platón es verdadera’. (Pero tampoco ha mostrado Tarski cómo eliminar este uso del predicado de verdad a menos que tenga una definición de la verdad para el lenguaje hablado por Platón).

simplemente las comillas de «‘Schnee ist weiss’».

Además, queda el hecho de que los métodos de Tarski nos permitan reemplazar los predicados de verdad que él define en cualquier contexto, y que el reemplazo no deje ningún predicado explícitamente semántico tras de sí; en este sentido, sus predicados de verdad son como la conectiva oracional ‘es verdadero que’, que puede eliminarse por simple supresión. Lo que es sorprendente, por supuesto, no es que la expresión ‘es verdadero’ puede reemplazarse, porque éste puede ser el punto de la definición; lo que es sorprendente es que no se reemplace por ninguna otra cosa, semántica o de otro tipo. Es presumiblemente este rasgo lo que lleva a Putnam a decir que, de acuerdo con tales teorías, la verdad no es una propiedad. (Esto no es completamente adecuado aplicado a las definiciones de verdad de Tarski, sin embargo. Los predicados de verdad de Tarski son predicados legítimos, con una extensión que no tiene ningún predicado en el lenguaje objeto. Pero se ve lo que Putnam quiere decir con su obser-[154]vación). Putnam concluye que los predicados de verdad de Tarski no tienen nada que ver con la semántica o con la concepción común de la verdad: «Como tratamiento filosófico de la verdad, la teoría de Tarski falla tanto como pueda fallar un tratamiento» (*op. cit.*, p. 64).

Lo que está claro es que Tarski no definió el concepto de verdad, ni siquiera aplicado a oraciones. Tarski mostró cómo definir un predicado de verdad para cada uno de entre un conjunto de lenguajes que se comportan bien, pero sus definiciones, por supuesto, no nos dicen qué tienen en común estos predicados. Dicho de una forma ligeramente diferente: él definió distintos predicados de la forma ‘s es

verdadero<sub>L</sub>’, cada uno aplicable sólo a un lenguaje, pero no consiguió definir un predicado de la forma ‘s es verdadero en L’ para ‘L’ variable. La observación fue hecha por Max Black<sup>18</sup> y posteriormente por Dummett<sup>19</sup>; pero por supuesto Tarski ha hecho esto atronadoramente claro desde el principio probando que ningún predicado único de este tipo podría definirse en un lenguaje consistente, dadas sus asunciones concernientes a los predicados de verdad. Dadas estas restricciones, nunca hubo ninguna posibilidad de que diera una definición general del concepto de verdad, ni siquiera para oraciones. Si consideramos la aplicación de verdad a creencias y fenómenos relacionados como afirmaciones y aserciones, es obvio de otra manera que Tarski no intentó una definición realmente general. Considerando lo evidente que es que Tarski no dio una definición general de verdad, y el hecho de que quizá su resultado más importante fue que esto no podía hacerse siguiendo las líneas que le hubieran satisfecho, es notable cuánto esfuerzo han puesto algunos críticos en el intento de persuadirnos que Tarski no consiguió ofrecernos una definición tal.

Dummett dice en el «Prefacio» a *Truth and Other Enigmas*<sup>20</sup> que el «argumento fundamental» de su artículo anterior «Truth» era que cualquier forma de teoría de la redundancia (y él incluye a las definiciones de verdad de Tarski en esta categoría) debe ser falsa porque ninguna teoría tal puede captar el sentido de introducir un predicado de verdad. Esto puede verse, argumenta él, en el hecho de que, si

---

<sup>18</sup> *Language and Philosophy*, Ithaca: Cornell, 1949, p. 104.

<sup>19</sup> «Truth», en *Proceedings of the Aristotelian Society*, LIX (1958–9): 141–162.

<sup>20</sup> Londres: *Duckworth*, 1978.

tenemos una definición tarskiana de verdad para un lenguaje que no entendemos, [155]

no tendremos idea del sentido de introducir el predicado [...] a menos que [...] sepamos ya de antemano cuál se supone que es el sentido de un predicado así definido. Pero, si sabemos de antemano el sentido de introducir el predicado «verdadero» entonces sabemos algo acerca del concepto de verdad expresado por el predicado que no está encarnado en esta [...] definición de verdad [*ibid.*, pp. xx-xxi].

Dummett añade que «aunque este argumento era tan obvio cuando se formuló creo que mereció la pena expresarlo en el momento» (*ibid.*). Tiene razón: el argumento era obvio, y mereció la pena expresarlo, al menos para mí<sup>21</sup>. La aplicación a las teorías del significado es importante; pero el asunto es más general: Tarski sabía que no podía dar una definición general de verdad, y así no había ninguna manera formal en la que él pudiera captar «el sentido» de introducir los predicados de verdad, tanto si el sentido concernía a la conexión entre la verdad y el significado o entre la verdad y algún otro concepto o conceptos.

---

<sup>21</sup> Mi confusión en este punto es más que evidente en «Truth and Meaning», en *Inquiries in to Truth and Interpretation* (Nueva York, Oxford, 1984). Mi equivocación fue pensar que podíamos *a la vez* tomar una definición de verdad de Tarski como si nos dijera todo lo que necesitamos saber acerca de la verdad *y* usar la definición para describir un lenguaje real. Pero en el mismo artículo incluso discutí (de manera inconsistente) cómo decir que una definición tal se aplicaba a un lenguaje. Pronto reconocí el error. (Véase la «Introducción», pp. xiv-xv, y otros artículos en *Inquiries into Truth and Interpretation*.)



Dummett y otros han intentado de varias maneras hacernos a los lentos de mente apreciar el fracaso de los predicados de verdad de Tarski para captar completamente el concepto de verdad. La dificultad central, como hemos visto, se debe simplemente al hecho de que las definiciones de Tarski no nos dan idea de cómo aplicar el concepto a un caso nuevo, tanto si el caso nuevo es un nuevo lenguaje o una palabra añadida de nuevas a un lenguaje [éstas son realmente la misma cuestión, señalada de ambas maneras por Dummett (*op.cit.*) y de la segunda manera por Hartry Field<sup>22</sup>]. Este rasgo de las definiciones de Tarski puede a su vez fácilmente conectarse con el hecho de que dependen de dar la extensión o referencia de los nombres o predicados básicos mediante la enumeración de casos: una definición dada de esta manera no puede ofrecer ninguna pista para el caso general o siguiente. [156]

Un número de críticas de, o comentarios sobre, el tratamiento de la verdad de Tarski dependen del aspecto enumerativo de sus definiciones. Una de este tipo es la afirmación de que las definiciones de Tarski no pueden explicar por qué, si la palabra ‘nieve’ hubiera significado «carbón», la oración ‘la nieve es blanca’ hubiera sido verdadera si y sólo si la nieve hubiera sido negra. Putnam y Soames hacen los dos este comentario, pero para Putnam es una crítica, mientras que para Soames ilustra lo absurdo de esperar mucho de una teoría o definición de la verdad. Otra queja es que las definiciones de Tarski no establecen la conexión entre verdad y significado que muchos filósofos consideran

---

<sup>22</sup> «Tarski's Theory of Truth», *The Journal of Philosophy*, LXIX, 13 (1972): 347–375.

esencial. (De nuevo, para Putnam esto muestra qué está básicamente equivocado en la concepción de la verdad de Tarski; para Soames es un ejemplo más del aspecto loablemente deflacionista de las definiciones de Tarski.) Un comentario estrechamente relacionado es que Tarski no relaciona la verdad con el uso o los usos del lenguaje (Field, Putnam, Soames, Dummett). Cualquiera que pudiera ser el valor de estos comentarios merece la pena mantener en la mente que todos ellos se remontan al mismo rasgo simple del trabajo de Tarski: al emplear una lista finita y exhaustiva de casos básicos en el curso de definir la satisfacción (en términos de la cual se define la verdad), él necesariamente no consiguió especificar cómo continuar con otros casos.

A pesar de las limitaciones que se han identificado o imaginado en el trabajo de Tarski sobre la verdad, un número de filósofos, como hemos visto, han mantenido este trabajo como si abarcara todos los rasgos esenciales de la verdad. Entre estos filósofos se incluye a Rorty, Leeds, Michel Williams, Horwich, Soames, y, de acuerdo con Putnam, Quine; también, de acuerdo con Rorty, a mí<sup>23</sup>.

Sin embargo yo no pertenezco a esta lista. El argumento básico, que pretendía descubrir a Tarski como un deflacionista, puede tomarse de dos maneras: como mostrando que él no captó aspectos esenciales

---

<sup>23</sup> No es sorprendente que las concepciones de la gente de esta lista difieran en el sentido en el que Tarski es un deflacionista. Horwich, por ejemplo, introdujo el término 'deflacionista' al hablar de Tarski, pero mantiene que el «esquema» de Tarski da las condiciones de verdad, y así los significados, de las expresiones de un lenguaje; su concepción es esencialmente la de mi «Truth and Meaning». La mayoría de los otros piensa que el enfoque deflacionista de Tarski muestra que la verdad tal como él la define no tiene nada que ver con el significado.

del concepto de verdad, o como mostrando que el concepto de verdad no es tan profundo e interesante como muchos han [157] pensado<sup>24</sup>. Como Dummett y Putnam, creo que debemos tomarlo en la primera de estas dos maneras. La razón es simple. Nada en las definiciones de verdad de Tarski sugiere lo que estas definiciones tienen en común. A menos que estemos preparados para decir que no hay un único concepto de verdad (incluso en cuanto aplicado a oraciones), sino sólo un número de conceptos diferentes para los que usamos la misma palabra, tenemos que concluir que hay más en el concepto de verdad —algo absolutamente básico, de hecho— que las definiciones de Tarski no tocan. Lo que es ligeramente sorprendente es que algunos filósofos que apelan a una versión del argumento básico para mostrar que los predicados de verdad de Tarski son deflacionistas acepten al mismo tiempo una teoría deflacionista. Pero si el argumento básico es correcto, muestra que definiciones como la de Tarski, o teorías construidas sobre las mismas líneas, no pueden captar el concepto de verdad.

Hay además otra afirmación o asunción sobre el trabajo de Tarski que, aunque a menudo va de la mano de algunos de los puntos que acabamos de ensayar, merece una discusión separada. El tema es que, si aceptamos una de las definiciones de verdad de Tarski, entonces los

---

<sup>24</sup> La primera actitud aparece en el comentario de Putnam de que la propiedad que Tarski define no es «ni siquiera dudosa o sospechosamente ‘cercana’ a la propiedad de la verdad es sólo que no es la verdad en absoluto» (*op. cit.*, p. 64). Soames representa el segundo punto de vista: «lo que parece correcto acerca del enfoque de Tarski es su carácter deflacionista.» Pero «la noción de la verdad de Tarski no tiene nada que ver con la comprensión o la interpretación semántica» («What is a Theory of Truth?», pp. 429, 424).

enunciados que deberían, si la verdad estuviera propiamente caracterizada, ser enunciados empíricos se convierten en verdades de la lógica. Así, de acuerdo con Putnam, una oración como «‘Schnee ist weiss’ es verdadera (en alemán) si y sólo si la nieve es blanca» debería ser una verdad sustantiva acerca del alemán, pero si por el predicado ‘s es verdadero (en alemán)’ sustituimos un predicado definido al estilo de Tarski, la evidente verdad sustantiva se convierte en una verdad de la lógica<sup>25</sup>. Es fácil ver que lo que quiera que haya en este argumento depende del mismo rasgo del método de Tarski que hemos estado discutiendo: si la extensión de un predicado se define mediante la lista de cosas a las que se aplica, el aplicar el predicado a un elemento de la lista dará como resultado un enunciado equivalente a una verdad lógica. (Por razones técnicas ésta es una ex-[158]plicación sobresimplificada de este aspecto del método de Tarski cuando el lenguaje objeto incluye cuantificadores, etc. La fuerza del comentario permanece<sup>26</sup>.) Ésta parece ser la razón principal de Putnam para decir que Tarski fracasó «tanto como es posible fracasar» al ofrecer un tratamiento filosófico de la verdad. Soames pudiera estar pensando en la misma línea cuando mantiene que la única manera de defender la interpretación filosófica de Tarski de su trabajo es rechazar la exigencia de que las aplicaciones de sus predicados de verdad y satisfacción tengan contenido empírico. Cumplir la exigencia sería, dice Soames, «incom-

---

<sup>25</sup> Para versiones de este argumento, véase Putnam, *op. cit.*, y «On Truth», en Leigh Cauman *et al*, eds., *How Many Questions* (Indianápolis, Hackett, 1983), pp. 35–56.

<sup>26</sup> Para el desarrollo de este tema, véanse los trabajos de Putnam a los que nos referimos en la última nota a pie de página; también Soames, *op. cit.*; y John Etchemendy, «Tarski on Truth and Logical Consequence», *The Journal of Symbolic Logic*, LII(1988): 51–79.

patible» con el trabajo de Tarski (*op. cit.*, p. 425).

El argumento es explicado con alguna extensión por John Etchemendy (*op.cit.*). De acuerdo con Etchemendy, el objetivo de Tarski era formular predicados con dos propiedades: primero, deberían relacionarse de una manera específica con el concepto intuitivo de verdad y, segundo, deberían de estar garantizados, en la medida de lo posible, contra la amenaza de paradoja e inconsistencia. La primera condición se cumplía al inventar un concepto que pudiera fácilmente mostrarse que se aplicaba a todas las oraciones verdaderas de un lenguaje y sólo a ellas. La relación con el concepto intuitivo de verdad se hace manifiesta mediante la convención-T. La convención-T exige que el predicado de verdad ‘s es verdadero<sub>L</sub>’ para un lenguaje L se caracterice de tal modo que implique, para toda oración s de L, un teorema de la forma ‘s es verdadero<sub>L</sub> si y sólo si p’, cuando ‘s’ se reemplaza por una descripción sistemática de s y p se reemplaza por una traducción de s al lenguaje de la teoría. Llamemos a estos teoremas oraciones-T. El predicado en las oraciones-T, ‘s es verdadero<sub>L</sub>’, es un predicado monádico; el subíndice no es una variable, sino el nombre o la descripción de un lenguaje particular y una parte no desligable del predicado. La relación con el concepto común de verdad es evidente a partir del hecho de que las oraciones-T siguen siendo verdaderas si por el predicado de verdad al estilo de Tarski sustituimos el predicado castellano ‘s es verdadero en L’. (Este es un predicado diádico: podemos sustituir nombres o descripciones de otros lenguajes en el lugar de ‘L’.) La exigencia de que el predicado de verdad no [159] amenace con introducir inconsistencias en la teoría o el lenguaje se cumple al dar una definición explícita del predicado sin usar conceptos semánticos; así, cualquier desafío a la

consistencia que estos conceptos pudieran presentar ha sido evitado. Si el metalenguaje es consistente antes de la introducción del predicado de verdad, está garantizado que permanecerá así después de la introducción.

Las oraciones-T que contienen predicados de verdad de Tarski *parecen* transmitir hechos sustantivos acerca del lenguaje objeto, a saber, que sus oraciones son verdaderas bajo las condiciones especificadas por la oración-T ('Schnee ist weiss' es verdadero en alemán si y sólo si la nieve es blanca), pero de hecho, dice Etchemendy, «no portan ninguna información acerca de las propiedades semánticas del lenguaje, *ni siquiera acerca de las condiciones de verdad de sus oraciones*» (su énfasis, *op. cit.*, p. 57). La razón de esto es que las oraciones-T son verdades de la lógica, y así no pueden decirnos nada que la lógica sola no pudiera decirnos. Las oraciones-T son verdades de la lógica, a su vez, porque se siguen de las definiciones de Tarski, y éstas son simplemente estipulaciones; nos hemos desorientado a causa de «la facilidad con la que leemos contenido sustantivo en lo que se pretende como definiciones estipulativas, la facilidad con la que reemplazamos el 'si y sólo si' de la definición por el 'si y sólo si' de los axiomas o teoremas» (*op. cit.*, p. 58). Si queremos afirmar hechos sustantivos acerca de un lenguaje, debemos sustituir en las oraciones-T y en todas partes un predicado que transmita algo como el concepto intuitivo de verdad. Si hacemos esto, «las oraciones que hacemos se parecerán a veces sorprendentemente a las cláusulas» en las definiciones de Tarski y (si esto es correcto) darán como resultado información genuina, acerca de las propiedades semánticas de un lenguaje.

Pero, y éste es el mensaje central de Etchemendy, las dos empresas —la de definir la verdad de acuerdo con los objetivos de Tarski, y la de proporcionar un tratamiento semántico formal pero sustantivo de un lenguaje— no sólo son empresas totalmente diferentes, sino que están en «una oposición bastante directa la una con la otra... Porque sin dejar al lado el objetivo principal de Tarski, hay un sentido en el que la semántica simplemente no puede hacerse» (*op. cit.*, pp. 52-3). La diferencia entre las dos es que la primera exige un predicado que pueda eliminarse sin residuos de todos los contextos, mientras que la segunda requiere una noción de verdad «fija, metateórica». Emplear el segundo concepto frustraría directamente el sentido del proyecto de Tarski. Así, la relación entre el logro pretendido y conseguido con éxito de Tarski, por un lado, y el proyecto de proporcionar una manera de describir la semántica de los lenguajes interpretados, por otro, es «poco más que un accidente fortuito» (*op. cit.*, pp. 52-3).

Putnam, Soames y Etchemendy están de acuerdo en que las oraciones-T de Tarski sólo parece que expresan verdades empíricas acerca de un lenguaje; son de hecho «tautologías» (Putnam). Ellos difieren en sus valoraciones de la tesis en la que están de acuerdo: Putnam<sup>27</sup> cree que lo que Tarski definió «simplemente no es la verdad en absoluto»; Soames y Etchemendy afirman que Tarski sí que hizo lo que se propuso. Soames mantiene que Tarski tenía razón al dar un tratamiento deflacionista de la verdad, mientras que Etchemendy cree que la semántica empírica es un estudio legítimo que Tarski no estaba persiguiendo.

---

<sup>27</sup> «A Comparison of Something with Else», p. 64.

¿Qué deberíamos pensar acerca de estas afirmaciones? Una cosa es cierta: Tarski no estaba de acuerdo con estas valoraciones de sus resultados. En «The Semantic Conception of Truth»<sup>28</sup>, hay una sección titulada «Conformity of the Semantic Conception of Truth with Philosophical and Common–Sense Usage». Permítanme que cite de ella:

En lo que a mi propia opinión concierne, no tengo ninguna duda de que nuestra formulación está conforme con el contenido intuitivo de la de Aristóteles... se han expresado algunas dudas acerca de si la concepción semántica refleja la noción de verdad en su uso común y cotidiano. Me doy cuenta claramente [...] de que el significado común de la palabra «verdadero» —como el de cualquier otra palabra del lenguaje de todos los días— es hasta cierto punto vago [...]. Así [...] toda solución a este problema implica necesariamente una cierta desviación de la práctica del lenguaje de todos los días.

A pesar de todo esto, ocurre que creo que la concepción semántica está conforme en una medida considerable con el uso de sentido común [...] [*ibid.*, p. 360].

Al plantear su problema, Tarski no se distancia del proyecto de caracterizar conceptos que pueden usarse como se usan los conceptos semánticos ordinarios; conceptos que expresan, como él dice, «conexiones entre las expresiones de un lenguaje y los objetos y estados de cosas referidos por esas expresiones»<sup>29</sup>. Él no se propone, [161] dice,

---

<sup>28</sup> *Philosophy and Phenomenological Research*, IV (1944): 341–375.

<sup>29</sup> «The Establishment of Scientific Semantics», en *Logic. Semantics, Metamathematics*. Nueva York: Oxford, 1956, p. 401.



asignar un significado nuevo a una palabra antigua, sino «capturar el significado real de una noción antigua»<sup>30</sup>. En otras palabras, él es bastante explícito en que no pretende, como mantiene Etchemendy, que sus definiciones sean puramente estipulativas.

Tarski describe su proyecto como «The Establishment of Scientific Semantics», y dice que «los conceptos semánticos expresan ciertas relaciones entre objetos (y estados de cosas) referidos por el lenguaje que se discute y expresiones del lenguaje que refieren a esos objetos»<sup>31</sup>. Él considera la verdad de una oración como su «correspondencia con la realidad» (*ibid.*). Tarski considera estas caracterizaciones de los conceptos semánticos como «vagas», pero claramente estarían totalmente equivocadas si los conceptos semánticos no tuvieran ninguna aplicación empírica. Cuando Tarski exige que sus definiciones sean «materialmente adecuadas y en concordancia con el uso ordinario», argumenta que la convención-T es justo lo que nos asegura que la condición se cumple. El argumento es éste: dado un lenguaje que entendemos, un lenguaje interpretado como el castellano, reconocemos como verdaderas todas las oraciones de la forma «‘la nieve es blanca’ es verdadera si y sólo si la nieve es blanca». Tarski llama a tales oraciones «definiciones parciales» de la verdad. Obviamente, una definición que implica todas estas oraciones tendrá la misma extensión que el concepto intuitivo de verdad con el cual empezamos. Admitir esto es contar a las oraciones-T como si tuvieran contenido empírico; de otra manera la convención-T no tendría sentido, ni lo tendría la insistencia

---

<sup>30</sup> «The Semantic Conception of Truth», p. 341.

<sup>31</sup> «The Establishment of Scientific Semantics», pp. 403-4.

de Tarski de que está interesado en definir la verdad sólo para lenguajes interpretados.

Debemos concluir, creo, que, si Etchemendy, Soames, y Putnam tienen razón, Tarski confundió completamente su objetivo y la naturaleza de lo que consiguió. Sin embargo, sorprendentemente se necesita poco para reconciliar a Tarski con Etchemendy. Etchemendy acepta, por supuesto, que «Tarski introdujo precisamente las técnicas matemáticas que se necesitaban para un tratamiento iluminador de las propiedades semánticas de ciertos lenguajes simples» y «llegar desde una definición tarskiana de verdad a un tratamiento sustantivo de las propiedades semánticas del lenguaje objeto puede involucrar tan poco como la reintroducción de una noción primitiva de verdad» (*op. cit.*, pp. 59-60). El truco es sólo añadir a la definición de Tarski [162] de un predicado de verdad para un lenguaje L (digamos, 's es verdadero<sub>L</sub>') el comentario de que el predicado de Tarski vale para todas las oraciones verdaderas de L y sólo para ellas. Aquí, por supuesto, la palabra 'verdadero' expresa el concepto no definido, sustantivo, de la vida real, que necesitamos para la semántica seria. Llamemos a este comentario el *Axioma de Verdad*.

La primera cosa que hay que notar es que, si el lenguaje era consistente antes de que añadiésemos el axioma de verdad, el axioma de verdad no puede hacerlo inconsistente en la medida en que no dotamos formalmente a nuestro nuevo predicado de ninguna propiedad. Puede *tener* todo tipo de propiedades interesantes y no se hará ningún daño formal si las propiedades no se meten explícitamente en la teoría; y no se hará ningún daño informal si las propiedades adicionales no

llevan a contradicción.

Añadir el axioma de verdad es, desde un punto de vista formal, inocuo; es además inútil. Porque podemos de la misma manera considerar el predicado de verdad de Tarski 's es verdadero<sub>L</sub>' como si tuviera las propiedades de nuestro predicado de la vida real 's es verdadero en L', siempre y cuando estas propiedades no creen inconsistencias. La objeción a este pensamiento es que ya no podemos sentirnos confiados de que, si tuviéramos que especificar todas las propiedades del predicado de la vida real, podrían resultar inconsistencias; no sabemos exactamente qué significa nuestro predicado de verdad. La «definición» de verdad ya no es una definición puramente estipulativa.

Consideremos un lenguaje objeto formalizado y un metalenguaje exactamente como aquellos descritos por Tarski en las secciones 2 y 3 de «The Concept of Truth in Fomalized Languages»<sup>32</sup>. Ahora añadamos las definiciones de Tarski que conducen a, e incluyen, la definición de verdad; pero no las *llamemos* definiciones, y pensemos en ellas como si emplearan expresiones empíricamente significativas apropiadas para describir la semántica del lenguaje objeto (que ha sido interpretado por Tarski como si fuera sobre el cálculo de clases). De acuerdo con Etchemendy, la diferencia entre este nuevo sistema y el original de Tarski es extrema: el nuevo sistema describe correctamente la semántica del lenguaje objeto, mientras que el sistema de Tarski meramente define un predicado que no puede usarse para afirmar nada, verdadero o falso, acerca de ningún lenguaje interpre-[163]tado particular. Las definiciones de Tarski convierten a las oraciones-T implicadas en

---

<sup>32</sup> En *Logic, Semantics, and Metamathematics*.

verdades lógicas; el nuevo sistema las conserva como comentarios instructivos acerca de las condiciones de verdad de las oraciones. Pero este cambio poderoso no toca el sistema formal de ninguna manera; es un cambio en cómo describimos el sistema, no en el sistema mismo. Si el sistema de Tarski es consistente, lo mismo lo es el nuevo.

Todo el asunto gira, pues, alrededor de cómo consideramos las definiciones. Algunas definiciones claramente pretenden introducir palabras nuevas; otras tienen como objetivo expresar verdades sustantivas de un tipo u otro. Como hemos visto, Tarski no pretendió que sus definiciones encajaran un significado nuevo en un término viejo, sino «captar el significado real de una noción vieja»<sup>33</sup>.

Deberíamos ahora echar una mirada retrospectiva al tema, que no sólo se encuentra en Etchemendy, sino también en Putnam y Soames, de que las definiciones de verdad de Tarski no pueden tener nada que ver con la semántica o la interpretación de lenguajes reales porque, dadas sus definiciones, los teoremas relevantes (es decir, las oraciones-T) son verdades lógicas. De hecho, son verdades lógicas sólo sobre la asunción de que las definiciones de verdad de Tarski son puramente estipulativas, que nos dicen todo lo que hay que saber acerca del predicado que él define. No hay razón para aceptar esta asunción. Una

---

<sup>33</sup> Etchemendy sugiere que el 'si y sólo si' de una definición no tiene el mismo significado que el 'si y sólo si' de una afirmación sustantiva, pero yo no creo que este comentario deba tomarse en serio puesto que la diferencia no produce ninguna diferencia en absoluto dentro del sistema, y si tuviéramos que marcar la supuesta diferencia introduciendo símbolos diferentes, las reglas de inferencia del sistema tendrían que alterarse. Etchemendy dice que su sugerencia no pretendía ser en serio (conversación privada).

analogía simple dejará esto claro. Supongamos que ofrecemos como una definición del predicado 'x es un planeta solar' lo siguiente: x es un planeta solar si y sólo si x es exactamente uno de los siguientes: Mercurio, Venus, La Tierra, Marte, Júpiter, Saturno, Urano, Neptuno, Plutón. Esto implica la oración-P 'Neptuno es un planeta solar'. ¿Es esto último una verdad lógica? Uno también podría decir eso si nuestra definición fuera puramente estipulativa, de otra manera no. La cuestión de si es puramente estipulativa no es una que pueda responderse estudiando el sistema formal; tiene que ver con las intenciones de la persona que hace la definición. Si simplemente se nos presentara la oración que define, a duras penas podríamos dejar de darnos cuenta de que, si interpretamos las palabras más o menos de la forma usual, expresa una verdad sustantiva. Apelando a la convención-T, Tarski nos invita a fijarnos en un rasgo análogo de sus definiciones de verdad.

¿Qué deberíamos concluir acerca de cómo Tarski pretendió que tomáramos sus definiciones? Las indicaciones pueden parecer ambiguas. Por un lado, tenemos su afirmación repetida y explícita de que él quería, y pensó que lo había hecho, «captar el significado real» del concepto intuitivo de verdad, en la medida en que esto era posible; por otra parte, él claramente dependía del hecho de que sus definiciones permitieran la eliminación de todo el vocabulario semántico explícito para garantizar que su concepto no introduciría inconsistencias en un lenguaje de otro modo consistente. Pero ¿muestra esto que Tarski estaba confundido? Creo que no. Aquí hay una manera de ver el asunto.

Las definiciones de Tarski dotan a sus predicados de verdad con

propiedades que aseguran que definen la clase de oraciones verdaderas de un lenguaje. Si los predicados no tienen otras propiedades, sabemos que no engendrarán inconsistencias. Esto hace útiles a los predicados para ciertos propósitos. Si pensamos en los predicados de verdad como si tuvieran otras propiedades no especificadas, no podemos estar seguros de que aquellas propiedades no causen problemas si se las hace explícitas. Pero no hay nada que nos prohíba trabajar dentro del sistema de Tarski y reconocer al mismo tiempo que los predicados de verdad pueden tener otras propiedades esenciales, siempre y cuando no hagamos uso de las propiedades no especificadas. De esta forma, podemos tomar todas las ventajas del logro técnico de Tarski y sin embargo no tratar a los contenidos de sus teorías como «vacíos» o «meramente» formales.

Considerar el trabajo de Tarski bajo esta luz es admitir que hay un sentido en el que él no define un concepto de verdad, ni siquiera para lenguajes particulares. Él definió la clase de las oraciones verdaderas dando la extensión del predicado de verdad, pero no dio el significado. Esto se sigue en el momento en que decidimos que las oraciones-T tienen contenido empírico, porque esto implica que hay más en concepto de verdad de lo que la definición de Tarski nos dice. Mi argumento no es que Tarski pudiera, después de todo, haber captado un concepto sustancial de verdad, sino que no necesariamente nos confundimos si interpretamos sus sistemas formales como teorías empíricas sobre los lenguajes. Al hacer eso, evitamos dos tesis potencialmente mutilantes acerca de la verdad, tesis que, como [165] hemos visto, son bastante comunes hoy. Una es que el trabajo de Tarski está bastante poco relacionado con el concepto de verdad como lo enten-

demos comúnmente, de modo que, si queremos estudiar la semántica de los lenguajes interpretados, debemos de tomar otro rumbo. Caribdis es la tesis de que, aunque la versión de la verdad de Tarski es meramente desentrecorolladora, dice todo lo que hay que decir acerca del concepto de verdad.

Mi propio punto de vista es que Tarski no ha dicho mucho de lo que queremos saber acerca del concepto de verdad, y de que debe haber más. Debe haber más porque no hay ninguna indicación en el trabajo formal de Tarski de qué es lo que sus distintos predicados de verdad tienen en común, y esto debe formar parte del contenido del concepto. No es suficiente señalar a la convención-T como esta indicación, porque no dice nada de la cuestión de cómo sabemos que una teoría de la verdad para un lenguaje es correcta. El concepto de verdad tiene conexiones esenciales con los conceptos de creencia y significado, pero esas conexiones no están tocadas por el trabajo de Tarski. Es aquí donde deberíamos esperar destapar lo que echamos en falta en las caracterizaciones de Tarski en los predicados de verdad.

Lo que Tarski ha hecho por nosotros es mostrar en detalle cómo describir el tipo de patrón que la verdad debe hacer, tanto en el lenguaje como en el pensamiento. Lo que necesitamos hacer ahora es decir cómo identificar la presencia de tal patrón o estructura en la conducta de la gente.

## II. VERDAD Y CONOCIMIENTO

Si no hubiera nada más que saber acerca del concepto de verdad

que lo que podemos aprender de las definiciones de Tarski de los predicados de verdad, no tendríamos un uso claro para el concepto de verdad aparte de la conveniencia menor de su función desentrecamiadora, puesto que Tarski ha mostrado cómo eliminar tales predicados sin residuo semántico. Cualquier conexión de la verdad con el significado o la creencia sería discutible. Si consideramos que las definiciones de Tarski son puramente estipulativas, los teoremas que tales predicados nos permiten probar, en particular las oraciones-T, son equivalentes a las verdades de la lógica; a menos que leamos más en los predicados de verdad de lo que las definiciones proporcionan, estos teoremas no pueden, por tanto, ofrecer verdades empíricas acerca de las oraciones de cualquier lenguaje, y no pueden tomarse como si dieran las condiciones de verdad de las oraciones.

Tarski nunca afirmó que sus predicados hicieran más que escoger la clase de las oraciones verdaderas en lenguajes particulares. Ciertamente él no creía que hubiera definido un predicado de verdad general, ni se propuso exceder los límites de la extensionalidad. Captar el significado, como algo distinto de la extensión, no era parte de su proyecto. Ni le importaba que pudiera haber otras maneras de caracterizar las mismas clases de oraciones —maneras que pudieran ser más iluminadoras para propósitos diferentes del suyo—.

Los dos puntos están relacionados puesto que no hay ninguna manera evidente de dar una caracterización general de la verdad sin introducir criterios de un tipo bastante diferente a aquellos a los que Tarski apeló. Se sugiere a veces por los partidarios de una concepción deflacionista de la verdad que la convención-T proporciona una



respuesta adecuada a la cuestión de lo que tienen en común los distintos predicados de verdad de Tarski. Pero no deberíamos darnos por satisfechos con esta idea. Porque en aquellos casos en los que el lenguaje objeto está contenido en el metalenguaje, el requisito es meramente sintáctico: nos dice algo acerca de los predicados, pero no mucho acerca del concepto. En otros casos su aplicación depende de nuestra previa comprensión de la noción de traducción, un concepto mucho más oscuro que el de la verdad. El punto central es éste: aparte de nuestra comprensión del concepto de traducción, la convención-T no nos da ninguna idea de cómo decir en general cuándo uno de los predicados de verdad de Tarski se aplica a un lenguaje particular. El no define el concepto de traducción<sup>34</sup>.

Todavía nos falta, pues, un enfoque satisfactorio del rasgo o rasgos generales del concepto de verdad que no podemos encontrar en Tarski. Sin embargo, podemos aprender mucho de Tarski. Sus construcciones hacen, por ejemplo, evidente que, para un lenguaje con algo como el poder expresivo de un lenguaje natural, la clase de las oraciones verdaderas no puede caracterizarse sin introducir una relación como la satisfacción, que conecta palabras (términos singulares, predicados) con objetos. Si consideramos la satisfacción como

---

<sup>34</sup> Michael Williams dice que un deflacionista piensa que «lo que se lleva de un lenguaje a otro... es la utilidad, para cada lenguaje, de tener su propio mecanismo desentrecomillador» —«Scepticism and Charity», *Ralio* (New Series), I (1988), p. 180—. Pero aparte de asignar un significado claro a la «utilidad» de un mecanismo, está el hecho de que en un lenguaje podemos hablar de la verdad en otro lenguaje; y aquí la generalización sugerida por Williams no puede hacerlo mejor que la convención-T, con su apelación esencial a la traducción.

una forma generalizada de referencia, Tarski ha mostrado cómo la verdad de las oraciones depende de los rasgos semánticos (i.e., referencia) de ciertas partes propias de las oraciones. (Por supuesto, Tarski no define el concepto general de referencia más de lo que lo hace con el de verdad.) Así, incluso sin una respuesta a la cuestión de cómo sabemos cuándo una definición de verdad se aplica a un lenguaje dado, Tarski ha mostrado cómo el concepto de verdad puede usarse para dar una descripción clara de un lenguaje. Por supuesto, para dar tal descripción, debemos captar el concepto de verdad primero; pero podemos captar eso sin ser capaces de formular una descripción sistemática de un lenguaje. La convención-T conecta nuestra comprensión ingenua del concepto con la ingeniosa maquinaria de Tarski; nos persuade de que los trabajos de la maquinaria están de acuerdo con el concepto como lo conocíamos.

Esto es, entonces, lo que podemos aprender de Tarski acerca del concepto de verdad: puesto que es obvio que no ha definido el concepto general de verdad, podemos ignorar la sugerencia de que sus definiciones estipulativas captan todo lo que hay de este concepto. Pero no hay razón para no hacer uso de la estructura que llevó a las definiciones de Tarski. Para hacer esto, no necesitamos hacer ningún cambio en los sistemas formales de Tarski; una vez que nos damos cuenta de que esos sistemas no reflejan aspectos importantes de los conceptos de verdad y referencia, podemos considerar a los predicados de verdad y referencia (satisfacción) como *primitivos* en las cláusulas que llevan a las caracterizaciones recursivas de Tarski de referencia y verdad. Si encontramos que la palabra ‘definición’ encaja mal con la idea de que los predicados son primitivos, podemos eliminar la palabra; esto no

cambiará el sistema. Pero para honrar el reconocimiento de que los predicados semánticos son primitivos, podemos eliminar el paso final que para Tarski convierte las caracterizaciones recursivas en definiciones explícitas, y considerar los resultados como teorías axiomatizadas de la verdad<sup>35</sup>. [168]

Una teoría axiomatizada de la verdad puede compararse con, digamos, la axiomatización de la probabilidad de Kolmogorov, que pone claras restricciones al concepto de probabilidad, pero deja abiertas cuestiones tales como si la probabilidad tiene que ser caracterizada posteriormente como frecuencia relativa, grado de creencia, o alguna otra cosa. El tratamiento axiomático de Ramsey de la preferencia ante la incertidumbre, cuando se aplica a un agente particular, es análogo a

---

<sup>35</sup> Tarski reconoció la posibilidad de dar teorías axiomáticas de la verdad, y señaló que ‘no hay nada esencialmente equivocado en tal procedimiento semántico y podría ser útil para varios propósitos’ («The Semantic Conception of Truth», p. 352). Tarski tenía un número de razones para preferir una definición explícita a un tratamiento axiomático del concepto de verdad. Primero, él nota que la elección de axiomas «tiene más bien un carácter accidental, dependiendo de factores inesenciales (tales como por ejemplo el estado real de nuestro conocimiento)». Segundo, sólo una definición explícita puede garantizar la consistencia del sistema resultante (dada la consistencia del sistema previa a la introducción de nuevos conceptos primitivos); y, tercero, sólo una definición explícita puede dominar las dudas de si el concepto está ‘en armonía con los postulados de la unidad de la ciencia y el fisicalismo’ («The Establishment of Scientific Semantics», pp. 405–6). El primer peligro se evita si los axiomas se restringen a las cláusulas recursivas que se necesitan para caracterizar la satisfacción; escapamos del segundo (menos concluyentemente) tan pronto como las maneras conocidas de producir paradojas no se introducen; y la amenaza de que la verdad podría resultar no ser reducible a conceptos físicos es una amenaza de la cual, en mi opinión, ni podemos ni deberíamos querer escapar.

una teoría axiomatizada de la verdad en el siguiente respecto, que ofrece una teoría separada para cada agente, del mismo modo en que las teorías de la verdad tarskianas son peculiares de un lenguaje, o, como propondré, de un individuo.

Así como una teoría tarskiana no nos dice cómo determinar el que la teoría se aplique a un lenguaje o a un hablante particulares, así nada en las teorías de Ramsey nos dice cuándo una teoría tal se aplica a un agente particular. La cuestión en el caso de la teoría de la decisión es, en parte, especificar las condiciones que un agente debe satisfacer para que se diga que prefiere un objeto o curso de acción a otro. En el caso de una teoría de la verdad, lo que queremos saber es cómo decir cuándo las oraciones-T (y así la teoría como un todo) describen el lenguaje de un grupo o de un individuo. Esto obviamente requiere especificar al menos parte del contenido del concepto de verdad que los predicados de verdad de Tarski no consiguen captar.

¿Qué añadimos, entonces, a las propiedades de verdad que Tarski ha delineado cuando aplicamos el concepto intuitivo de verdad? Aparte de la posición de que Tarski dijo todo lo que puede o debería decirse acerca de la verdad, una posición que discutí y rechacé en la primera sección de este artículo, creo que la mayoría de las propuestas contemporáneas caen en dos categorías amplias: aquellos que hu-[169]manizan la verdad haciéndola básicamente epistémica, y aquellos que promocionan alguna forma de teoría de la correspondencia.

Muchos filósofos, en particular recientemente, han mantenido que la verdad es un concepto epistémico: incluso cuando no han mantenido explícitamente esta tesis, sus posiciones la han implicado a

menudo. Las teorías de la verdad como coherencia se mueven habitualmente por un motor epistémico, como lo hacen las caracterizaciones más pragmáticas de la verdad. El antirrealismo de Dummett y Crispin Wright, la idea de Peirce de que la verdad es donde la ciencia acabara si continúa el tiempo suficiente, la afirmación de Richard Boyd de que la verdad es lo que explica la convergencia de las teorías científicas y el realismo interno de Putnam, todas incluyen o implican un enfoque epistémico de la verdad. Quine también ha mantenido, al menos a veces, que la verdad es interna a una teoría del mundo y así que en esta medida depende de nuestra postura epistemológica. El relativismo acerca de la verdad es quizás siempre un síntoma de infección por el virus epistemológico; esto parece en cualquier caso ser verdad para Quine, Nelson Goodman y Putnam.

Aparentemente opuestas a estas concepciones está la idea intuitiva de que la verdad, quitando algunos pocos casos especiales, es completamente independiente de nuestras creencias; como se dice a veces, nuestras creencias podrían ser exactamente como son y sin embargo la realidad —y así la verdad acerca de la realidad— ser muy diferente. De acuerdo con esta intuición, la verdad es ‘radicalmente no epistémica’ (así caracterizó Putnam el ‘realismo transcendental’), o ‘transcendente a la evidencia’ (para usar la expresión de Dummett para el realismo). (Tanto Putnam como Dummett se oponen, por supuesto, a estas concepciones.) Si estuviéramos buscando etiquetas para estas dos concepciones de la verdad, podríamos quedarnos con los adjetivos ‘epistémico’ y ‘realista’; la aserción de una atadura esencial a la epistemología introduce una dependencia de la verdad de lo que de algún modo puede ser verificado por criaturas racionales finitas, mientras

que la negación de cualquier dependencia de la verdad de la creencia u otra actitud humana define un uso filosófico de la palabra ‘realismo’.

En la siguiente y última sección de este artículo, esbozo una aproximación al concepto de verdad que rechaza ambas concepciones de la verdad. No me propongo reconciliar las dos posiciones. Considero insostenibles las concepciones epistémicas, e ininteligibles en último extremo a las concepciones realistas. Que ambas concepciones, que sin duda responden a intuiciones poderosas, están [170] fundamentalmente equivocadas está al menos sugerido por el hecho de que ambas invitan al escepticismo. Las teorías epistémicas son escépticas de la misma manera en que son escépticos el idealismo o el fenomenalismo; son escépticos no porque hagan a la realidad incognoscible, sino porque reducen la realidad a mucho menos de lo que creemos que es. Las teorías realistas, por otra parte, parecen arrojar en la duda no sólo nuestro conocimiento de lo que es «transcendente a la evidencia», sino a todo el resto de lo que creemos que conocemos, porque tales teorías niegan que lo que es verdad esté conectado conceptualmente de alguna manera con lo que creemos.

Consideremos el proyecto de dar contenido a una teoría de la verdad. Las definiciones de Tarski se alcanzan normalmente a través de varios pasos. Primero, hay una definición de lo que es ser una oración en el lenguaje objeto; después una caracterización recursiva de una relación de satisfacción (la satisfacción es una versión de la referencia altamente generalizada); la caracterización recursiva de la satisfacción se convierte en una definición explícita a la manera de Gottlob Frege y Dedekind; después la verdad se define sobre la base de los conceptos de

oración y satisfacción. Estamos eliminando el paso que convierte a la caracterización recursiva de la satisfacción en una definición, haciendo así explícito el hecho de que estamos tratando a los predicados de verdad y satisfacción como primitivos.

Desde un punto de vista formal, es una cuestión de elección cuál de los dos conceptos semánticos, satisfacción o verdad, tomemos como básico. La verdad, como Tarski mostró, se define fácilmente sobre la base de la satisfacción; pero, alternativamente, la satisfacción puede considerarse como cualquier relación que ofrezca un enfoque correcto de la verdad. El trabajo de Tarski puede parecer que da señales inciertas. El hecho de que la verdad de las oraciones se defina apelando a las propiedades semánticas de las palabras sugiere que, si pudiéramos dar un enfoque satisfactorio de las propiedades semánticas de las palabras (esencialmente, de la referencia o de la satisfacción), entenderíamos el concepto de verdad. Por otro lado, el papel clave de la convención-T para determinar que la verdad, como se caracteriza por la teoría, tiene la misma extensión que el concepto intuitivo de verdad, hace parecer que es la verdad más que la referencia lo que es el primitivo básico. La segunda es, creo, la concepción correcta. En su apelación a la convención-T, Tarski asume, como hemos visto, una captación previa del concepto de verdad; entonces muestra cómo esta intuición puede completarse en detalle para lenguajes particulares. Esta compleción requiere la introducción de un [171] concepto referencial, una relación entre palabras y cosas —alguna relación como la satisfacción—. La historia acerca de la verdad genera un patrón en el lenguaje, el patrón de las formas lógicas, o gramática propiamente concebida, y el entramado de dependencias semánticas. No hay forma de contar esta

historia, que, siendo acerca de la verdad, es acerca de oraciones o de sus ocasiones de uso, sin asignar papeles semánticos a las partes de las oraciones. Pero no se apela a una comprensión previa del concepto de referencia.

Esta manera de concebir una teoría de la verdad va en contra de la tradición. De acuerdo con la tradición, nunca podríamos llegar a entender oraciones en su amplia o incluso infinita colección a menos que entendamos las palabras, tomadas de un vocabulario finito, de las que están compuestas; por tanto, las propiedades semánticas de las palabras deben aprenderse antes de que entendamos las oraciones y las propiedades semánticas de las palabras tienen prioridad conceptual porque son ellas las que explican las propiedades semánticas —por encima de todo las condiciones de verdad— de las oraciones. Creo que esta línea de argumento, que comienza con una perogrullada, termina con una conclusión falsa; así que algo debe estar mal. El error es confundir el orden de la explicación que es apropiado una vez que la teoría está, con la explicación de por qué la teoría es correcta. La teoría es correcta porque ofrece las oraciones-T correctas; su corrección se contrasta contra nuestra captación del concepto de verdad tal como se aplica a oraciones. Puesto que las oraciones-T no dicen nada en absoluto acerca de la referencia, la satisfacción, o de las expresiones que no son oraciones, la contrastación de la corrección de la teoría es independiente de las intuiciones que conciernen a estos conceptos. Una vez que tenemos la teoría, sin embargo, podemos explicar la verdad de las oraciones sobre la base de sus estructuras y de las propiedades semánticas de las partes. La analogía con las teorías de la ciencia es completa: para organizar y explicar lo que observamos



directamente, postulamos objetos y fuerzas no observadas u observadas indirectamente; la teoría se contrasta mediante lo que se observa directamente.

La perspectiva sobre el lenguaje y la verdad que hemos ganado es ésta: lo que está abierto a la observación es el uso de las oraciones en contexto, y la verdad es el concepto semántico que mejor entendemos. La referencia y las nociones semánticas relacionadas como la satisfacción son, por comparación, conceptos teóricos (como lo son las nociones de término singular, predicado, conectiva oracional, y el resto). No puede cuestionarse la corrección de estos conceptos teóri- [172]cos más allá de la cuestión de si ofrecen un enfoque satisfactorio del uso de las oraciones.

Un efecto de estas reflexiones es fijarse en la centralidad del concepto de verdad en la comprensión del lenguaje; es nuestra captación de este concepto lo que nos permite dar sentido a la cuestión de si una teoría de la verdad para un lenguaje es correcta. No hay razón para buscar un enfoque previo, o independiente, de alguna relación referencial. La otra consecuencia principal de la presente posición es que ofrece una oportunidad para decir con bastante exactitud lo que falta en una teoría de la verdad al estilo de Tarski en cuanto enfoque de la verdad.

Lo que falta es la conexión con los usuarios del lenguaje. Nada contaría como una oración, y el concepto de verdad no tendría por tanto aplicación, si no hubiera criaturas que usaran oraciones al proferir o inscribir ejemplares de ellas. Cualquier enfoque completo del concepto de verdad debe relacionarlo con el intercambio lingüístico

real. Más precisamente: la cuestión de si una teoría de la verdad es verdadera para un lenguaje dado (esto es, para un hablante o grupos de hablantes) tiene sentido sólo si las oraciones de este lenguaje tienen un significado que es independiente de la teoría (de otra forma la teoría no es una teoría en el sentido usual, sino una descripción de un lenguaje posible). O para volver a la forma definicional preferida por Tarski: si puede plantearse la cuestión de si una definición de verdad realmente define la verdad para un lenguaje dado, el lenguaje debe tener una vida independiente de la definición (de otro modo la definición es meramente estipulativa: específica, pero no es verdadera de, un lenguaje).

Si supiéramos en general lo que hace que una teoría de la verdad se aplique correctamente a un hablante o grupo de hablantes, podría plausiblemente decirse que entendemos el concepto de verdad; y si pudiéramos decir exactamente qué es lo que hace que una teoría tal sea verdadera, podríamos dar un tratamiento explícito —quizás una definición— de la verdad. La evidencia última, como opuesta a un criterio, para la corrección de una teoría de la verdad debe descansar en los hechos disponibles acerca de cómo los hablantes usan el lenguaje. Cuando digo disponibles, quiero decir públicamente disponibles —disponibles no sólo en principio, sino disponibles en la práctica para cualquiera que sea capaz de entender al hablante o hablantes del lenguaje—. Puesto que todos nosotros entendemos a algunos hablantes de algunos lenguajes, todos nosotros debemos tener evidencia adecuada para atribuir condiciones de verdad a las preferencias de algunos hablantes; todos nosotros tenemos, por tanto, una captación competente del concepto de verdad tal como se

aplica a la conducta del habla de otros.

¿Hemos sentado ahora la cuestión de si la verdad es radicalmente no epistémica, como los realistas declaran, o básicamente epistémica, como mantienen otros? Podría decirse que la cuestión se ha sentado en favor de la concepción subjetiva o epistémica, puesto que hemos seguido una línea de argumento que llega a la conclusión de que lo que decide si una teoría de la verdad para un lenguaje es verdadera es cómo se usa este lenguaje, Pero de hecho la cuestión no está sentada, porque los realistas podrían considerar que la cuestión de si la teoría es verdadera para un lenguaje o grupo de hablantes dado es de hecho empírica, pero sólo porque la cuestión de qué significan las palabras es empírica; el problema de la verdad, puede considerarse, tiene todavía que responderse bien por la teoría misma o de alguna otra manera.

¿Contiene la teoría ya la respuesta? La contiene si hay fundamento para la afirmación de que una teoría de la verdad tipo Tarski es una teoría de la correspondencia, porque entonces la teoría debe en efecto definir a la verdad como correspondencia con la realidad —la forma clásica de realismo con respecto a la verdad—. Tarski mismo dijo que quería que sus definiciones de verdad «hagan justicia a las intuiciones que apoyan a la *concepción clásica de la verdad*»; entonces cita la *Metafísica* de Aristóteles («decir de lo que es que es, o de lo que no es que no es, es verdadero»), y ofrece como una formulación alternativa

*La verdad de una oración consiste en su acuerdo (o correspondencia) con la realidad.*

(Tarski<sup>36</sup> añade que la expresión ‘teoría de la correspondencia’ ha sido sugerida por esta manera de ver las cosas.) Yo mismo he argumentado en el pasado que las teorías del tipo que Tarski enseñó a producir eran teorías de la correspondencia de un tipo<sup>37</sup>. Dije esto [174] sobre la base de que no hay manera de dar tal teoría sin emplear un concepto como el de referencia o satisfacción que relaciona expresiones con objetos en el mundo.

Me parece ahora que ha sido un error llamar a tales teorías teorías de la correspondencia. Aquí está la razón por la que creo que fue un error. La queja habitual acerca de las teorías de la correspondencia es que no tiene sentido sugerir que es posible de algún modo comparar las palabras o las creencias de uno con el mundo, puesto que el intento debe siempre concluir simplemente con la adquisición de más creencias. Esta queja fue expresada, por ejemplo, por Otto Neurath<sup>38</sup>, quien por esta razón adoptó una concepción de la verdad como coherencia;

---

<sup>36</sup> «The Semantic Conception of Truth», pp. 342–3. Tarski también habla de oraciones «que describen» «estados de cosas», *ibid.*, p. 345. Cf. «The Concept of Truth in Formalized Languages», p. 153, y «The Establishment of Scientific Semantics», p. 403.

<sup>37</sup> En «True to the Facts», en *Inquiries into Truth and Interpretation*. El argumento es éste. La verdad se define sobre la base de la satisfacción: una oración del lenguaje objeto es verdadera si se satisface por cualquier sucesión de objetos sobre los que varíen las variables de cuantificación del lenguaje objeto. Tómese ‘corresponde con’ por «satisface» y se habrá definido la verdad como correspondencia. La extrañeza de esta idea se hace evidente por la naturaleza antiintuitiva y artificial de las entidades a las que las oraciones «corresponden» y del hecho de que todas las oraciones verdaderas corresponderían a las mismas entidades.

<sup>38</sup> «Protokollsätze», *Erkenntnis*, III (1932/33): 204–214.

Carl Hempel<sup>39</sup> ha expresado la misma objeción, hablando de la «fatal confrontación de enunciados y hechos» (*ibid.*, p. 51). Rorty<sup>40</sup> ha insistido repetidamente, declarando simpatía por Dewey, en que una concepción de la verdad como correspondencia hace inútil al concepto de verdad. Yo he dicho más o menos lo mismo<sup>41</sup>.

Esta queja contra las teorías de la correspondencia no es correcta. Una razón por la que no es correcta es que depende de asumir que alguna forma de teoría epistémica es correcta; por tanto, sólo sería una queja legítima si la verdad fuera un concepto epistémico. Si ésta fuera la única razón para rechazar las teorías de la correspondencia, el realista podría simplemente replicar que su posición no ha [175] sido tocada; él siempre mantuvo que la verdad era independiente de nuestras creencias o de nuestra habilidad para aprender la verdad.

---

<sup>39</sup> «On the Logical Positivist theory of Truth», *Analysis*, II (1935): 49–59 (ed. cast.: «La teoría de la verdad de los positivistas lógicos», en este mismo volumen).

<sup>40</sup> *Consequences of Pragmatism*, «Introduction»; también en «Pragmatism, Davidson and Truth», en Ernest Lepore, ed., *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (Nueva York: Blackwell, 1986).

<sup>41</sup> La posición que tomo en el presente artículo estuvo influida por un intercambio entre Rorty y yo en la reunión de 1982 de la *Pacific Division of the American Philosophical Association*. Rorty me persuadió de que no llamara a mi posición ni una teoría de la correspondencia ni una teoría de la coherencia; creo que yo le persuadí a él de abandonar la teoría pragmática de la verdad. «Pragmatism, Davidson and Truth» es una versión revisada de la conferencia de Rorty de 1982 en la reunión de la Pacific Division.

Para un ejemplo de uso de ‘correspondencia’ que ahora deploro, véase mi «A Coherence Theory of Truth and Knowledge», en *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*.

La objeción real a las teorías de la correspondencia es más simple; es que no hay nada interesante o instructivo a lo que las oraciones verdaderas pudieran corresponder. Este punto fue señalado hace algún tiempo por C. I. Lewis<sup>42</sup>; él desafió a los teóricos de la correspondencia a *localizar* el hecho o parte de la realidad, o del mundo, al que una oración verdadera correspondía. Uno puede localizar objetos individuales, si sucede que la oración los nombra o describe, pero incluso tal localización sólo tiene sentido relativamente a un marco de referencia, y así presumiblemente el marco de referencia debe estar incluido en lo que quiera que sea aquello a lo que la oración verdadera corresponde. Perseguir esta línea de pensamiento llevó a Lewis a concluir que, si las oraciones verdaderas no corresponden a nada en absoluto, debe ser al universo como un todo; así, todas las oraciones verdaderas corresponden a la misma cosa. Frege, como sabemos, alcanzó la misma conclusión a través de una línea de razonamiento de algún modo similar. El argumento de Frege, si Alonzo Church<sup>43</sup> tiene razón, puede formalizarse: empezando con las asunciones de que una oración verdadera no puede hacerse corresponder a algo diferente mediante la sustitución de los términos singulares correferenciales, o mediante la sustitución de oraciones lógicamente equivalentes, es fácil mostrar que, si las oraciones verdaderas corresponden a algo, todas ellas corresponden a la misma cosa. Pero esto es trivializar el concepto de correspondencia

---

<sup>42</sup> *An Analysis of Knowledge and Valuation*, La Salle, YL: Open Court, 1946, pp. 50–55.

<sup>43</sup> El argumento, atribuido a Frege por Church, puede encontrarse en Church: *Introduction to Mathematical Logic*, Vol. 1, Princeton University Press, 1956, pp. 24–25. El argumento de Frege se ensaya en mi «True to the Facts».

completamente; la relación de correspondencia no tiene ningún interés si sólo hay una cosa a la que corresponder, puesto que, como en cualquier caso de este tipo, la relación podría colapsar también en una propiedad simple: así, ‘o corresponde al universo’, como ‘o corresponde a (o nombra) lo Verdadero’, o ‘o corresponde a los hechos’ puede leerse de manera menos desorientadora como ‘o es verdadera’. Peter Strawson<sup>44</sup> ha observado que las partes de una oración podrían corresponder a las partes del mundo (esto es, referir a ellas), pero añade, [176]

Es evidente que no hay nada más en el mundo con lo que el enunciado mismo pueda relacionarse. [...] Y es evidente que la demanda de que haya tal *relatum* es lógicamente absurda. [...] Pero la demanda de algo en el mundo *que haga verdadero al enunciado* [...], o *a lo que el enunciado corresponda si es verdadero*, es exactamente esta demanda [*ibid.*, pp. 194-95].

Continúa él afirmando correctamente que, «mientras que decimos ciertamente que un enunciado corresponde a (encaja, está soportado por, concuerda con) los hechos», esto es meramente «una variante de decir que es verdadero» (*ibid.*).

La objeción correcta a las teorías de la correspondencia no es, entonces, que hagan de la verdad algo a lo que los humanos no pueden nunca aspirar legítimamente; la objeción real es más bien que tales teorías no proporcionan las entidades a las cuales los vehículos de verdad (tanto si consideramos que éstos son enunciados, oraciones o

---

<sup>44</sup> «Truth», en *Logico-Linguistic Papers*, Londres: Methuen, 1971.

preferencias) puede decirse que correspondan. Si esto es correcto, y estoy convencido de que lo es, deberíamos cuestionar también la asunción popular de que las oraciones, o sus ejemplares hablados, o las entidades o configuraciones en nuestros cerebros del tipo de las oraciones, puedan llamarse propiamente «representaciones», puesto que no hay nada que ellas representen. Si abandonamos los hechos como entidades que hacen a las oraciones verdaderas, deberíamos abandonar las representaciones al mismo tiempo, porque la legitimidad de cada una de ellas depende de la legitimidad de la otra.

Hay así una razón seria para sentir el haber dicho que una teoría de la verdad al estilo de Tarski era una forma de la teoría de la correspondencia. Mi razón básica para decirlo no era que hubiera cometido el error de suponer que las oraciones o preferencias de oraciones correspondiesen a algo en un sentido interesante. Sino que yo estaba todavía bajo la influencia de la idea de que hay algo importante en la concepción realista de la verdad; la idea de que la verdad, y por tanto la realidad, son (excepto en casos especiales) independientes de lo que cualquiera crea o pueda conocer. Así, yo promocioné mi concepción como un tipo de realismo, realismo con respecto al «mundo exterior», con respecto al significado, y con respecto a la verdad<sup>45</sup>.

Los términos ‘realismo’ y ‘correspondencia’ estaban mal elegidos porque sugerían el apoyo positivo a una posición, o a una asunción de que hay una tesis positiva clara que adoptar, mientras que [177] todo lo que yo estaba justificado para mantener, y todo lo que mi posición realmente implicaba con respecto al realismo y la verdad, era la

---

<sup>45</sup> «A Coherence Theory of Truth and Knowledge», p. 307.



concepción negativa de que las concepciones epistémicas eran falsas. La concepción realista de la verdad, si tiene algún contenido, debe basarse en la idea de la correspondencia, correspondencia tal como se aplica a oraciones o a creencias o a preferencias —entidades que tienen un carácter preposicional—; y tal correspondencia no puede hacerse inteligible. Yo simplemente cometí el error de asumir que el realismo y las teorías epistémicas eran las únicas posiciones posibles. La única razón legítima que tenía para llamar a mi posición una forma de realismo era rechazar posiciones como el antirrealismo de Dummett; estaba preocupado en rechazar la doctrina de que la realidad o la verdad dependían directamente de nuestras capacidades epistémicas. Hay una finalidad en este rechazo. Pero es inútil tanto rechazar como aceptar el *slogan* de que lo real y lo verdadero son «independientes de nuestras creencias». El único sentido positivo evidente que esta expresión puede tener, el único uso que se ajusta a las intenciones de aquellos que la valoran, deriva de la idea de la correspondencia, y ésta es una idea sin contenido<sup>46</sup>.

Rechazar la doctrina de que lo real y lo verdadero son independientes de nuestras creencias, no es, por supuesto, rechazar la perogrullada de lo que equivocadamente puede pensarse que expresa: creer algo no es en general hacerlo verdadero. Porque aceptar que la perogrullada es verdadera no nos compromete a decir que no hay ninguna

---

<sup>46</sup> Arthur Fine rechazó el realismo por algunas de las mismas razones que yo, y añadió una refutación espléndida de la tesis de que una concepción realista de la verdad explica la práctica y el avance de la ciencia: «The Natural Ontological Attitude», en *The Shaky Game: Einstein, Realism and the Quantum Theory*, Chicago: University Press, 1986.

conexión en absoluto entre la creencia y la verdad; debe haber alguna conexión si tenemos que relacionar la verdad de las preferencias con su uso. La cuestión es qué puede ser esta conexión.

Distintas formas de subjetivismo —esto es, de posiciones que construyen la verdad a partir de un concepto epistémico— conectan los pensamientos, deseos, e intenciones humanas con la verdad de maneras bastante diferentes, y no puedo fingir haber hecho justicia a todas esas concepciones aquí. Lo mejor que puedo hacer es indicar por qué, a pesar de las diferencias entre las distintas posiciones, tiene sentido estar insatisfecho con todas ellas. [178]

He clasificado a las teorías de la verdad como coherencia como epistémicas, y esto necesita una explicación. Una teoría pura de la verdad como coherencia debería mantener, supongo, que todas las oraciones de un conjunto consistente de oraciones son verdaderas. Quizás nadie ha mantenido nunca una teoría de tal tipo, porque es una locura. Aquellos que han propuesto teorías de la coherencia, por ejemplo, Neurath y Rudolf Carnap (en un tiempo), han dejado claro habitualmente que eran conjuntos de *creencias*, o de oraciones tenidas por verdaderas, cuya consistencia era suficiente para hacerlas verdaderas; por eso clasifico a las teorías de la coherencia con las concepciones epistémicas: ligan la verdad directamente con lo que se cree. Pero a menos que se añada algo más, esta concepción parece tan equivocada como Moritz Schlick<sup>47</sup> mantuvo que era (la llamó un «error asombroso»); la objeción obvia es que son posibles muchos conjuntos consis-

---

<sup>47</sup> «Über das Fundament der Erkenntnis», *Erkenntnis*, IV (1934): 79–99.

tentes diferentes de creencias que no son consistentes entre sí<sup>48</sup>.

Hay teorías, similares en ciertos aspectos a la teoría de la coherencia, que tienen más o menos el mismo problema. Quine mantiene que la verdad de algunas oraciones, a las que llama oraciones observacionales, está directamente ligada a la experiencia (más precisamente, a patrones de terminaciones nerviosas excitadas); otras oraciones derivan su contenido empírico de sus conexiones con las oraciones observacionales y sus mutuas relaciones lógicas. La verdad de la teoría resultante depende sólo de en qué medida sirva para explicar o predecir oraciones observacionales verdaderas. Quine plausiblemente mantiene que podría haber dos teorías igualmente capaces de dar cuenta de todas las oraciones observacionales verdaderas, y sin embargo que ninguna de las teorías pueda ser reducida a la [179] otra (cada teoría contiene al menos un predicado que no puede definirse usando los recursos de la otra teoría). Quine ha mantenido en diferentes momentos diferentes formas de pensar esta situación. De acuerdo con una de ellas, ambas teorías son verdaderas. No veo ninguna razón para

---

<sup>48</sup> No toda teoría que relaciona la verdad con conjuntos consistentes de creencias está equivocada. Lo que debe añadirse a las teorías estándar de la coherencia es una apreciación no sólo de cómo se relacionan causalmente y lógicamente las creencias entre sí, sino de cómo dependen los contenidos de una creencia de su conexión causal con el mundo. Discuto estos asuntos en la sección siguiente. Véase también mi «A Coherence Theory of Truth and Knowledge» y «Empirical Content», en *Truth and interpretation: Perspectiva on the Philosophy of Donald Davidson*. Ahora me parece que es otro error terminológico más haber llamado a la tesis de «A Coherence Theory» una teoría de la coherencia. Explico por qué con más amplitud en «Afterthoughts, 1987», añadido a la reimpresión de «A Coherence Theory» que aparecerá en A. Malichowski, ed., *Reading Rorty*, Nueva York: Blackwell, 1990, pp. 136–8.

objetar el punto de vista de que teorías empíricamente equivalentes (como quiera que se caracterice el contenido empírico) sean verdaderas o falsas a la vez. De acuerdo con otro punto de vista de Quine, un hablante o un pensador en un momento dado opera con una teoría y, para él en ese momento, la teoría que está usando es verdadera y la otra teoría falsa. Si cambia a la teoría alternativa, entonces *ésta* se convierte en verdadera y la teoría previamente aceptada en falsa. La posición puede ilustrar lo que Quine quiere decir cuando dice que la verdad es «inmanente»<sup>49</sup>. Esta concepción de la inmanencia o relatividad de la verdad no debe confundirse con el sentido pedestre en que la verdad de una oración es relativa al lenguaje en el que aparece. Las dos teorías de Quine pueden pertenecer a, y estar expresadas en, el mismo lenguaje; de hecho, deben estarlo si hemos de entender la afirmación de que las teorías están en conflicto. No es fácil ver cómo la misma oración (sin elementos deícticos), sin cambiar la interpretación, puede ser verdadera para una persona y no para otra, o para una persona dada en un momento y no en otro. La dificultad parece debida al intento de importar consideraciones epistemológicas al concepto de verdad.

El «realismo interno» de Putnam también hace a la verdad inmanente, aunque no, como ocurre en la concepción de Quine, relativa a

---

<sup>49</sup> Véase *Ontological Relativity and Other Essays* (Nueva York: Columbia, 1969). Para el problema de Quine acerca de las teorías empíricamente equivalentes y mutuamente irreductibles véase su «On Empirically Equivalent Systems of the World», *Erkenntnis*, IX (1975): 313–328; *Theories and Things*, Cambridge: Harvard, 1981, pp. 29–30; L. E. Hahn y P. A. Schilpp, eds., *The Philosophy of W. V. Quine*, La Salle, IL: Open Court, 1986, pp. 156–7.

una teoría, sino al lenguaje y al esquema conceptual completos que una persona acepta. Por supuesto si todo esto significa que la verdad de las oraciones o preferencias es relativa a un lenguaje, esto resulta familiar y trivialmente correcto. Pero Putnam parece tener algo más en mente —por ejemplo, que una oración tuya y una oración mía pueden contradecirse entre sí, y sin embargo cada una ser verdadera «para el hablante»—. Es difícil pensar en qué lenguaje puede expresarse esta posición coherentemente, no digamos ya persuasivamente. [180]

La fuente del problema es de nuevo la necesidad que sentimos de hacer a la verdad accesible. Putnam tiene claro que ésta es la consideración que le concierne. Identifica explícitamente la verdad con la asertabilidad justificada idealizada. Llama a esto una forma de realismo porque hay «una cuestión acerca de cuál sería el veredicto si las condiciones fueran suficientemente buenas, un veredicto acerca de qué opinión ‘convergería’ si fuéramos razonables»<sup>50</sup>. Añade que su punto de vista es «un tipo de realismo *humano*, una creencia de, que hay una cuestión acerca de lo que es correctamente asertable por nosotros, como algo opuesto a lo que es correctamente asertable desde la perspectiva del ojo de Dios tan querida por el realismo metafísico clásico» (*ibid.*). Uno sospecha que, si las condiciones bajo las cuales alguien está justificado idealmente para asertar algo se explicaran completamente, se haría patente que tales condiciones o bien permiten la posibilidad de error o que son tan ideales que hacen inútiles las pretendidas conexiones con las habilidades humanas. También es

---

<sup>50</sup> *Realism and Reason: Philosophical Papers*, Vol. 3, Nueva York: Cambridge, 1983, p. XVIII.

sorprendente que Putnam parece no tener ningún argumento a favor de su posición excepto que la alternativa («el realismo metafísico» —esto es, una teoría de la correspondencia—) es inaceptable. Él no argumenta que no puede haber otra posición.

Putnam describe su posición cercana a la de Dummett en el punto central —el status epistemológico de la verdad—. Una diferencia es que Putnam está menos seguro que Dummett de que la verdad está limitada a lo que es definitivamente determinable, y por tanto está menos seguro de que el principio de bivalencia deba abandonarse; esto explica quizás por qué Putnam llama a su concepción una forma de realismo mientras que Dummett llama antirrealista a su posición. Putnam piensa también que se diferencia de Dummett en que liga la verdad a la asertabilidad justificada *idealizada* en vez de a la asertabilidad justificada; pero aquí creo que una lectura cuidadosa de Dummett mostraría que él tiene más o menos la misma idea. Si Dummett no insiste en algo similar a las condiciones ideales de Putnam, creo entonces que se aplica una crítica a Dummett que Putnam formuló una vez: si la verdad depende de la asertabilidad justificada, la verdad puede «perderse», esto es, una oración puede ser verdadera para una persona en un momento y más tarde convertirse en falsa porque cambien las condiciones de justificación. Esto debe estar [181] equivocado<sup>51</sup>. Dummett dice que está de acuerdo en que la verdad no puede perderse, pero fracasa al dar una idea clara de cómo la asertabilidad garantizada puede ser al mismo tiempo una propiedad fija y una

---

<sup>51</sup> Putnam, «Reference and Understanding» y «Reply to Dummett's Comment», en A. Margalit, ed., *Meaning and Use*, Dordrecht: Reidel, 1979, pp. 226–8.

propiedad que depende de la capacidad real de los hablantes humanos para reconocer que se satisfacen ciertas condiciones. Las capacidades reales aumentan y disminuyen, y difieren de persona a persona; la verdad no.

¿Por qué sostiene Dummett esta concepción de la verdad? Hay muchas razones, pero una parece ser esta. Hemos visto que una teoría de la verdad al estilo de Tarski ni define la verdad ni la caracteriza completamente; no hay forma de decir si la teoría se aplica a un hablante o grupo de hablantes a menos que se añada algo que relacione a la teoría con los usos humanos del lenguaje. Dummett piensa que la única manera de hacer esto es hacer a la verdad humanamente reconocible. El uso humano del lenguaje debe ser una función de cómo entiende la gente el lenguaje, así si la verdad tiene que jugar algún papel en la explicación de qué es entender un lenguaje, debe haber algo, piensa Dummett, que cuente como el que una persona tenga «evidencia concluyente» de que un enunciado es verdadero. Uno puede apreciar la fuerza de esta idea y encontrarla al mismo tiempo difícil de aceptar. He dado mi razón principal para rechazarla; que o bien es vacía o hace de la verdad una propiedad que puede perderse. Pero es importante darse cuenta que hay otras intuiciones fuertes que tendrían que sacrificarse si Dummett tuviera razón. Una es la conexión de la verdad con el significado: en la concepción de Dummett, podemos entender una oración como ‘Nunca se construirá una ciudad en este lugar’ sin tener idea de qué haría a esta oración verdadera (puesto que la oración, o una preferencia de ella, no tiene valor de verdad para Dummett). Otra es la conexión de la verdad con la creencia: en la concepción de Dummett, puedo entender y creer que nunca se cons-

truirá una ciudad en este lugar, pero mi creencia no tendrá ningún valor de verdad. Parecería que, para Dummett, tener una creencia que uno expresa mediante una oración dada no es necesariamente creer que la oración es verdadera.

Estaría tentado a seguir con Dummett si pensara que debemos elegir entre lo que Putnam llama realismo trascendental, esto es, la concepción de que la verdad es «radicalmente no epistémica», que [182] todas nuestras teorías y creencias mejor investigadas y establecidas podrían ser falsas, y la identificación de Dummett de la verdad con la asertabilidad garantizada, puesto que encuentro a la primera concepción —esencialmente la concepción de la correspondencia— incomprensible, mientras que encuentro a la concepción de Dummett meramente falsa. Pero no veo razón para suponer que realismo y antirrealismo, explicados en los términos del carácter radicalmente no epistémico o radicalmente epistémico de la verdad, sean las únicas maneras de dar fundamento a una teoría de la verdad o del significado.

Recapitulemos brevemente. En la primera sección de este artículo, rechacé las concepciones deflacionistas de la verdad, aquellas que enseñan que no hay nada más en el concepto de lo que Tarski ha mostrado cómo definir para lenguajes particulares. En esta sección, he argumentado que ciertos intentos familiares de caracterizar la verdad que van más allá de dar contenido empírico a una estructura del tipo de las que Tarski nos enseñó a describir son vacíos, falsos, o confusos. No deberíamos decir que la verdad es correspondencia, coherencia, asertabilidad garantizada, asertabilidad justificada idealmente, lo que es aceptado en la conversación de la gente adecuada, lo que la ciencia



acabará manteniendo, lo que explica la convergencia hacia teorías simples en la ciencia, o el éxito de nuestras creencias comunes. En la medida en que realismo y antirrealismo dependen de una u otra de estas concepciones de la verdad deberíamos rechazar el sostener ninguno de los dos. El realismo, con su insistencia en la correspondencia radicalmente no epistémica, pide más a la verdad de lo que podemos entender; el antirrealismo, con su limitación de la verdad a lo que puede determinarse, priva a la verdad de su papel de *standard* intersubjetivo. Debemos encontrar otra manera de considerar el asunto.

### III. LOS CONTENIDOS DE LA VERDAD

Una teoría de la verdad, en contraste con una definición estipulativa de la verdad, es una teoría empírica acerca de las condiciones de verdad de todas las oraciones de algún corpus de oraciones. Pero, por supuesto, las oraciones son objetos abstractos, formas, digamos, y no tienen condiciones de verdad excepto cuando hablantes y garabateadores los encaman en sonidos y garabatos. Al final, una teoría de la verdad debe tratar con proferencias y escrituras de los usuarios [183] del lenguaje; el papel de las oraciones en una teoría es meramente hacer posible el tratar con *tipos* de proferencias e inscripciones, tanto si estos tipos particulares se realizan como si no. Introducir oraciones sirve así para dos propósitos: nos permite hablar de todas las proferencias e inscripciones reales del mismo tipo de una vez; y nos permite estipular cuáles serían las condiciones de verdad de una preferencia o inscripción de un tipo dado en el caso de que fuera proferida. (Por

cuestiones de brevedad, a partir de ahora me referiré a los actos de escribir como preferencias de la misma manera que a sus contrapartidas audibles.)

Aunque a veces podemos decir que un grupo habla con una sola voz, las preferencias son esencialmente personales; cada preferencia tiene su agente y su tiempo. Una preferencia es un suceso de un tipo especial, una acción intencional. Las teorías de la verdad se ocupan en primer lugar de las preferencias oracionales, preferencias que, cualquiera que sea su gramática superficial, deben tratarse como preferencias de oraciones. La primacía de las oraciones o de las preferencias oracionales la dicta el hecho de que la teoría ofrece condiciones de verdad para, y la verdad se predica de, oraciones, en cuanto proferidas en ocasiones particulares por hablantes particulares. A parte de las condiciones verbales de éxito, no hay razón para no llamar a la preferencia de una oración, bajo las condiciones que hacen verdadera a la oración, una preferencia verdadera.

Una teoría de la verdad hace más que describir un aspecto de la conducta hablada de un agente, porque no sólo da las condiciones de verdad de las preferencias reales del agente; también especifica las condiciones bajo las que la preferencia de una oración sería verdadera si fuera proferida. Esto se aplica tanto a las oraciones proferidas realmente, diciéndonos lo que hubiera sido el caso si aquellas oraciones hubieran sido proferidas en otros momentos o bajo otras circunstancias, como a oraciones no proferidas nunca. La teoría describe así una cierta habilidad compleja.

Una preferencia tiene ciertas condiciones de verdad sólo si el

hablante pretende que sea interpretada como teniendo aquellas condiciones de verdad. Consideraciones morales, sociales, o legales pueden a veces invitarnos a negar esto, pero no creo que las razones para tales excepciones revelen nada importante acerca de lo que es básico para la comunicación. Alguien podría decir algo que fuera normalmente ofensivo o insultante en un lenguaje que cree que sus oyentes no entienden; pero en este caso su audiencia para el propósito de interpretación es sólo, obviamente, el hablante mismo. Un malapropismo o desliz de lengua, si significa algo, significa lo que el que lo promulga pretende que signifique. A algunos les gusta mantener que los significados de las palabras son mágicamente independientes de las intenciones del hablante; por ejemplo, que dependen de cómo habla la mayoría de, o los mejor informados, o los mejor nacidos de la comunidad en la que el hablante vive, o quizás de cómo hablarían si tuvieran el suficiente cuidado<sup>52</sup>. Esta doctrina implica que un hablante puede ser perfectamente inteligible para sus oyentes, puede ser interpretado exactamente como él pretende que se le interprete, y sin embargo puede no saber lo que quiere decir mediante lo que dice. Creo que esta concepción, aunque ha sido ingeniosamente elaborada y defendida no revela ningún interés filosóficamente serio acerca de la naturaleza de la

---

<sup>52</sup> Saul Kripke atribuye una concepción de este tipo a Wittgenstein en *Wittgenstein on Rules and Private Language*, Nueva York: Blackwell, 1982, y la acepta tentativamente. Para una versión diferente, véanse los numerosos trabajos de Tyler Burge sobre antiindividualismo, por ejemplo, «Individualism and the Mental», en P. French, T. Uehling, H. Wettstein, eds., *Midwest Studies in Philosophy*, volumen 4, Minneápolis: Minnesota UP, 1979, pp. 73–121; «Individualism and Psychology», *Philosophical Review*, XCV (1986): 3–46; «Wherein is Lenguaje Social?» en A. George, ed., *Reflections on Chomsky*, Nueva York: Blackweil, 1989, pp. 176–191.

verdad o del significado (aunque puede tener mucho que ver con maneras buenas o aceptables, y puede representar una intención, o incluso algún tipo de responsabilidad social por parte de algunos hablantes)<sup>53</sup>. Para el propósito de la empresa presente, el de entender la verdad y el significado, deberíamos, creo, mantenemos tan cerca como sea posible de lo que el hablante pone directamente a disposición de la audiencia, y éste es el estado relevante de la mente del hablante. Lo que importa para la comunicación lingüística con éxito es la intención del hablante de ser interpretado de una cierta manera, por una parte, y la interpretación real de las palabras del hablante en las líneas pretendidas a través del reconocimiento del intérprete de las intenciones del hablante, por otra<sup>54</sup>. [185]

La aproximación que estoy siguiendo no coloca ningún peso especial en el concepto de un lenguaje en cuanto algo compartido por hablante e intérprete, o por un hablante y su comunidad de habla,

---

<sup>53</sup> Véase mi «Knowing One's Own Mind», *Proceedings and Addresses of the American Philosophical Association*, LX (1987): 441–458.

<sup>54</sup> La influencia de H. P. Grice: «Meaning», *The Philosophical Review*, LXVI (1957): 377–388, será evidente aquí. Mi caracterización de la comunicación con éxito deja abierto un rango de posibilidades acerca de la cuestión de qué quiere decir un hablante mediante sus palabras en una ocasión. Puesto que el hablante debe pretender ser interpretado de una cierta manera, debe creer que su audiencia está equipada para interpretar sus palabras de esta manera. Pero ¿cómo de justificada tiene que ser esta creencia y cuan aproximadamente correcta? No creo que nuestros estándares para decidir lo que significan las palabras de alguien, en cuanto habladas en una ocasión dada, sean lo suficientemente firmes como para permitimos trazar una línea nítida entre una intención fallida de que las palabras de uno tengan un cierto significado y un éxito en el significado acompañado por una intención fallida de ser interpretado como se pretendía.

excepto en este sentido: aunque la comunicación mediante el habla no requiere, hasta donde puedo ver, que ninguno de los dos hablantes *hable* de la misma manera, sí exige, por supuesto, una coincidencia entre cómo los hablantes pretenden que se les interprete y cómo los entienden sus intérpretes. Esta exigencia tiende sin duda a animar a la convergencia en la conducta hablada entre aquellos que intercambian palabras, dependiendo del grado de factores como el *status* económico y social compartido, los antecedentes étnicos y educacionales, etc. Que la convergencia exista es de tan amplia importancia práctica que podríamos exagerar tanto su grado como su significación filosófica. Pero creo que hacemos bien en ignorar esta cuestión práctica al construir teorías del significado, de la verdad, y de la comunicación lingüística<sup>55</sup>. Trataré, por tanto, a las teorías de la verdad como si se aplicaran en primer lugar a hablantes individuales en distintos períodos o incluso momentos de sus vidas.

Una teoría de la verdad liga al hablante con el intérprete: describe a la vez las habilidades y prácticas lingüísticas del hablante y da contenido a lo que el intérprete erudito conoce que le permite captar el significado de las preferencias del hablante. Esto no es decir que o el hablante o el intérprete sean conscientes o tengan conocimiento proposicional de los contenidos de una teoría tal. La teoría describe las condiciones bajo las cuales una preferencia de un hablante es verdadera, y así no *dice* nada directamente acerca de lo que el hablante sabe. La teoría, sin embargo, *implica* algo acerca del contenido proposicional de

---

<sup>55</sup> Véase mi «Communication and Convention», en *Inquiries into Truth and Interpretation*

ciertas intenciones del hablante, a saber, las intenciones de que sus preferencias se interpreten de una cierta manera. Y aunque ciertamente el intérprete no necesita tener conocimiento explícito de la teoría, la teoría proporciona la única manera de especificar la infinidad de cosas que el intérprete sabe acerca del hablante, a saber, las condiciones bajo las cuales cualquiera de entre [186] un número indefinidamente amplio de oraciones del hablante sería verdadera si fuera proferida.

Debe haber por supuesto *algún* sentido en que hablante e intérprete hayan internalizado una teoría; pero esto no es más que el hecho de que el hablante es capaz de hablar como si creyera que el intérprete lo interpretaría de la manera en que la teoría describe, y el hecho de que el intérprete está preparado para interpretarlo así. Todo lo que necesitaríamos de una teoría de la verdad para un hablante es que sea tal que, *si* un intérprete tuviera conocimiento proposicional explícito de la teoría, sabría las condiciones de verdad de las preferencias del hablante<sup>56</sup>.

---

<sup>56</sup> Esto es, por supuesto, mucho más de lo que ofrece cualquier teoría que nadie haya sido capaz de ofrecer para ningún lenguaje natural. La condición no es, por tanto, una que sabemos que puede satisfacerse. Sabemos, por otra parte, cómo producir una teoría tal para un fragmento poderoso, quizá autosuficiente, del inglés y de otros lenguajes naturales, y esto es suficiente para dar contenido a la idea de que la incorporación del concepto de verdad a una teoría ofrece una intuición acerca de la naturaleza del concepto. Podríamos tener que conformarnos al final con un sentido mucho menos preciso de 'teoría' de los que Tarski tenía en la mente.

Me estoy saltando un grupo de problemas bien trabajado, tales como proporcionar las condiciones de verdad de los condicionales subjuntivos, de los imperativos, interrogativas, enunciados éticos, etc. He discutido (aunque ciertamente no solucionado) la mayoría de estos problemas en otra parte.

Una teoría de la verdad para un hablante es una teoría del significado en este sentido, que el conocimiento explícito de la teoría bastaría para entender las preferencias de este hablante. Consigue esto al describir el núcleo crítico de la conducta lingüística real y potencial del hablante, en efecto, cómo el hablante pretende que sus preferencias sean interpretadas. El tipo de comprensión involucrada se restringe a lo que podríamos también llamar el significado literal de las palabras, mediante lo cual quiero decir, más o menos, el significado que el hablante pretende que el intérprete capte, cualquiera que sea la significación o la fuerza posterior que el hablante quiera que el intérprete desentrañe<sup>57</sup>. [187]

La tesis de que una teoría de las condiciones de verdad ofrece un enfoque adecuado de lo que se necesita para entender los significados literales de las preferencias está, por supuesto, muy discutida, pero puesto que he argumentado a favor de ella ampliamente en otro sitio, trataré en su mayor parte la tesis aquí como una asunción. Si la asun-

---

<sup>57</sup> Hay una intención no tocada por una teoría de la verdad que un hablante debe pretender que un intérprete perciba, *la fuerza* de la preferencia. Un intérprete debe, si es que entiende al hablante, ser capaz de decir si una preferencia pretende ser un chiste, una aserción, una orden, una pregunta, y así sucesivamente. No creo que haya reglas o convenciones que gobiernen este aspecto esencial del lenguaje. Es algo que los usuarios del lenguaje pueden transmitir a los oyentes y que los oyentes pueden detectar suficientemente a menudo; pero esto no muestra que estas habilidades puedan regimentarse. Creo que hay razones fundadas para pensar que no es posible nada como una teoría seria concerniente a esta dimensión del lenguaje. Todavía menos hay convenciones o reglas para crear o entender metáforas, ironía, humor, etc. Véase mi «What Metaphors Mean?» y «Convention and Communication», en *Inquiries into Truth and Interpretation*.

ción está equivocada, muchos de los detalles a los que voy a descender acerca de la aplicación del concepto de verdad se verán amenazados, pero el enfoque general, creo, permanecerá válido.

Una teoría de la verdad, considerada como una teoría empírica, se contrasta por sus consecuencias relevantes, y éstas son las oraciones-T implicadas por la teoría. Una oración-T dice de un hablante particular que, en cualquier momento que él profiera la oración dada, la preferencia será verdadera si y sólo si se satisfacen ciertas condiciones. Así las oraciones-T tienen la forma y la función de leyes naturales; son bicondicionales universalmente cuantificados, y como tales se entiende que se aplican contrafácticamente y que se confirman mediante sus instancias<sup>58</sup>. Así, una teoría de la verdad es una teoría para describir, explicar, entender, y predecir un aspecto básico de la conducta verbal. Puesto que el concepto de verdad es central a la teoría, tenemos justificación para decir que la verdad es un concepto explicativo de importancia crucial.

La cuestión que queda es: ¿cómo confirmamos la verdad de una oración-T? La cuestión es un tipo de cuestión que se plantea con respecto a muchas teorías, tanto en las ciencias físicas como en psicología. Una teoría de la medida fundamental del peso, por ejemplo, afirma en forma axiomática las propiedades de la relación entre  $x$  e  $y$

---

<sup>58</sup> Esto de alguna manera responde a una crítica frecuente a las teorías de la verdad como teorías del significado. Por ejemplo, dado el caso (inusual) de dos predicados no estructurados con la misma extensión, una teoría de la verdad podría hacer una distinción si hubiera circunstancias que nunca se dan pero bajo las cuales las condiciones de verdad fueran diferentes.



que se dan cuando  $x$  es al menos tan pesado como  $y$ ; esta relación debe, entre otras cosas, ser transitiva, reflexiva, y no-simétrica. Una teoría de la preferencia podría estipular que la relación de preferencia débil tiene las mismas propiedades formales. Pero en ninguno de los dos casos los axiomas definen la relación central ( $x$  es al menos [188] tan pesado como  $y$ ,  $x$  es débilmente preferido a  $y$ ), ni nos instruyen en cómo determinar cuándo se da la relación. Antes de que la teoría pueda contrastarse o usarse, debe decirse algo acerca de la interpretación de los conceptos no definidos. Lo mismo se aplica al concepto de verdad<sup>59</sup>.

Es un error buscar una definición conductista, o cualquier otra clase de definición explícita o reducción completa del concepto de verdad. La verdad es uno de los conceptos más básicos y claros que tenemos, así es inútil soñar en eliminarlo en favor de algo más simple o más fundamental. Nuestro procedimiento es más bien éste: nos hemos preguntado cuáles son las propiedades formales del concepto cuando se aplica a estructuras relativamente bien comprendidas, a saber, a lenguajes. Aquí el trabajo de Tarski ofrece la inspiración. Queda por indicar cómo una teoría de la verdad puede aplicarse a hablantes o grupos de hablantes particulares. Dada la complejidad de las estructuras a las que el concepto de verdad ayuda a caracterizar, trozos compa-

---

<sup>59</sup> Expliqué en la sección previa por qué creo que no debemos preocuparnos separadamente acerca de la referencia o la satisfacción. Dicho brevemente, la razón es que las oraciones-T no contienen conceptos referenciales. Puesto que las implicaciones contrastables de la teoría son oraciones-T en cuanto aplicadas a casos, cualquier manera de caracterizar la satisfacción que ofrezca oraciones-T confirmables era tan buena como cualquier otra.

rativamente anémicos de evidencia, aplicados a una infinidad potencial de puntos, pueden ofrecer resultados ricos e instructivos. Pero no puede esperarse la formalización completa de la relación entre la evidencia para la teoría y la teoría misma.

Lo que *deberíamos* exigir, sin embargo, es que la evidencia para la teoría sea en principio accesible públicamente, y esto no es asumir de antemano los conceptos que tienen que ilustrarse. El requisito de que la evidencia sea públicamente accesible no se debe a una añoranza atávica de fundamentaciones conductistas o verificacionistas, sino al hecho de que lo que hay que explicar es un fenómeno social. Los fenómenos mentales en general podrían ser privados o no, pero la interpretación correcta del habla de una persona por otra debe *en principio* ser posible. La intención de un hablante de que sus palabras se entiendan de una cierta manera podrían por supuesto permanecer opaca para los oyentes más capacitados y eruditos, pero lo que tiene que ver con la interpretación correcta, con el significado, y con las condiciones de verdad tiene que basarse necesariamente en evidencia disponible. Como Ludwig Wittgenstein, por no mencionar a [189] Dewey, G. H. Mead, Quine y muchos otros han destacado, el lenguaje es intrínsecamente social. Esto no implica que la verdad y el significado *puedan definirse* en términos de conducta observable o que no sea «nada más que» conducta observable; pero si implica que el significado se determina completamente mediante conducta observable, incluso mediante conducta fácilmente observable. Que los significados sean descifrables no es una cuestión de suerte; la disponibilidad pública es un aspecto constitutivo del lenguaje.

Los conceptos usados para expresar la evidencia no deben cometer petición de principio; deben ser suficientemente remotos de lo que la teoría produce en último extremo. Esta conclusión final no es más que lo que pedimos de cualquier análisis revelador, pero es difícil, al menos en este caso, satisfacerla. Cualquier intento de entender la comunicación verbal debe considerarla en su lecho natural como parte de una empresa más amplia. Al principio parece que esto no puede ser difícil, no teniendo el lenguaje más que transacciones públicas entre hablantes e intérpretes, y las aptitudes para tales transacciones. Sin embargo la tarea nos elude. Porque el hecho de que los fenómenos lingüísticos no sean más que fenómenos conductuales, biológicos, o físicos descritos en un vocabulario exótico de significado, referencia, verdad, aserción, y así sucesivamente —la mera superveniencia de esta clase de un tipo de hecho o descripción sobre otro— no garantiza, o ni siquiera alarga la promesa de la posibilidad de la reducción conceptual.

Aquí descansa nuestro problema. Ahora bosquejaré lo que creo que es al menos la *clase* correcta de solución. El entorno psicológico inmediato de los logros y aptitudes lingüísticos tiene que encontrarse en las actitudes, estados, y eventos que se describen en expresiones intensionales: acción intencional, deseos, creencias, y sus parientes próximos tales como esperanzas, miedos, apetencias, e intentos. No sólo las distintas aptitudes preposicionales, y sus servidores conceptuales forman el lecho en el que ocurre el habla, sino que no hay posibilidad de llegar a una comprensión profunda de los hechos lingüísticos excepto si esta comprensión se acompaña mediante un enfoque entrelazado de las actitudes cognitivas y conativas centrales.

Es pedir demasiado que estas nociones intensionales básicas se reduzcan a otra cosa —a algo más conductual, neurológico, o fisiológico, por ejemplo—. No que podamos analizar ninguno de estos tres básicos —creencia, deseo, y significado— en términos de uno o dos de los otros; o eso creo, y lo he argumentado en otra [190] parte<sup>60</sup>. Pero incluso si pudiéramos efectuar una reducción en este trío básico, los resultados no alcanzarían lo que podría esperarse simplemente porque el punto final —la interpretación, digamos, del habla— estaría demasiado próxima a donde empezamos (con creencia y deseo, o con intención, que es el producto de la creencia y el deseo). Un tratamiento básico de cualquiera de estos conceptos debe empezar más allá o por debajo de todos ellos, o en algún punto equidistante de todos ellos.

Si esto es así, un análisis del significado lingüístico que asuma la identificación previa de intenciones o propósitos no lingüísticos será radicalmente incompleto. Y no ayudará el apelar a reglas o convenciones explícitas o implícitas, aunque sólo sea porque éstas deben entenderse en términos de intenciones y creencias. Las convenciones y reglas no explican el lenguaje; el lenguaje las explica a ellas. No hay duda, por supuesto, de la importancia de mostrar cómo están conectados significados e intenciones. Tales conexiones dan estructura a las actitudes proposicionales y permiten un tratamiento sistemático de ellas. Pero la interdependencia de las actitudes intensionales básicas es tan completa que carece de base esperar entender una independientemente del

---

<sup>60</sup> Para consideraciones en apoyo de estas afirmaciones, véase mi «Belief and the Basis of Meaning», *Synthese*, XXVII (1974): 309–323; «Radical Interpretation», *Dialectica*, XVII (1973): 313–328; y «Thought and Talk», en Samuel Guttenplan, ed., *Mind and Language*, Nueva York: Oxford, 1975, pp. 7–23.

entendimiento de las otras. Lo que se busca, entonces es un tratamiento que ofrezca una interpretación de las palabras de los hablantes al mismo tiempo que proporcione una base para atribuir al hablante creencias y deseos. Un tratamiento tal pretende proporcionar una base para, mejor que asumir, la individuación de las actitudes proposicionales.

La teoría bayesiana de la decisión, tal como la desarrolló Ramsey<sup>61</sup>, trata dos de los tres aspectos intencionales de la racionalidad que parecen los más fundamentales, la creencia y el deseo. La elección de un curso de acción sobre otro, o la preferencia de que se dé un estado de cosas mejor que otro, es el producto de dos consideraciones: el valor que se coloca en las distintas consecuencias posibles, y cómo se juzga que serán esas consecuencias, dado que la acción se realice o que el estado de cosas llegue a darse. Al elegir una acción o estado de cosas, por tanto, un agente racional seleccionará una, el [191] valor relativo de cuyos posibles resultados, cuando se equilibra mediante la probabilidad que el agente asigna a esas consecuencias, es el mayor. Actuar es siempre un juego de azar, puesto que un agente no puede nunca estar seguro de cómo resultarán las cosas. Así en la medida en que un agente es racional tomará lo que crea que es la mejor apuesta disponible (él «maximiza la utilidad esperada»).

Un rasgo de una teoría tal es que lo que está diseñada para explicar —las preferencias o elecciones ordinales entre opciones— está relativamente abierto a la observación, mientras que el mecanismo

---

<sup>61</sup> «Truth and Probability», en *The Foundations of Mathematics*, Nueva York: Humanities, 1950, pp. 156–198.

explicativo, que involucra grado de creencia y valores cardinales, no se considera observable. La cuestión que se plantea, por tanto, es cuándo una persona tiene un cierto grado de creencia en alguna proposición, o cuáles son las fuerzas relativas de sus preferencias. El problema evidente es que lo que se conoce (la preferencia ordinal, o simple) es la resultante de dos desconocidos, el grado de creencia y la fuerza relativa de la preferencia. Si las preferencias cardinales de una persona por los resultados se conocieran, entonces sus elecciones entre cursos de acción revelarían su grado de creencia; y si su grado de creencia se conociera, sus elecciones dejarían al descubierto los valores relativos que él coloca en los resultados. Pero ¿cómo pueden ambos desconocidos determinarse a partir de las elecciones simples o las preferencias únicamente? Ramsey solucionó este problema mostrando cómo, sobre la única base de elecciones simples, es posible encontrar una proposición que se trata como si tuviera la misma probabilidad de ser verdadera que su negación. Esta proposición simple puede usarse entonces para construir una serie infinita de apuestas, elecciones entre las cuales ofrece una medida del valor para todas las opciones y eventualidades posibles. Es entonces rutina el calcular los grados de creencia de todas las proposiciones.

Ramsey fue capaz de cambiar este truco especificando las restricciones sobre los patrones permisibles de preferencias o elecciones simples. Estas restricciones no son arbitrarias, sino que son parte de un tratamiento satisfactorio de las razones a favor de las preferencias y la conducta electiva de una persona. Las restricciones explican la exigencia de que un agente sea racional, no en sus valores particulares y últimos, sino en los patrones que éstos forman unos con otros y en

combinación con sus creencias. La teoría tiene así un fuerte elemento normativo, pero un elemento que es esencial si los conceptos de preferencia, creencia, razón, y acción intencional han de tener aplicación. [192]

El patrón en lo que se observa es central para la inteligibilidad de la conducta electiva de un agente —determina nuestra habilidad para entender acciones en cuanto hechas por una razón—. El mismo patrón es central para el poder de la teoría para extraer, de los hechos que tomados de uno en uno están conectados de una manera relativamente directa con lo que puede observarse, hechos de una clase más sofisticada (grados de creencia, comparaciones de diferencias de valor). Desde el punto de vista de la teoría, los hechos sofisticados explican los simples, más observables, mientras que los observables constituyen la base evidencial para contrastar o aplicar la teoría.

La teoría de la decisión bayesiana no proporciona una definición de los conceptos de creencia y preferencia sobre la base de nociones no intensionales. Más bien, hace uso de una noción intensional la preferencia ordinal entre apuestas o resultados, para dar contenido a otras dos nociones, grado de creencia y comparaciones de diferencias de valor. Así sería una equivocación creer que la teoría proporciona una reducción de conceptos intensionales a otra cosa. Sin embargo, es un paso importante en la dirección de reducir conceptos intensionales complejos y relativamente teóricos a conceptos intensionales que en aplicación están más cerca de la conducta públicamente observable. Por encima de todo, la teoría muestra cómo es posible asignar un contenido a dos actitudes preposicionales básicas e interrelacionadas

sin asumir que ninguna de las dos está entendida de antemano.

Como teoría para explicar las acciones humanas, una teoría de la decisión bayesiana del tipo de la que he estado describiendo está abierta a la crítica de que presupone que podemos identificar e individuar las proposiciones a las que se dirigen actitudes como la creencia y el deseo (o la preferencia). Pero como se dijo hace algunas páginas, nuestra habilidad para identificar, y distinguir entre, las proposiciones que un agente mantiene no puede separarse de nuestra habilidad para entender lo que dice. En general, descubrimos exactamente lo que alguien quiere, prefiere, o cree sólo mediante la interpretación de su habla. Esto es particularmente obvio en el caso de la teoría de la decisión, donde los objetos que se escogen o prefieren son a menudo apuestas complejas, con resultados que se describen como contingentes sobre la ocurrencia de eventos específicos. Claramente, una teoría que intenta eliminar las actitudes y creencias que explican las preferencias o elecciones debe incluir una teoría de la interpretación verbal si no ha de hacer asunciones mutiladas. Lo que debemos añadir a una teoría de la decisión, o incorporar a [193] ella, es una teoría de la interpretación verbal, una manera de decir lo que un agente quiere decir mediante sus palabras. Sin embargo esta adición debe hacerse en ausencia de información detallada acerca de los contenidos proposicionales de creencias, deseos, o intenciones.

En aspectos importantes, el tratamiento de Quine del significado es sorprendentemente similar al tratamiento de Ramsey del tomar decisiones. Nótese que, mientras que no hay una manera directa de observar lo que los hablantes quieren decir, toda la evidencia requerida



para llevar a cabo la comunicación debe estar disponible públicamente, Quine examina la evidencia disponible relevante, y pregunta cómo podría usarse para eliminar los significados. Lo que puede observarse, por supuesto, es la conducta del habla en relación al entorno, y desde esto ciertas actitudes hacia oraciones pueden inferirse de manera bastante directa, del mismo modo en que las preferencias pueden inferirse a partir de elecciones. Para Quine, los observables clave son actos de *asentimiento* y *disentimiento*, en cuanto causados por eventos dentro del ámbito del hablante. A partir de tales actos es posible inferir que ciertos tipos de eventos causan el que el hablante mantenga que una oración es verdadera<sup>62</sup>.

Exactamente aquí aparece un desafío básico. Un hablante mantiene que una oración es verdadera como resultado de dos consideraciones: lo que él considera que la oración significa, y lo que él cree que es el caso. El problema es que lo que es directamente observable relativamente para un intérprete es el producto de dos actitudes inobservables, creencia y significado. ¿Cómo pueden distinguirse los papeles de estos dos factores explicativos y extraerse a partir de la evidencia? Curiosamente el problema es parecido al problema de desenredar los papeles de la creencia y la preferencia al determinar elecciones y preferencias.

La solución de Quine se parece, en principio si no en detalle, a la de Ramsey. El paso crucial en ambos casos es encontrar la manera de mantener un factor fijo en ciertas situaciones mientras se determina el

---

<sup>62</sup> El paso desde los asentimientos observados a la actitud inferida de mantener que algo es verdadero no está, creo, explícito en Quine.

otro. La idea clave de Quine es que la interpretación correcta de un agente por otro no puede admitir inteligiblemente ciertas clases y grados de diferencia entre el que interpreta y lo interpretado con respecto a la creencia. Como resultado, un intérprete está justificado a hacer ciertas asunciones acerca de las creencias de un agente [194] antes de que comience la interpretación. Como una restricción sobre la interpretación, a esto se le llama a veces por el nombre que Neil Wilson<sup>63</sup> le dio, el *Principio de Caridad*. Como una estrategia para separar el significado y la creencia sin asumir ninguno, es una alternativa brillante a cualquier tratamiento del significado que tome a los significados por garantizados o que asuma la distinción analítico-sintético.

En lo que sigue, uso el inspirado método de Quine de maneras que se desvían, a veces sustancialmente, de la suya. Una diferencia relevante para el presente tópico es ésta. Mientras que a Quine le conciernen las condiciones de traducción con éxito desde el lenguaje de un hablante al de un intérprete, yo pongo el énfasis en lo que el intérprete necesita conocer de la semántica del lenguaje del hablante, esto es, lo que se transmite mediante las oraciones-T implicadas por una teoría de la verdad. La relación entre estos dos proyectos, el de Quine y el mío, es obvia; dada una teoría de la verdad para el lenguaje de un hablante L expresado en el lenguaje del intérprete M, es bastante simple producir un manual que traduzca (al menos aproximadamente) desde *L* a *M*<sup>64</sup>. Pero la conversa es falsa; hay muchas oraciones que

---

<sup>63</sup> «Substances without Substrata», *Review of Metaphysics*, XII (1959): 521–539.

<sup>64</sup> La navegación puede no ser completamente recta; es fácil imaginar un lenguaje

podemos traducir sin tener ni idea de cómo incorporarlas a una teoría de la verdad. Exigir que una teoría de la interpretación satisfaga la restricción de una teoría de la verdad significa que debe hacerse manifiesta más estructura de la que se necesita para la traducción.

Si suponemos, como el principio de caridad dice que inevitablemente debemos, que el patrón de oraciones al que el hablante asiente refleja la semántica de las constantes lógicas, es posible detectar e interpretar aquellas constantes. Los principios que guían aquí, como en la teoría de la decisión, derivan de consideraciones normativas. Las relaciones entre creencias juegan un papel constitutivo decisivo; un intérprete no puede aceptar desviaciones grandes u obvias de sus propios *estándares* de racionalidad sin destruir el fundamento de la inteligibilidad sobre el que descansa toda interpretación. La posibilidad de entender el habla o las acciones de un agente depende de la [195] existencia de un patrón fundamentalmente racional, un patrón que debe, en líneas generales, ser compartido por todas las criaturas racionales. No tenemos más elección, entonces, que proyectar nuestra propia lógica sobre el lenguaje y las creencias de otro. Esto significa que es una restricción sobre las interpretaciones posibles de las oraciones mantenidas como verdaderas el que sean (dentro de la razón) lógicamente consistentes unas con otras.

La consistencia lógica no ofrece más que la interpretación de las constantes lógicas, sin embargo (cualesquiera que sean los límites de la lógica y la lista de las constantes lógicas). Mayor interpretación requie-

---

que no contenga traducción alguna de la palabra castellana 'ahora' pero que pueda dar las condiciones de verdad de las oraciones castellanas que contengan la palabra 'ahora'.

re mayores formas de acuerdo entre el hablante y el intérprete. Asumiendo que la identificación de las constantes lógicas requerida para la estructura cuantificacional de primer orden se ha conseguido, es posible identificar como tales a los términos singulares y a los predicados. Esto plantea la cuestión de cómo tienen que interpretarse éstos. Aquí el progreso depende de prestar atención, no sólo a qué oraciones un agente considera verdaderas, sino también a los eventos y objetos del mundo que causan el que él considere a las oraciones verdaderas. Las circunstancias, observables del mismo modo por hablante e intérprete, que causan que un agente acepte oraciones como 'está lloviendo', 'eso es un caballo', o 'me duele un pie' como verdaderas proporcionan la evidencia más obvia para la interpretación de esas oraciones y de los predicados en ellas. El intérprete, al darse cuenta de que el agente acepta o rechaza regularmente la oración 'el café está preparado' cuando el café está o no está preparado tenderá (aunque sea tentativamente dependiendo de resultados relacionados) a una teoría de la verdad que diga que una preferencia de un agente de la oración 'el café está preparado' es verdadera si y sólo si el agente puede observar que el café está preparado en el momento de la preferencia.

La interpretación de los nombres y predicados comunes depende fuertemente de los elementos deícticos en el habla, tales como demostrativos y tiempos verbales, puesto que son éstos los que de manera más directa permiten conectar predicados y términos singulares con objetos y eventos en el mundo. (Para acomodar a los elementos deícticos, las teorías de la verdad del tipo propuesto por Tarski deben completarse; la naturaleza de estas modificaciones ha sido discutida en

otro lugar<sup>65</sup>.) El método que propongo para interpretar los predi-  
[196]cados y las oraciones más observacionales es similar en algunos  
aspectos al método de Quine en *Word and Object* (§7–10), pero es  
diferente en otros. La diferencia más importante concierne a los  
objetos o eventos que determinan el contenido comunicable. Para  
Quine, son los patrones de las terminaciones nerviosas los que provo-  
can el asentimiento a una oración; una oración de observación de un  
hablante es «estimulativamente sinónima» de una oración de observa-  
ción de un intérprete si los mismos patrones de estimulación proximal  
provocaran el aceptar o rechazar las oraciones respectivas de hablante  
e intérprete. La idea de Quine es captar en una forma científicamente  
respetable la idea empirista de que el significado depende de la eviden-  
cia directamente disponible para cada hablante. En contraste, mi  
enfoque es externalista: sugiero que la interpretación depende (en las  
situaciones más simples y básicas) de los objetos y eventos externos  
sobresalientes tanto para el hablante como para el intérprete, los  
mismos objetos y eventos son entonces considerados por el intérprete  
como el tema de las palabras del hablante. Es el estímulo distal lo que  
cuenta para la interpretación<sup>66</sup>. El significado de este punto será ahora  
valorado.

---

<sup>65</sup> El tipo de modificación requerida se discute en *Inquiries into Truth and Interpretation*.

<sup>66</sup> He discutido este aspecto de la teoría del significado de Quine en «Meaning, Truth and Evidence», en R. Gibson, ed., *Perspectives on Quine*, Nueva York: Blacwell, 1989. Allí señalo que Quine a veces parece también subscribir la teoría «distal», especialmente en *The Roots of Reference*, La Salle, YL: Open Court, 1973.

La dificultad con lo que podríamos llamar *Teoría Distal de la Referencia* es que hace difícil explicar el error, el hueco crucial entre lo que uno cree que es verdadero y lo que es verdadero; puesto que la teoría distal basa la verdad en la creencia, el problema es crucial. La solución depende de dos estrategias interpretativas íntimamente relacionadas. Un intérprete dedicado a trabajar sobre los significados de un hablante se da cuenta de más cosas que las que causan asentimiento y disentimiento; se da cuenta de lo bien colocado y equipado que está el hablante para observar aspectos de su entorno, y de acuerdo con eso le da más peso a algunas respuestas verbales que a otras. Esto le da los rudimentos de una explicación de los casos desviados donde el hablante llama a una oveja una cabra porque está equivocado acerca del animal más que acerca de la palabra. La estrategia más sutil y más importante depende de la interanimación de oraciones. Con esto quiero decir la medida en la que un hablante cuenta la verdad de una oración como apoyo a la verdad de otras. [197]

Hemos visto un ejemplo de cómo la evidencia de tales dependencias lleva a la interpretación de las constantes lógicas. Pero las cuestiones de apoyo evidencial pueden también ayudar en la interpretación de los así llamados términos observacionales, ayudando a explicar el error.

La interpretación de los términos menos directamente dependientes de la observación no atenta debe depender también en una amplia medida de probabilidades condicionales, que muestran lo que el agente cuenta como evidencia para la aplicación de sus predicados más teóricos. Si queremos identificar y así interpretar el papel de los

conceptos teóricos o su expresión lingüística, debemos saber cómo se relaciona con otros conceptos y palabras. Estas relaciones son en general holistas y probabilísticas. Podemos, por tanto, localizarlas sólo si podemos detectar el grado en el que un agente considera que una oración es verdadera, sus probabilidades subjetivas. El asentimiento y disentimiento simples están en los límites extremos y opuestos de una escala; necesitamos colocar las actitudes que son intermedias en cuanto a fuerza. El grado de creencia, sin embargo, no puede ser directamente diagnosticado por un intérprete; como vimos al discutir la teoría de la decisión, el grado de creencia es una construcción basada en actitudes más elementales.

La teoría de la interpretación verbal y la teoría de la decisión bayesiana están evidentemente hechas la una para la otra. La teoría de la decisión debe liberarse de la asunción del acceso independiente de los significados; la teoría del significado necesita una teoría del grado de creencia para hacer un uso serio de las relaciones de apoyo evidencial. Pero afirmar estas dependencias mutuas no es suficiente, porque ninguna teoría puede desarrollarse primero como una base para la otra. No hay ninguna forma de añadir simplemente una a la otra porque cada una para empezar requiere un elemento extraído de la otra. Lo que se busca es una teoría unificada que ofrezca el grado de creencia, las deseabilidades en una escala de intervalos, y una interpretación del habla, una teoría que no asuma que los deseos o las creencias tienen que individualizarse de antemano, mucho menos que se cuantifiquen.

Una teoría tal debe basarse en alguna actitud simple que un in-

térprete pueda reconocer en un agente antes de que el intérprete tenga conocimiento detallado de ninguna de las actitudes proposicionales del agente. La actitud siguiente servirá: la actitud que un agente tiene hacia dos de sus oraciones cuando prefiere la verdad de una a la verdad de la otra. Las oraciones deben estar dotadas de significado para [198] el hablante, por supuesto, pero interpretar las oraciones es parte de la tarea del intérprete. Lo que el intérprete tiene que interpretar, entonces, es la información acerca de qué episodios y situaciones en el mundo causan que un agente prefiera que esta oración mejor que otra sea verdadera. Claramente un intérprete puede saber esto sin saber lo que las oraciones significan, qué estados de cosas valora el agente, o qué cree. Pero es igual de claro que el que un agente prefiera la verdad de las oraciones es una función de lo que el agente considere que las oraciones significan, el valor que coloque sobre los distintos estados del mundo reales o posibles, y la probabilidad que ligue a aquellos estados que son contingentes sobre la verdad de las oraciones relevantes. Así no es absurdo pensar que estas tres actitudes del agente puedan abstraerse del patrón de preferencias entre oraciones de un agente.

Podría objetarse que una preferencia por la verdad de una oración más que por otra es en si misma un estado intencional, y uno que podría saberse que se da sólo sobre la asunción de que están presentes muchos factores psicológicos. Esto es verdad (como lo es también del asentimiento a, o del considerar verdadera, una oración). Pero el objetivo no era el evitar estados intencionales; era evitar estados intencionales *individuativos*, estados *intensionales*, estados con un objeto proposicional (como se dice). Una preferencia por la verdad de una oración sobre otra es una relación extensional que relaciona a un



agente y a dos oraciones (y un tiempo). Porque puede detectarse sin saber lo que las oraciones significan, una teoría de la interpretación basada en ella puede esperar dar el paso crucial desde lo no proposicional a lo proposicional.

Aquí, en esbozo, está cómo creo que la esperanza puede satisfacerse. Hemos visto ya (de nuevo en forma esquemática) cómo llegar a una teoría del significado y la creencia sobre la base del conocimiento acerca de los grados en los cuales las oraciones se consideran verdaderas. Así, si pudiéramos derivar el grado de creencia en las oraciones apelando a la información acerca de las preferencias de que las oraciones sean verdaderas, tendríamos una teoría unificada con éxito.

La versión de Ramsey de la teoría de la decisión bayesiana hace un uso esencial de los juegos de azar o las apuestas, y esto crea una dificultad para mi proyecto. Porque ¿cómo podemos decir que un agente considera que una oración presenta una apuesta hasta que hayamos llegado bastante lejos en el proceso de interpretar su lenguaje? Una apuesta, después de todo, especifica una conexión, probablemente causal entre la ocurrencia de un cierto evento (una moneda que cae de cara) y un resultado específico (ganas un caballo). Incluso si asumimos que podemos decir cuándo un agente acepta tal conexión, la aplicación clara de la teoría depende también de que el evento que es la causa (la moneda cayendo de cara) no tenga valor, positivo o negativo, en sí mismo. También es necesario asumir que la probabilidad que el agente asigna a que la moneda caiga de cara no está contaminada por pensamientos acerca de la probabilidad de ganar un caballo. En pruebas experimentales de teorías de la decisión, uno

trata de proporcionar entornos en los que estas asunciones tengan una posibilidad de ser verdaderas; pero la aplicación general que tenemos en mente ahora no puede ser tan detallada.

Le debemos a Richard Jeffrey<sup>67</sup> una versión de la teoría de la decisión bayesiana que no hace uso directo de apuestas, sino que trata a los objetos de la preferencia, los objetos a los que se asignan las probabilidades subjetivas, y a los objetos a los que se asignan valores relativos todos como proposiciones. Jeffrey ha mostrado en detalle cómo extraer probabilidades subjetivas y valores a partir de las preferencias de que las proposiciones sean verdaderas.

Queda un problema obvio. Jeffrey muestra cómo conseguir resultados más o menos como Ramsey sustituyendo preferencias entre proposiciones por preferencias entre apuestas. Pero las proposiciones

---

<sup>67</sup> *The Logic of Decision* (Chicago: University Press, 2.<sup>a</sup> ed., 1983). La teoría de Jeffrey no determina las probabilidades y utilidades hasta los mismos conjuntos de transformaciones que la teoría *standard*. En vez de una función de utilidad determinada hasta una transformación lineal, en la teoría de Jeffrey la función de utilidad es única sólo hasta una transformación lineal fraccional; y las asignaciones de probabilidad, en vez de ser únicas una vez que se ha escogido un número para medir la certeza (siempre Uno), son únicas sólo dentro de una cierta cuantización. Estas disminuciones en la determinación son conceptual y prácticamente apropiadas: equivalen, entre otras cosas, a permitir algo como el mismo tipo de indeterminación en la teoría de la decisión que hemos llegado a esperar en una teoría de la interpretación lingüística. En la misma medida en que se puede dar razón de los mismos datos en la teoría de la decisión usando distintas funciones de utilidad haciendo los correspondientes cambios en la función de probabilidad, así se pueden cambiar los significados que se atribuyen a las palabras de una persona (dentro de unos límites) siempre que se hagan los cambios compensatorios en las creencias que se le atribuyen.

son significados, u oraciones con significados, y si sabemos las proposiciones entre las que un agente está escogiendo, nuestro problema original de interpretar el lenguaje e individuar actitudes proposicionales se asume que ha sido solucionado desde el principio. Lo que queremos es alcanzar los resultados de Jeffrey, pero empezando con preferencias entre oraciones sin interpretar, no proposiciones.

Esto resulta ser un problema soluble. El método de Jeffrey para encontrar las probabilidades subjetivas y las deseabilidades relativas de las proposiciones sólo depende de la estructura veritativo–funcional de las proposiciones —de cómo se hacen las proposiciones a partir de proposiciones simples mediante la aplicación repetida de la conjunción, disyunción, y de las otras operaciones definibles en términos de éstas—. Si empezamos con oraciones en vez de proposiciones, entonces la dificultad crucial se vencerá suponiendo que las conectivas veritativo–funcionales puedan identificarse. Porque una vez que las conectivas veritativo–funcionales hayan sido identificadas, Jeffrey ha mostrado cómo fijar, hasta el grado deseado, las deseabilidades subjetivas y las probabilidades de todas las oraciones; y esto, he argumentado, es suficiente para ofrecer una teoría para la interpretación de las oraciones. El conocer las actitudes evaluativas y cognitivas de un agente hacia las oraciones interpretadas no es algo que se discierna (al menos en el contexto de este enfoque) a partir del conocimiento de las creencias y deseos del agente. Los pasos esenciales en este procedimiento, particularmente el procedimiento que saca la interpretación de las conectivas veritativo–funcionales de hechos acerca de la preferencia, se describen en el apéndice de este artículo.

El tratamiento de los problemas del significado, la creencia y el deseo que he esbozado no pretende, estoy seguro de que esto está claro, arrojar ninguna luz directa sobre cómo llegamos a entendernos unos a otros en la vida real, ni sobre cómo dominamos nuestros primeros conceptos y nuestro primer lenguaje<sup>68</sup>. He estado comprometido en un ejercicio conceptual dirigido a revelar las dependencias entre nuestras actitudes preposicionales básicas en un nivel suficientemente fundamental como para evitar la asunción de que podemos llegar a captarlas —o a atribuírselas inteligiblemente a otros— una cada vez. Realizar el ejercicio ha requerido mostrar cómo es posible en principio llegar a todas ellas a la vez. Mostrar esto equivale a presentar una prueba informal de que hemos dotado al pensamiento, el deseo, y el habla de una estructura que haga posible la interpretación. Por supuesto, sabíamos que era posible de antemano. La cuestión filosófica era ¿*qué* la hace posible?

---

<sup>68</sup> Dado lo intrincado de cualquier sistema interpretable de pensamiento y lenguaje, he asumido que debe haber muchos tratamientos alternativos a la interpretación. He bosquejado uno; otros podrían ser menos artificiales o más cercanos a nuestras intuiciones concernientes a la práctica interpretativa. Pero no deberíamos dar por garantizado que el procedimiento que he esbozado es totalmente remoto de lo que es practicable. Para empezar, obsérvese que toda preferencia que pueda tratarse como una petición o demanda sincera debe tomarse como si expresara la preferencia del que la profiere de que una oración, mejor que su negación, sea verdadera. La mayor parte del trabajo experimental en teoría de la decisión toma como datos las elecciones que los sujetos hacen entre alternativas que se describen por escrito o en el habla. Se asume normalmente que los sujetos entienden estas descripciones que los experimentadores hacen. Abandonar esta asunción da como resultado datos exactamente del tipo requerido por el tratamiento presentado aquí.

Lo que hace a la tarea practicable en absoluto es la estructura que el carácter normativo del pensamiento, el deseo, el habla y la acción imponen sobre las atribuciones correctas de actitudes a los otros, y así sobre las interpretaciones de su habla y las explicaciones de sus acciones. Lo que he dicho acerca de las normas que gobiernan nuestras teorías de la atribución intencional es tosco, vago, e incompleto. La manera de mejorar nuestra comprensión de tal comprensión es mejorar nuestra captación de los standards de racionalidad implícitos en toda interpretación del pensamiento y la acción.

La idea de que el contenido proposicional de las oraciones de observación se determina (en la mayoría de los casos) mediante lo que es común y sobresaliente tanto al hablante como al intérprete es un correlato directo de la concepción de sentido común del aprendizaje del lenguaje. Tiene profundas consecuencias para la relación entre el pensamiento y el significado, y para nuestra concepción del papel de la verdad, porque no sólo asegura que hay un nivel básico en el cual los hablantes comparten sus concepciones, sino que también que lo que ellos comparten es una visión ampliamente correcta de un mundo común. La fuente última tanto de la objetividad como de la comunicación es el triángulo que, al relacionar hablante, intérprete, y el mundo, determina los contenidos del pensamiento y del habla. Dada esta fuente, no hay lugar para un concepto relativizado de verdad.

Reconocemos que la verdad debe de alguna manera relacionarse con las actitudes de las criaturas racionales; esta relación se revela ahora como si surgiera de la naturaleza del entendimiento interpersonal. La comunicación lingüística, el instrumento indispensable del

entendimiento interpersonal de grano fino, descansa en las preferencias mutuamente entendidas, cuyos contenidos se fijan finalmente mediante los patrones y las causas de las oraciones consideradas verdaderas. El apuntalamiento conceptual de la interpretación es una [202] teoría de la verdad; la verdad descansa así, al final, en la creencia y, más al final incluso, en las actitudes afectivas.

## APÉNDICE

El método de Jeffrey para encontrar las probabilidades subjetivas y las deseabilidades relativas de las proposiciones depende sólo de la estructura veritativo–funcional de las proposiciones —de cómo están construidas las proposiciones a partir de proposiciones simples por la aplicación repetida de la conjunción, disyunción, negación, y las otras operaciones definibles en términos de éstas—. Si empezamos con oraciones en vez de con proposiciones, entonces nuestro problema se solucionará siempre y cuando las conectivas veritativo–funcionales puedan ser identificadas. Porque una vez que las conectivas veritativo–funcionales han sido identificadas, Jeffrey ha mostrado cómo fijar, con el grado deseado, las probabilidades y deseabilidades subjetivas de todas las oraciones; y esto, he argumentado, basta para dar una teoría para interpretar las oraciones.

El primitivo empírico básico en el método que hay que describir es la preferencia (débil) del agente de que sea verdadera una oración mejor que otra; uno podría por tanto pensar en los datos como si fueran del mismo tipo que los datos habitualmente reunidos en una

prueba experimental de cualquier teoría bayesiana de la decisión, siempre y cuando la interpretación de las oraciones entre las que el agente elige no se asuma como conocida de antemano por el intérprete.

La uniformidad y simplicidad de la ontología empírica del sistema, que comprende sólo las preferencias y las oraciones, es esencial para lograr el objetivo de combinar la teoría de la decisión con la interpretación. Seguiré a Jeffrey, cuya teoría trata sólo de proposiciones, tan cerca como sea posible, sustituyendo oraciones sin interpretar donde él asume proposiciones. Aquí, entonces, está el análogo del *Axioma de Deseabilidad* (D) de Jeffrey, aplicado a oraciones en vez de a proposiciones:

(D) Si  $\text{prob}(s \text{ y } t) = 0$  y  $\text{prob}(s \text{ o } t) > 0$ , entonces

$$\text{des}(s \text{ o } t) = \frac{\text{prob}(s)\text{des}(s) + \text{prob}(t)\text{des}(t)}{\text{prob}(s) + \text{prob}(t)}$$

[203] [Escribo ‘prob (s)’ para la probabilidad subjetiva de s y ‘des(s)’ para la deseabilidad de s.] Al relacionar la preferencia y la creencia, este axioma hace el tipo de trabajo que habitualmente se hace en las apuestas; la relación es, sin embargo, diferente. Los eventos se correlacionan con oraciones que bajo la interpretación resulta que dicen que el evento ocurre (‘la carta siguiente es un trébol’). Las acciones y los resultados también están representados por oraciones (‘el agente apuesta un dólar’, ‘el agente gana cinco dólares’). Las apuestas no entran directamente, pero el elemento de riesgo está presente, puesto que elegir que una oración sea verdadera es habitualmente correr un riesgo acerca de lo que será concomitantemente verdadero. (Se asume

que uno no puede elegir una oración lógicamente falsa.) Así vemos que, si el agente elige hacer verdadera en vez de falsa la oración ‘el agente apuesta un dólar’, está tomando una posibilidad sobre un resultado, que podría pensarse, por ejemplo, que depende de si la siguiente carta es o no un trébol. Entonces la deseabilidad de la (verdad de) la oración ‘el agente apuesta un dólar’ será la deseabilidad de las distintas circunstancias en las que la oración es verdadera, sopesadas de la forma habitual por las probabilidades de esas circunstancias. Supongamos que el agente cree que ganará cinco dólares si la carta siguiente es un trébol y que no ganará nada si la carta siguiente no es un trébol; tendrá entonces un especial interés en si la verdad de ‘el agente apuesta un dólar’ se emparejará con la verdad o falsedad de ‘la siguiente carta es un trébol’. Abreviemos estas dos oraciones por ‘s’ y ‘t’. Entonces

$$\text{des}(s) = \frac{\text{prob}(s \text{ y } t)\text{des}(s \text{ y } t) + \text{prob}(s \text{ y } \bar{t})\text{des}(s \text{ y } \bar{t})}{\text{prob}(s)}$$

Esto es, por supuesto, algo parecido a las apuestas de Ramsey. Difiere, sin embargo, en que no hay ninguna asunción de que «los estados de la naturaleza» que podría pensarse que determinan los resultados sean, en la terminología de Ramsey, «moralmente neutrales», esto es, que no tengan efecto sobre las deseabilidades de los resultados. Ni hay tampoco la asunción de que las probabilidades de los resultados dependan de nada más que de las probabilidades de los «estados de naturaleza» (el agente podría creer que tiene una posibilidad de ganar cinco dólares incluso si la siguiente carta no es un trébol,



y una posibilidad de que no ganará cinco dólares incluso si la próxima carta es un trébol). [204]

El axioma de deseabilidad puede usarse para mostrar cómo las probabilidades dependen de las deseabilidades en el sistema de Jeffrey. Tomemos el caso especial donde  $t = s$ . Entonces tenemos

$$(1) \quad \text{des}(s \text{ o } s) = \text{des}(s)\text{prob}(s) + \text{des}(s)\text{prob}(s)$$

Puesto que  $\text{prob}(s) + \text{prob}(s) = 1$ , podemos resolver por  $\text{prob}(s)$ :

$$(2) \quad \text{prob}(s) = \frac{\text{des}(s \text{ o } s) - \text{des}(s)}{\text{des}(s) - \text{des}(s)}$$

Así, la probabilidad de una proposición depende de la deseabilidad de esta proposición y de su negación. Además, es fácil ver que, si una oración  $s$  es más deseable que una verdad lógica arbitraria (tal como ' $t$  o  $\neg t$ '), entonces su negación ( $\neg s$ ) no puede también ser más deseable que una verdad lógica. Supongamos que asignamos el número 0 a cualquier verdad lógica. (Esto es intuitivamente razonable puesto que un agente es indiferente a la verdad de una tautología.) Entonces (2) puede reescribirse:

$$(3) \quad \text{prob}(s) = 1 - \frac{\text{des}(s)}{\text{des}(s)}$$

De repente es evidente que  $des(s)$  y  $des(\neg s)$  no pueden ser ambas mayores, o ambas menores, que 0, la deseabilidad de cualquier verdad lógica, si  $prob(s)$  tiene que caer en el intervalo de 0 a 1. Si (siguiendo a Jeffrey) llamamos buena a una opción si es preferida a una verdad lógica y mala si una verdad lógica es preferida a ella, entonces (3) muestra que es imposible que una opción (oración) y su negación sean ambas buenas o ambas malas.

Tomando '(s y s)' como nuestra muestra de verdad lógica (podemos afirmar este principio en términos puramente de preferencias:

$$(4) \quad \text{Si } des(s) > des(\neg(s \text{ y } \neg s)) \text{ entonces} \\ des(\neg(s \text{ y } \neg s)) \geq des(\neg s), \text{ y}$$

Puesto que tanto la negación como la conjunción pueden definirse en términos de la barra de Sheffer '|' («no a la vez»), (4) puede reescribirse: [205]

$$(5) \quad \text{Si } des(s) > des((t|u)|((t|u)|(t|u))) \text{ entonces} \\ des((t|u)|((t|u)|(t|u))) \geq des(s|s), \text{ y} \\ \text{si } des((t|u)|((t|u)|(t|u))) > des(s) \text{ entonces} \\ des(s|s) \geq des((t|u)|((t|u)|(t|u))).$$

El interés de (5) para los propósitos actuales es éste. Si asumimos que '|' es algún operador veritativo-funcional arbitrario que forma oraciones a partir de pares de oraciones, entonces se cumple lo siguiente: si (5) es verdadero para todas las oraciones s, t, y u, y para algún s y t,  $des(s|s) \geq des(t|t)$ , entonces '|' debe ser la barra de Sheffer

(debe tener las propiedades lógicas de «no a la vez»); ninguna otra interpretación es posible<sup>69</sup>.

Así, los datos que involucran sólo preferencias entre oraciones, cuyos significados son desconocidos para el intérprete, ha llevado (dada las restricciones de la teoría) a la identificación de una conectiva oracional. Puesto que todas las oraciones lógicamente equivalentes son iguales en deseabilidad, es ahora posible interpretar todas las otras conectivas oracionales veritativo–funcionales, puesto que todas son definibles en términos de la barra de Sheffer. Por ejemplo, si se encuentra que para toda oración  $s$ ,

$$\text{des}(s|s) = \text{des}(\neg s)$$

podemos concluir que la tilde es el signo de negación.

Ahora es posible medir la deseabilidad y probabilidad subjetiva de todas las oraciones, porque la aplicación de fórmulas como (2) y (3) exige la identificación sólo de las conectivas oracionales veritativo–funcionales. Así, está claro a partir de (3) que, si dos oraciones son iguales en deseabilidad (y se prefieren a una verdad lógica) y sus negaciones son también iguales en deseabilidad, las oraciones deben tener la misma probabilidad. Por lo mismo, si dos oraciones son iguales en deseabilidad (y se prefieren a una verdad lógica), pero la negación de una se prefiere a la negación de la otra, entonces la proba-

---

<sup>69</sup> Estoy en deuda con Stig Kangeen por mostrarme por qué un intento anterior para solucionar este problema no funcionaría. También él añadió algunos refinamientos necesarios a la propuesta actual.

bilidad de la primera es menor que la de la segunda. Esto, junto con los apropiados axiomas de existencia, es suficiente para establecer una escala de probabilidad. Entonces es fácil determinar las deseabilidades relativas de todas las oraciones<sup>70</sup>. [206]

En este punto las probabilidades y deseabilidades de todas las oraciones han sido en teoría determinadas. Pero ninguna oración completa ha sido todavía interpretada, aunque las conectivas oracionales veritativo–funcionales han sido identificadas, y así las oraciones lógicamente verdaderas o falsas en virtud de la lógica oracional pueden reconocerse.

Hemos mostrado cómo interpretar las oraciones más simples sobre la base de (grados de) creencia en su verdad. Dados los grados de creencia y fuerzas de deseo relativas de la verdad de las oraciones interpretadas, podemos dar un contenido proposicional a las creencias y deseos de un agente.

---

<sup>70</sup> Para los detalles véase Jeffrey, *The Logic of Decision*.