Paradojas Semánticas y Cambio de Lógica[1]

Federico Marulanda
Instituto de Investigaciones Filosóficas
Universidad Nacional Autónoma de México
federico@filosoficas.unam.mx

Resumen: Versiones informales y formales de la paradoja del mentiroso son discutidas, junto con posibles respuestas, con miras a evaluar el argumento según el cual dichas paradojas apuntan hacia la necesidad de reformar la lógica clásica.

Palabras clave: paradoja del mentiroso, lógica paraconsistente, pluralismo lógico.

Abstract: Informal and formal versions of the Liar's Paradox are discussed, together with possible responses, in order to evaluate the argument according to which these paradoxes indicate a need to revise are classical logic.

Keywords: Liar's Paradox, paraconsistent logic, pluralism logical.

1. La paradoja del mentiroso en el lenguaje natural

La paradoja del mentiroso es comúnmente introducida de manera informal, derivando una conclusión aparentemente inaceptable a partir de la oración 'La presente oración es falsa' —oración a la que, para facilitar la discusión, llamaré 'M1'. Dicha derivación se basa en el supuesto de que el valor de verdad de toda oración afirmativa es o verdadero o falso. Si M1 es verdadera, lo que dice es el caso: luego, de acuerdo con lo que dice, es falsa. Pero si, por el contrario, M1 es falsa, lo que dice es, de hecho, cierto, y la oración es verdadera. Así pues, M1 parece ser una oración que es verdadera si, y sólo si, es falsa.

Una reacción natural al argumento anterior es la de rechazar la premisa según la cual toda oración afirmativa es o verdadera o falsa,

---

[1] Una versión preliminar de este trabajo fue presentada bajo el mismo título en el Simposio 'Lógicas y Filosofía de las Lógicas', en el marco del XIV Congreso Internacional de Filosofía de la Asociación Filosófica de México, celebrado en noviembre 2007 en Mazatlán, Sinaloa.

y mantener, en su lugar, que ciertas oraciones afirmativas, M1 entre ellas, no son ni verdaderas ni falsas. Pero a esta reacción inicial la aquejan al menos dos problemas. El primero es que vierte escasa luz sobre el porqué oraciones como M1 desembocan en una paradoja: simplemente se responde a una conclusión aparentemente inaceptable abandonando el principio semántico de la bivalencia, piedra angular de la lógica clásica. El segundo problema con mantener que M1 no es ni verdadera ni falsa es que esto no cierra definitivamente las puertas a la paradoja. La razón es que si M1 no es ni verdadera ni falsa, *a fortiori* no es falsa, y por ende, lo que dice es, de hecho, falso.[2] Lo que nos conduce a la nueva paradoja de que si M1 no es ni verdadera ni falsa, es falsa.

Una segunda reacción a la paradoja original es la de anotar que el argumento que a ella conduce emplea un método de evaluación poco común, y posiblemente ilegítimo: normalmente no se evalúa una oración suponiendo ahora su verdad, ahora su falsedad. ¿Es acaso este método evaluativo el culpable del problema? La respuesta es negativa: hace varias décadas que Tarski, siguiendo el ejemplo de £ukasiewicz, demostró cómo obtener una versión en lenguaje natural de la paradoja del mentiroso sin apelar a una evaluación cuestionable, sino a tres suposiciones aparentemente correctas (Tarski 1933 [1983]: 157):

> (i) El hecho empírico que una oración aparece inscrita en un lugar concreto.
>
> (ii) La validez de convenciones tradicionales de denominación (en particular, el hecho que podamos darle a un objeto cualquiera un nombre cualquiera, y de ahí en adelante podamos hacer referencia a él mediante ese nombre).
>
> (iii) La validez del esquema de verdad de Tarski para todas las oraciones afirmativas del lenguaje natural.

[2] Repito el razonamiento para garantizar su comprensión. M1 dice de sí misma que es falsa. Si suponemos que M1 no es ni verdadera ni falsa, estamos suponiendo, en particular, que no es falsa. Pero en tal caso lo que M1 dice es falso.

De las anteriores suposiciones, únicamente la tercera requiere comentario. El esquema de verdad de Tarski puede ser visto como la conjunción de dos principios: el principio de *desencomillado,* según el cual, de la atribución de verdad a la cita de una oración, se sigue una afirmación de la misma (*e.g.*, si la oración 'Mazatlán está a orillas del Pacífico' es verdadera, se sigue que Mazatlán está a orillas del Pacífico), y el principio converso de *ascenso semántico*, según el cual, de la afirmación de una oración, se sigue la atribución de verdad a su cita (*e.g.*, de la afirmación que Mazatlán está a orillas del Pacífico, se sigue que 'Mazatlán está a orillas del Pacífico' es verdadera). Así pues, según el esquema de verdad de Tarski, la oración 'Mazatlán está a orillas del Pacífico' es verdadera si, y sólo si, Mazatlán está a orillas del Pacífico (y algo análogo puede decirse de toda oración afirmativa).

Por razonables que parezcan, las suposiciones (i) – (iii) conllevan a una nueva versión de la paradoja del mentiroso. Para ver cómo, basta considerar la siguiente variante de M1, a la que denominaremos 'M2':

La única oración inscrita en un recuadro en el artículo *Paradojas Semánticas y Cambio de Lógica* de F: Marulanda es falsa.

De acuerdo con el esquema de verdad de Tarski:

> 'La única oración inscrita en un recuadro en el artículo *Paradojas Semánticas y Cambio de Lógica* de F. Marulanda es falsa' es verdadera si, y sólo si, la única oración inscrita en un recuadro en el artículo *Paradojas Semánticas y Cambio de Lógica* de F. Marulanda es falsa.

Apelando al nombre de la oración que nos ocupa, podemos afirmar lo anterior de otra manera:

> M2 es verdadera si, y sólo si, la única oración inscrita en un recuadro en el artículo *Paradojas Semánticas y Cambio de Lógica* de F. Marulanda es falsa.

Pero 'M2' denota a la única oración inscrita en un recuadro en el artículo *Paradojas Semánticas y Cambio de Lógica* de F. Marulanda. En consecuencia:

> La única oración inscrita en un recuadro en el artículo *Paradojas Semánticas y Cambio de Lógica* de F. Marulanda es verdadera si, y sólo si, la única oración inscrita en un recuadro en el artículo *Paradojas Semánticas y Cambio de Lógica* de F. Marulanda es falsa.

En otras palabras, M2 es verdadera, si, y sólo si, es falsa.

¿Cómo responder a la paradoja del mentiroso, si no es ni rechazando el principio de bivalencia, ni rechazando el método de evaluación al cual se somete M1 en la primera versión expuesta de la paradoja? Existen varias propuestas, ninguna de las cuales cuenta con el apoyo unánime de aquellos que se han ocupado a profundidad del asunto. Aunque no es mi objetivo principal defender una respuesta al fenómeno paradójico en el contexto del lenguaje natural, expondré brevemente aquella que, a mi modo de ver, explica y resuelve de manera más satisfactoria lo que aquí ocurre (a este respecto véanse Wittgenstein 1939 [1976], Gaifman 2000). La respuesta comienza por hacer una distinción entre una *oración tipo*, las oraciones que son *instancias* de este tipo, y las *afirmaciones* hechas a través de tales instancias (ver Strawson 1950). Una oración tipo es un objeto abstracto que puede ser instanciado de manera concreta, física, por múltiples instancias. Ahora, según la posición que busco delinear, ni las oraciones tipo, ni sus instancias, tienen valor de verdad: las que son verdaderas o falsas son las afirmaciones hechas a través de una instancia de una oración tipo. Por ejemplo, la oración tipo 'Tengo frío' ha sido instanciada un sinnúmero de veces. Si en este momento profiriera una nueva instancia del tipo, estaría haciendo una afirmación falsa, pues actualmente no tengo frío; en un futuro otra instancia del tipo seguramente me servirá para hacer una afirmación verdadera.

Continuando con la respuesta a la paradoja, prestemos ahora atención a las características de M1 y de M2. En ambas hay una atribución semántica, en este caso de falsedad, a cierta oración (ya sea a una oración tipo, o a una instancia de oración: el sustantivo 'oración' es usado de manera ambigua). Aquí nos encontramos ante una dificultad, pues si nos circunscribimos a la distinción entre oraciones tipo, instancias de oración, y afirmaciones, únicamente las últimas de las cuales son o verdaderas o falsas, encontramos que en M1 y M2 se comete un error de categoría, pues se le atribuye falsedad a una oración. Pero evidentemente no es ésta la fuente del problema, pues bien podemos enunciar una nueva variante de M1 (llamémosla 'M3') que desemboque de nuevo en la paradoja sin recaer en el error de categoría mencionado —a saber, 'La presente afirmación es falsa'.[3]

Preguntémonos ahora cómo se procede, en el curso normal de las cosas, cuando se evalúan aquellas afirmaciones en las que se hace una atribución de verdad o falsedad a una afirmación (en aras del análisis, y corriendo el riesgo de alejarnos un poco del uso común, acordemos en utilizar locuciones como 'La afirmación que . . . es verdadera' o 'La afirmación que . . . es falsa', para tales efectos[4]). Tomemos, por ejemplo, el caso de:

> (1) La afirmación que Mazatlán está a orillas del Pacífico es verdadera.

El procedimiento a seguir para evaluar la afirmación hecha a través de (la anterior instancia de)[5] (1) es sencillo y no genera

[3] Otra variante más precisa, pero menos coloquial, sería 'La afirmación hecha a través de la presente oración es falsa'.

[4] Debido a que el valor de verdad de afirmaciones de instancias de oraciones con componentes indéxicos puede variar (pensemos de nuevo en afirmaciones de 'Tengo frío' hechas por distintas personas, o por la misma persona en momentos diferentes), una atribución semántica debería, como mínimo, hacer mención explícita de la persona que hace la afirmación en cuestión, y del momento en que la hace. Pero estos elementos serán omitidos aquí para simplificar la discusión, pues es claro que la paradoja del mentiroso no es causada por indexicalidad de este tipo (en M2, por ejemplo, tal indexicalidad es eliminada por completo).

[5] Puesto que una afirmación se hace siempre a través de una instancia de una oración, y no a través de su tipo, de aquí en adelante no continuaré señalando la distinción.

controversia. Primero se verifica si Mazatlán está a orillas del Pacífico. Habiendo comprobado que sí lo está, se concluye que la afirmación hecha a través de (1) es verdadera.

Pasemos ahora al caso de:

(2) La afirmación que Mazatlán está a orillas del Pacífico es falsa.

Procediendo de igual manera que antes, concluimos sin dificultad que la afirmación hecha a través de (2) es falsa. Continuemos con:

(3a) Mazatlán está a orillas del Pacífico.

(3b) La anterior afirmación es falsa.

La situación no difiere mucho de la que se presenta en (2). Bajo el supuesto que el demostrativo 'la anterior' que figura en (3b) se refiere a la afirmación hecha a través de (3a), y dado que esta última afirmación es verdadera, no es controversial concluir que la afirmación hecha a través de (3b) es falsa.

Hasta aquí no hay sorpresas. El procedimiento recién delineado parece, pues, servir para evaluar cualquier afirmación acerca del valor de verdad de otra afirmación. Pero veamos qué sucede en el caso de:

(M3) La presente afirmación es falsa.

De acuerdo con el procedimiento recién expuesto, para evaluar la afirmación hecha a través de M3, necesitamos evaluar primero la afirmación a la cual se refiere el demostrativo que figura en M3. Pero el demostrativo 'la presente' que aparece en M3 se refiere a la afirmación hecha a través de M3. En consecuencia, para conocer el valor de verdad de la afirmación hecha a través de M3, tendríamos primero que conocer el valor de verdad de la afirmación hecha a través de M3. Pero esto es algo que resulta, en principio, imposible de hacer.

La anterior discusión revela que M3 desencadena un ciclo que imposibilita llevar a cabo su evaluación semántica, lo que justifica concluir que M3 no es propicia para hacer afirmación alguna. Nótese bien que el punto no es que una afirmación hecha a través de M3 no es ni verdadera ni falsa —esto daría lugar a una nueva versión de la paradoja[6]— sino que no se logra hacer una afirmación, ya sea verdadera, ya sea falsa, profiriendo una oración como M3. Con esta conclusión se desarma la paradoja, pues si no se ha hecho ninguna afirmación por medio de M3, no se da por iniciado el argumento según el cual la afirmación hecha por medio de M3 es verdadera si, y sólo si, es falsa. Y lo mismo puede decirse de cualquier variante de M3 que se vea envuelta, al hacer referencia al valor semántico de sí misma o de otra afirmación o serie de afirmaciones, en un ciclo que impida determinar su valor semántico por medio de un procedimiento normal de evaluación como el recién expuesto.

2. La paradoja del mentiroso en teorías axiomáticas de verdad

La paradoja del mentiroso no surge exclusivamente en el lenguaje natural, sino que puede aparecer también en teorías deductivas expresadas en lenguajes formales, cuando estos lenguajes contienen vocabulario semántico. En particular, si se toma una teoría matemática axiomatizada *T* con un mínimo de alcance, y se expande el lenguaje en el cual *T* está expresada con un predicado de verdad (cuya interpretación pretendida es que estén en su extensión todos y sólo los enunciados verdaderos de *T*), todo esto con el objetivo de formular una teoría axiomática de verdad, existe un alto riesgo de que *T* sea una teoría inconsistente.

Por una teoría matemática con un mínimo de alcance entiendo una teoría que cuente con los recursos para (i) describir

[6] Ya hemos visto más arriba que no se exorciza la versión de la paradoja del mentiroso expuesta al principio de esta sección con decir que la oración M1 no es verdadera ni falsa; lo mismo sucede con M3. O, para exponer el punto con un ejemplo más, podemos enunciar una nueva variante de M1, a la que llamaremos 'M4': 'La presente afirmación no es verdadera'. Si se mantuviera que una afirmación hecha a través de M4 no es ni verdadera ni falsa, *a fortiori* no sería verdadera. Pero esto último es lo que se afirma con M4. Luego una afirmación de M4 resultaría verdadera, y al mismo tiempo ni verdadera ni falsa. M4 es la variante de la paradoja conocida como *el mentiroso reforzado*, cuya versión formalizada nos ocupará en la siguiente sección.

completamente la sintaxis del lenguaje en el cual está expresada, y (ii) probar un lema de diagonalización. Como veremos, estos dos ingredientes, junto con el esquema de verdad de Tarski, son suficientes para generar la versión formalizada de la paradoja del mentiroso. En el presente artículo, usaré como ejemplo de teoría matemática con un mínimo de alcance la aritmética de Peano (*PA*); pero todo lo dicho aquí seguiría siendo válido si habláramos de teorías deductivamente bastante más débiles que *PA*, como la aritmética primitiva recursiva, o la aritmética de Robinson. Por otra parte, es claro que teorías axiomáticas más potentes que *PA*, como las teorías de conjuntos de Zermelo o de Zermelo-Fraenkel, también podrían servir como base para una teoría axiomática de verdad.

¿Qué motiva la formulación de una teoría axiomática de verdad? Enumeraré tres razones, sin pretender ser exhaustivo ni connotar primacía por el orden de enunciación. La primera es la siguiente: el primer teorema de incompletud de Gödel establece que en el lenguaje de una teoría deductiva con alcance mínimo, T, existen enunciados que podemos reconocer como verdaderos, pero que no pueden ser demostrados en T. No obstante, Feferman 1991 demuestra que es posible construir una extensión de T, en la cual es posible dar una caracterización formal de la *totalidad* de las consecuencias de T, minimizando así los efectos de la incompletud: la idea básica es construir una teoría T^* que expande los axiomas de T con axiomas que gobiernan un predicado de verdad aplicable a los enunciados tanto de T, como de su extensión T^*. Una segunda razón para querer formular teorías axiomáticas de verdad es que en ellas se puede simular la cuantificación sobre propiedades, obviando así el uso de la lógica de segundo orden, y en consecuencia reduciendo el compromiso ontológico. La idea central aquí es que, en un lenguaje con un predicado de verdad, en vez de expresar, *e.g.*, que cierto objeto denotado por la constante a posee la propiedad ϕ, puede decirse que la fórmula $\phi(a)$ es verdadera. En consecuencia, en tal lenguaje, lo que normalmente se expresaría por medio de cuantificación de segundo orden sobre propiedades, puede expresarse por medio de cuantificación de primer orden sobre fórmulas (*e.g.*, en vez de decir que todas las propiedades de *a* son propiedades de *b* y que todas las propiedades de *b* son propiedades de *a*, o en otras palabras, que *a* y *b* comparten las mismas propiedades, puede decirse que una fórmula $\phi(a)$ es verdadera si, y sólo si, $\phi(b)$ es verdadera). Una tercera aplicación de las teorías axiomáticas de verdad es la de servir como complemento para sus contrapartes semánticas. En una teoría semántica de verdad, la interpretación (posiblemente parcial) del predicado de verdad de un lenguaje objeto se especifica en un metalenguaje (ver, *e.g.*, Kripke 1975, Gaifman 2000). Mientras que el teorema de Tarski establece que el metalenguaje en el cual se efectúa la definición debe tener recursos superiores a los del lenguaje objeto, el análisis de teorías axiomáticas de verdad permite establecer exactamente en qué consiste dicha superioridad. A su vez, las teorías semánticas de verdad son usadas para guiar la búsqueda de axiomas en sus contrapartes axiomatizadas, así como para examinar sus modelos.

Pero regresemos a la paradoja del mentiroso. En la sección anterior vimos que, al ser confrontados con ciertas oraciones del lenguaje natural que desembocan en una consecuencia paradójica, es plausible responder que estas oraciones no son propicias para hacer una afirmación —para, como diría Wittgenstein, "hacer una movida en el juego del lenguaje". Pero cuando nos trasladamos a lenguajes y teorías formales, ya no es posible hacer una distinción entre lo proferido y lo afirmado. En una teoría deductiva, un enunciado es una fórmula bien formada que puede ser o un axioma, o deducida de los axiomas como teorema, y en ambos casos las consecuencias deductivas de tal enunciado forman necesariamente parte de las consecuencias de la teoría. En otras palabras, si en una teoría deductiva un razonamiento análogo al de la paradoja del mentiroso en el lenguaje natural permite la deducción de una contradicción, la teoría es inconsistente, y esto es algo que tiene consecuencias ineludibles. Por esto todo aquel que busque enunciar una teoría axiomática de verdad debe tener claro, desde la partida, cómo va a confrontar el fenómeno de la paradoja. Veamos cómo se presenta este fenómeno en el contexto de una teoría aritmética como *PA* expandida por axiomas que gobiernen un predicado de verdad.

Un lenguaje $\mathcal{L}$ adecuado para la aritmética es un lenguaje de primer orden con identidad, que usualmente contiene como símbolos no lógicos: una constante individual para denotar al número 0; signos funcionales para denotar las funciones de sucesor, adición y multiplicación; y un signo diádico de predicado para denotar la relación "menor que". Supongamos que el vocabulario de $\mathcal{L}$ es expandido por dos signos de predicado monádicos adicionales, $V(\ )$ y $F(\ )$, con los que se pretende denotar predicados de verdad y falsedad, respectivamente. A primera vista, la función que puedan tener estos predicados parece oscura, pues en una teoría aritmética interpretada en el modelo estándar, cuyo universo de cuantificación es el conjunto de los enteros naturales, los predicados monádicos toman como argumentos miembros del universo, es decir números enteros, de los que no es coherente afirmar que sean verdaderos o falsos. Pero gracias a la técnica de codificación de Gödel —codificación que se lleva a cabo en la aritmética misma— toda expresión lingüística formada con el vocabulario de $\mathcal{L}$, y en particular, toda fórmula cerrada de $\mathcal{L}$, puede ser asociada con un número natural único. La idea es, pues, utilizar a los predicados $V(\ )$ y $F(\ )$ para expresar la verdad y falsedad de las fórmulas cerradas de $\mathcal{L}$, interpretando a $V(x)$ como "x es (el número Gödel de) una fórmula cerrada verdadera de $\mathcal{L}$", y a $F(x)$ como "x es (el número Gödel de) una fórmula cerrada falsa de $\mathcal{L}$".[7]

Concentrémonos por el momento en el predicado de verdad de $\mathcal{L}$. ¿Cómo controlar su comportamiento a través de axiomas de una teoría aritmética, T? Es razonable pensar que si $V(\ )$ ha de jugar el papel de predicado de verdad para los enunciados de $\mathcal{L}$, debe cumplir, como mínimo, con las siguientes características: (i) si es

[7] Omito los detalles de cómo interpretar los predicados de verdad y falsedad cuando su argumento no es (el número Gödel de) una fórmula cerrada de $\mathcal{L}$.

demostrable en T que un enunciado aritmético es verdadero (es decir, que el número Gödel del enunciado pertenece a la extensión del predicado $V(\ )$), el enunciado en sí debe ser demostrable en T, y, conversamente, (ii) si un enunciado es demostrable en T, también debe ser demostrable en T que el enunciado es verdadero. En otras palabras, es razonable esperar que los axiomas de T permitan deducir, para toda fórmula cerrada A de $\mathcal{L}$,

$$\vdash_T \ V(\langle A\rangle) \leftrightarrow A$$

en donde $\langle A\rangle$ es el numeral que denota al número Gödel de A.

Supongamos, pues, que los axiomas de T consisten simplemente de (i) los axiomas de PA y (ii) el esquema de axioma $V(\langle A\rangle) \leftrightarrow A$, que no es otro que el esquema de verdad de Tarski, cuya versión informal fue discutida en la sección anterior. Con base en esta axiomatización, y bajo el supuesto que T está regida deductivamente por la lógica clásica, es posible demostrar que T es inconsistente. La razón es que los axiomas aritméticos de T permiten la demostración de un teorema conocido como el *lema de diagonalización*, según el cual, si $G(v)$ es cualquier fórmula abierta de $\mathcal{L}$ con variable libre, v, existe una fórmula cerrada B de $\mathcal{L}$ tal que

$$\vdash_T \ B \leftrightarrow G(\langle B\rangle).$$ [8]

Consideremos ahora la fórmula abierta de $\mathcal{L}$, $\neg V(x)$. El lema de diagonalización garantiza la existencia de una fórmula cerrada B tal que

$$\vdash_T \ B \leftrightarrow \neg V(\langle B\rangle).$$

B es, pues, una oración de $\mathcal{L}$ que es intersustituible con la adscripción de la negación de verdad a su nombre (su análoga en el

[8] Para mayores detalles, ver, *e.g.*, Boolos *et al.* 2007: 221.

lenguaje natural es la oración M4, o el mentiroso reforzado, mencionada en pie de página al final de la primera sección). Pero el esquema de verdad de Tarski, que fue estipulado como un esquema de axioma de T, garantiza que

$$\vdash_T \ V(\langle B \rangle) \leftrightarrow B.$$

En consecuencia,

$$\vdash_T \ V(\langle B \rangle) \leftrightarrow \neg V(\langle B \rangle)$$

lo que en lógica clásica conlleva a

$$\vdash_T \ V(\langle B \rangle) \,\&\, \neg V(\langle B \rangle)$$

es decir, a una contradicción explícita (la conjunción de un enunciado y su negación), lo que hace de T una teoría trivial, pues una contradicción tiene como consecuencia clásica a todas las fórmulas cerradas del lenguaje.

Hasta aquí, la versión formal de la paradoja del mentiroso reforzado. Para derivar en T la paradoja del mentiroso propiamente dicha, basta con definir el predicado de falsedad tal y como se hace tradicionalmente: la falsedad de toda fórmula cerrada A de $\mathcal{L}$ equivale a la verdad de su negación, *i.e.*,

$$F(\langle A \rangle) \leftrightarrow V(\langle \neg A \rangle).$$

Ahora bien, el lema de diagonalización garantiza la existencia de una fórmula cerrada C tal que

$$\vdash_T \ C \leftrightarrow F(\langle C \rangle).$$

C es, pues, una oración de $\mathcal{L}$ que es intersustituible con la atribución de falsedad a su nombre (su análoga en el lenguaje natural es la oración M1 de la sección anterior). Adicionalmente, el esquema de verdad de Tarski, garantiza que

$$\vdash_T \ V(\langle C \rangle) \leftrightarrow C.$$

Sustituyendo equivalentes llegamos a

$$\vdash_T \ V(\langle C \rangle) \leftrightarrow F(\langle C \rangle).$$

Esto quiere decir que es demostrable en T que C es verdadera, si, y sólo si, C es falsa. La anterior expresión de la paradoja del mentiroso es fácilmente convertible en una contradicción explícita. Apelando a la definición de $F(\ )$, y sustituyendo equivalentes, se deriva

$$\vdash_T \ V(\langle C \rangle) \leftrightarrow V(\langle \neg C \rangle)$$

de lo que clásicamente se sigue que

$$\vdash_T \ V(\langle C \rangle) \,\&\, V(\langle \neg C \rangle)$$

y, finalmente se concluye, por medio del esquema de verdad de Tarski, que

$$\vdash_T \ C \,\&\, \neg C$$

lo cual trivializa a T.

En suma: a diferencia del caso del lenguaje natural, en el contexto de una teoría formalizada de verdad T, las paradojas del mentiroso y del mentiroso reforzado arrojan un veredicto inapelable: la combinación de los axiomas de una teoría matemática con recursos mínimos por una parte y el esquema de verdad de Tarski por la otra es inconsistente. Lo que es más, si la lógica que subyace a T es la lógica clásica, T es trivial: todas las fórmulas cerradas bien formadas de su lenguaje $\mathcal{L}$ son deducibles como teoremas, lo que hace de T una teoría inútil.

3. La respuesta paraconsistente a la paradoja del mentiroso

Los axiomas aritméticos que permiten describir la sintaxis de $\mathcal{L}$ y probar el teorema de diagonalización no están en tela de juicio (y en cualquier caso, existen varias otras teorías que se podrían utilizar para el mismo efecto, y es inconcebible desecharlas a todas). En consecuencia, los proponentes de una teoría axiomática de verdad se han visto obligados a abandonar el intento de gobernar el predicado de verdad por medio de un axioma o axiomas que permitan la derivación de todas las instancias del esquema de verdad de Tarski, por intuitivo que éste parezca.

Pero ¿qué axiomas postular en su lugar? La respuesta está lejos de ser evidente. En McGee 1992 se establece que hay un número infinito de teorías que contienen un máximo de instancias del esquema de Tarski sin ser inconsistentes, y que estas teorías pueden ser incompatibles entre sí (en otras palabras, se establece que imponer la limitación que una teoría axiomática de verdad contenga el mayor número posible de instancias del esquema de Tarski sin que esto conlleve a una contradicción, no constituye un criterio suficiente para identificar una teoría única o preferida).

Adicionalmente, en Montague 1966 y McGee 1985 se demuestra que si una teoría aritmética contiene ya no todas las instancias del esquema de verdad de Tarski, pero todas las instancias de ciertas debilitaciones más o menos naturales del mismo, la teoría sigue siendo inconsistente, o en el mejor de los casos, ω-inconsistente (*i.e.*, no puede tener un modelo cuyo universo de cuantificación sea el conjunto de los números naturales).

En Friedman y Sheard 1987 se analizan varias otras combinaciones de axiomas y leyes de inferencia que al parecer podrían gobernar un predicado de verdad, encontrando que entre aquellas teorías resultantes que son consistentes u ω-consistentes, un gran número de generalizaciones aparentemente correctas que contienen el predicado de verdad son, por fuerza, indemostrables.

En Feferman 1991 y en trabajos inspirados en éste se encuentran variantes de una teoría axiomática aritmetizada de verdad de mayor capacidad deductiva que las analizadas por Friedman y Sheard, pero el precio a pagar es que ciertas fórmulas que son caracterizadas por las teorías en cuestión como no verdaderas resultan demostrables como teoremas,[9] lo que corrompe el vínculo entre demostrabilidad y verdad—siendo este el vínculo cuyo estudio originalmente motivaba, por lo menos en parte, la formulación de una teoría axiomática de verdad.

Visto lo anterior, cobra valor explorar una opción diferente: ¿qué pasaría si se guardara el esquema de verdad de Tarski como axioma (o algún otro u otros que permitieran deducir todas sus instancias), pero se cambiaran las reglas deductivas que gobiernan la teoría, de tal manera que la contradicción que surge a raíz de la paradoja del mentiroso no desencadene la trivialización? Esta es precisamente la estrategia propuesta en Priest 1987 [2006], donde se defiende el uso de una lógica paraconsistente como lógica subyacente de una teoría axiomática aritmetizada de verdad (las *lógicas paraconsistentes* son aquellas que no avalan el uso de axiomas o leyes de inferencia que permitan la deducción de un enunciado arbitrario a partir de una contradicción).

Para evaluar la propuesta de Priest, es importante empezar por observar que sus motivaciones para proponer una versión paraconsistente de la teoría axiomática aritmetizada de verdad difieren fundamentalmente de las vistas en la sección anterior. El primer teorema de incompletud de Gödel impone limitaciones en teorías matemáticas clásicas de alcance mínimo, demostrando que en sus lenguajes pueden expresarse enunciados reconocibles metateóricamente como verdaderos pero que no pueden ser demostrables en dichas teorías, pues si lo fueran éstas serían inconsistentes. Pero si la lógica subyacente a una cierta teoría permite tolerar su inconsistencia, es decir, permite que teoremas

[9] Por ejemplo, si B es una fórmula del lenguaje de la teoría de Feferman tal que $B \leftrightarrow \neg V(\langle B \rangle)$ es derivable, entonces $\neg V(\langle B \vee \neg B \rangle)$ también será derivable. Ahora, esta última oración dice que $B \vee \neg B$ no es verdadera (o más precisamente: que su número Gödel no está en la extensión del predicado de verdad). Sin embargo, $B \vee \neg B$ es derivable en la teoría, por ser una instancia de la ley del tercio excluso.

contradictorios sean probados en ella sin desencadenar la trivialidad, el objetivo de formular una supra-teoría axiomática de verdad para caracterizar formalmente las verdades de la teoría original pierde interés, pues hacer esto ya hubiera sido posible en la teoría original (el precio, por supuesto, es el de tener que explicar cómo dos teoremas contradictorios pueden ser conjuntamente verdaderos).

De manera similar, la motivación para formular una teoría axiomática de verdad según la cual hacerlo permitiría establecer qué recursos adicionales debe tener un metalenguaje en el cual se especifica la interpretación de un lenguaje objeto, desaparece en el contexto paraconsistente, pues la necesidad de hacer una distinción entre lenguaje objeto y metalenguaje con recursos superiores proviene únicamente de la necesidad de evitar contradicciones generadas a raíz de la presencia de predicados semánticos.

La motivación de Priest para formular una teoría aritmetizada paraconsistente de verdad es, pues, otra. Más allá de buscar formular una teoría en la que puedan ser deducidas todas las instancias del esquema de verdad de Tarski (aunque éstas sean inconsistentes), su motivación es la de abogar por un cambio de lógica: afirmar que el fenómeno de las paradojas semánticas hace evidentes las limitaciones de la lógica clásica, y por ende hace necesaria la búsqueda de un sistema deductivo paraconsistente adecuado para remplazarla.

En lo que resta de este artículo quiero exponer algunas razones por las cuales considero que una teoría aritmetizada paraconsistente de verdad no es atractiva. Acto seguido, quiero argumentar que, independientemente de si la lógica paraconsistente propuesta por Priest es defendible, la posibilidad de enunciarla y adoptarla no cuenta como evidencia de un presunto pluralismo lógico.

Empecemos por señalar que no cualquier lógica paraconsistente es adecuada para regular la deducción en una teoría axiomática aritmetizada de verdad. Por una parte, una lógica adecuada para esta tarea debe ser lo suficientemente fuerte para permitir la deducción de teoremas aritméticos estándar, sin lo cual el ejercicio carecería de interés. Pero por otra parte, la paradoja de Curry demuestra que una teoría axiomática aritmetizada de verdad es trivial, sin tener que pasar por la derivación de una fórmula arbitraria a partir de una contradicción, siempre que su lógica subyacente avale ciertos pasos o leyes de inferencia básicos.

El argumento de Curry parte de la suposición (que compartimos aquí) que una teoría T tiene como axiomas el esquema de verdad de Tarski, así como los axiomas de PA. En ese caso, es posible deducir en pocos pasos la fórmula cerrada arbitraria A, a partir de la fórmula abierta $V(x) \rightarrow A$, de la siguiente manera:

1. $C \leftrightarrow (V(\langle C \rangle) \rightarrow A)$	Teorema de T (por el lema de diagonalización)
2. $V(\langle C \rangle) \leftrightarrow C$	Teorema de T (por el esquema de verdad de Tarski)
3. $V(\langle C \rangle) \leftrightarrow (V(\langle C \rangle) \rightarrow A)$	1, 2, sustitución de equivalentes
4. $V(\langle C \rangle) \rightarrow (V(\langle C \rangle) \rightarrow A)$ & $(V(\langle C \rangle) \rightarrow A) \rightarrow V(\langle C \rangle)$	3, definición del bicondicional
5. $(V(\langle C \rangle)$ & $(V(\langle C \rangle) \rightarrow A)) \rightarrow A$	Instancia del principio de *Afirmación*, *i.e.*, de la tautología clásica $(\varphi$ & $(\varphi \rightarrow \psi)) \rightarrow \psi$
6. $(V(\langle C \rangle)$ & $V(\langle C \rangle)) \rightarrow A$	3, 5, sustitución de equivalentes
7. $V(\langle C \rangle) \rightarrow A$	6, idempotencia de la conjunción
8. $(V(\langle C \rangle) \rightarrow A) \rightarrow V(\langle C \rangle)$	4, eliminación de la conjunción
9. $V(\langle C \rangle)$	7, 8, *modus ponens*
10. A	7, 9, *modus ponens*

Lo anterior demuestra que quien quiera construir una teoría de verdad de la cual sean teoremas todas las instancias del esquema de verdad de Tarski y quiera evitar su trivialización a manos de la paradoja del mentiroso gracias a la adopción de una lógica paraconsistente, está forzado a definir dicha lógica de tal manera que al menos uno de los principios empleados en los pasos 3 a 9 de la prueba anterior no sea válido.[10]

Puesto que la sustitución de equivalentes, la idempotencia de la conjunción, la eliminación de la conjunción y el *modus ponens* son principios básicos difícilmente desechables, Priest opta por definir la lógica paraconsistente que subyace a su teoría axiomática de verdad de tal manera que el principio de Afirmación no sea una verdad lógica. Para conseguirlo, formula su lógica desde una perspectiva semántica, dotándola de un operador intensional, $\rightarrow$, que invalida al principio de Afirmación pero que en otros respectos se comporta como un condicional, y luego genera un sistema deductivo correcto y completo con respecto a dicha semántica.

Hay varios ángulos desde los cuales criticar la estrategia recién descrita.

Primero, la semántica de la lógica de Priest es *ad hoc*, un artilugio que permite detener la derivación de principios que llevarían a la trivialización de su teoría de verdad, pero que apela a nociones de cuestionable justificación filosófica (la semántica del operador condicional intensional hace uso, por ejemplo, de la noción de mundos imposibles, mundos en los que las leyes de la lógica son inaplicables).

Segundo, la lógica en sí impide el uso de patrones de razonamiento tradicionales como *ex falso quodlibet*, el silogismo disyuntivo (*modus tollendo ponens*), y la introducción del

[10] Un camino alterno para generar la paradoja de Curry se vale de una instancia de una tautología clásica diferente a la invocada en el paso 5, el principio de *Contracción*, $(\varphi\rightarrow(\varphi\rightarrow\psi)) \rightarrow (\varphi\rightarrow\psi)$. Pero para los propósitos presentes no es necesario discutir todas las variantes de esta paradoja.

condicional, pues todos ellos trivializarían una teoría de verdad inconsistente.[11]

Que la lógica impida el uso de *ex falso quodlibet* es de esperarse; en todo caso esta regla no parece ser usada en el razonamiento informal, y por ello ha sido frecuentemente tenida como contraintuitiva. Más difícil parece justificar el abandono del silogismo disyuntivo, que sí es utilizado corrientemente, y el de introducción del condicional, que permite el razonamiento hipotético. La pregunta que surge de manera inmediata es: ¿cómo estar seguros que al debilitar la lógica clásica de la manera que Priest propone no se pierde gran capacidad deductiva? En particular, ¿son todos los teoremas de la aritmética de Peano demostrables en la teoría aritmetizada de verdad de Priest? Si no lo son, dicha teoría perdería interés, pues difícilmente estaríamos dispuestos a abandonar leyes aritméticas a cambio de acomodar los relativamente remotos usos del predicado de verdad aplicado a los enunciados del lenguaje de la misma.

[11] Es evidente que la lógica de Priest, como cualquier lógica paraconsistente, no puede permitir el uso de *ex falso quodlibet* (de una contradicción se sigue cualquier fórmula). El silogismo disyuntivo tampoco puede ser permitido, pues en compañía de otras leyes de inferencia básicas éste trivializaría una teoría inconsistente, por medio de un argumento como el siguiente:

1. $A \mathbin{\&} \neg A$	Hipótesis
2. A	1, eliminación de la conjunción
3. $\neg A$	1, eliminación de la conjunción
4. $A \vee B$	2, introducción de la disyunción
5. B	3, 4, silogismo disyuntivo

Por último, la regla de introducción del condicional (a veces llamada 'prueba condicional') debe ser rechazada por Priest pues permite la derivación como teorema de Afirmación, lo que como hemos visto expondría su teoría de verdad a la trivialidad, a través de la paradoja de Curry, del siguiente modo:

1. $A \mathbin{\&} (A \rightarrow B)$	Hipótesis
2. A	1, eliminación de la conjunción
3. $(A \rightarrow B)$	1, eliminación de la conjunción
4. B	2, 3, *modus ponens*
5. $(A \mathbin{\&} (A \rightarrow B)) \rightarrow B$	1–4, introducción del condicional

Un tercer motivo para criticar la estrategia de Priest es que existen argumentos que llevan a pensar que ciertos enunciados del lenguaje de $\mathcal{L}$ en los cuales aparece el predicado de verdad —por ejemplo, el enunciado que expresa que toda fórmula que sea demostrable en la teoría, es verdadera— no pueden ser probados en la teoría, bajo pena de trivialidad (Field 2006: 593, 597). Pero en tal caso la teoría de verdad de Priest no representa ventajas sobre una teoría clásica de verdad, en la cual el enunciado correspondiente tampoco es demostrable (puesto que una prueba de dicho enunciado sería una prueba de la consistencia de la teoría, y el segundo teorema de incompletud de Gödel establece que una teoría clásica no puede demostrar su propia consistencia). Un cuarto y último motivo por el cual la teoría de verdad de Priest es criticable es que no es claro cuál es su modelo pretendido.

Lo dicho hasta ahora no constituye un argumento irrefutable en contra de la teoría de verdad de Priest —como tampoco es mi objetivo actual proveer uno. Sí es mi objetivo, en cambio, poner en duda la aseveración hecha por Priest que la versión formalizada de la paradoja del mentiroso constituye evidencia para pensar que la lógica clásica es defectuosa y debe ser modificada. Si bien es cierto que formular una teoría de verdad axiomatizada clásica presenta dificultades serias, formular una teoría de verdad axiomatizada paraconsistente presenta otras no menos considerables, y la resolución de tanto las unas como las otras requiere de compromisos que son *prima facie* poco atractivos. Podría pensarse, sin embargo, que la construcción de Priest provee por lo menos un nuevo ejemplo de la posibilidad de formular provechosamente lógicas diferentes a la clásica, apuntando así hacia un pluralismo lógico. Pero esta conclusión es precipitada. La semántica de la lógica formulada por Priest con el propósito de razonar en su teoría axiomática de verdad ha de ser expresada en un lenguaje formal, y la corrección y completud de dicha lógica con respecto a su semántica han de ser demostradas en una teoría formal. ¿Cuáles son esta teoría y este lenguaje? Quien lea los pasajes relevantes de la propuesta de Priest puede fácilmente tener la impresión que, como es usual, la semántica está formulada en teoría clásica de conjuntos. Pero en tal caso la construcción de Priest sólo es comprensible a través de la lógica y matemática clásicas, y no provee un ejemplo que sirva para apoyar una tesis fuerte de pluralismo lógico (es un hecho incontestable que existe un gran número de lógicas no clásicas, y en ese sentido débil hay un pluralismo lógico: la pregunta es si algunas de estas lógicas pueden ser vistas como sistemas globales e independientes de razonamiento).

Ahora, en ciertas ocasiones (ver, *e.g.*, Priest 2006: 258–260) Priest aduce que la teoría en la cual se demuestran la corrección y completud de la lógica empleada en su teoría axiomática de verdad es una teoría paraconsistente de conjuntos, cuya lógica subyacente es la misma lógica que la de la teoría de verdad. Si esto fuera el caso, la lógica paraconsistente de Priest tendría la globalidad e independencia de la que parecen gozar la lógica clásica y la lógica intuicionista. Desafortunadamente, las afirmaciones que hace Priest en este sentido no están respaldadas por una demostración detallada. Pero aun si lo estuvieran, y la lógica de Priest tuviera la globalidad e independencia necesarias, esto no constituiría un argumento a favor de una tesis fuerte de pluralismo lógico. La razón es que, desde el punto de vista de Priest, es *su* lógica y no la clásica o la intuicionista la que caracteriza correctamente la validez de una forma de argumento, y la que articula cuáles pasos deductivos son permisibles.

En otras palabras, el mismo Priest no es un pluralista lógico. Tampoco tienen que serlo un lógico clásico o un lógico intuicionista: el lógico clásico puede reinterpretar, dentro de la perspectiva de la matemática y la lógica clásicas, la actividad del intuicionista, y juzgarla como una parte propia de la suya. A su vez, el intuicionista, dotado como está de una demostración en la matemática intuicionista de la corrección y completud de la lógica intuicionista con respecto a su semántica, puede adoptar la posición que los teoremas de la lógica y matemática clásicas que no sean intuicionísticamente reformulables carecen estrictamente de sentido, pues en ellas se permite ilegítimamente juzgar como verdaderos o falsos enunciados por medio de métodos que sobrepasan nuestras capacidades cognitivas, como el uso del principio del tercio excluso en teorías cuyo modelo pretendido tiene un dominio infinito.

En conclusión, el fenómeno de la paradoja del mentiroso, ya sea en el lenguaje natural o en un lenguaje formalizado, no constituye evidencia de la necesidad de "cambiar de lógica". Aunque nada impide la elaboración de teorías de verdad axiomáticas cuya lógica subyacente no sea clásica, la viabilidad y utilidad de las mismas tiene que ser demostrada. Y aunque no se descarta que pueda serlo en un caso particular, esto no contaría necesariamente como apoyo de una tesis fuerte de pluralismo lógico: para que así fuera, la lógica no clásica utilizada tendría que ser global e independiente, y, además, la posición filosófica que sustenta su semántica tendría que tolerar y explicar la posibilidad de que la validez de los argumentos pueda ser determinada de maneras incompatibles.

Referencias

Boolos, George, Burgess, John P. y Jeffrey, Richard C. (2007) *Computability and Logic* (Fifth Edition). Cambridge: Cambridge University Press.

Feferman, S. (1991) "Reflecting on Incompleteness", *Journal of Symbolic Logic,* 56: 1–49.

Field, H. (2006) "Truth and the unprovability of consistency", *Mind,* 115 (459): 567–605.

Friedman, H. y Sheard, M. (1987) "An Axiomatic Approach to Self-Referential Truth", *Annals of Pure and Applied Logic*, 33: 1–21.

Gaifman, H. (2000) "Pointers to Propositions". Chapuis A. y Gupta A. (eds.), *Circularity, Definition, and Truth,* 79–121, New Delhi: Indian Council of Philosophical Research.

Kripke, S. (1975) "Outline of a Theory of Truth", *Journal of Philosophy*, 72 (19): 690–716.

McGee, V. (1985) "How Truthlike can a Predicate Be? A Negative Result", *Journal of Philosophical Logic* ,14 (4): 299–410.

Montague, R. (1966) "Syntactic treatments of modality, with corollaries on reflexion principles and finite axiomatizability", *Acta Philosophica Fennica*, 16: 154–167. También en Montague, R. (1974) *Formal Philosophy*, 286–302, New Haven: Yale University Press.

Priest, G. (2006) *In Contradiction*. (Second Edition). New York: Oxford University Press.

Strawson, P. F. (1950) "On Referring". *Mind*, 59 (235): 320–344.

Tarski, A. (1933) "The Concept of Truth in Formalized Languages", en Tarski, A. (1983) *Logic, Semantics, Metamathematics: Papers from 1923 to 1938* (Second Edition), 152–278. Indianapolis, Hackett.

Wittgenstein, L (1976) *Lectures on the Foundations of Mathematics: Cambridge, 1939*. Ithaca, NY: Cornell University Press.

Recibido el 10 de julio, revisado el 20 de julio, aprobado el 30 de julio 2010.