

LA TEORÍA REVISIONISTA DE LA VERDAD

Natalia M. Buacar y Lavinia M. Picollo

La teoría revisionista de la verdad pretende resolver los problemas planteados por las paradojas semánticas generadas por el concepto de verdad sin adoptar jerarquías ni limitar, en modo alguno, el poder expresivo del lenguaje. El modo de hacerlo es reconocer el carácter circular del concepto de verdad. En la primera sección introducimos el marco filosófico en el que se desarrolla la teoría y delineamos el problema específico al que responde. Ofrecemos aquí una primera aproximación a la idea de verdad como concepto circular y a su significación por medio de una regla de revisión. En la segunda sección ofrecemos una presentación formal de la teoría en sus tres variantes: $\mathbf{T}^*$, $\mathbf{T}^\#$ y $\mathbf{T}^c$. En la tercera sección llevamos adelante un estudio crítico de la propuesta. Finalmente, en el apéndice se incluye una versión axiomática de la teoría.

1 Introducción

1.1 Preliminares

La ‘teoría de la revisión de la verdad’ o ‘teoría revisionista de la verdad’ fue desarrollada de modo independiente por Anil Gupta y Hans G. Herzberger. En el mismo año estos autores publicaron los artículos seminales: Herzberger (1982a) y (1982b) y Gupta (1982). Este último fue discutido críticamente por Nuel Belnap (1982). El trabajo conjunto de Gupta y Belnap cristalizó en *The Revision Theory of Truth*, texto clave en donde se despliega de modo acabado la teoría de la revisión y que será el eje de nuestra presentación.

La teoría revisionista surge como una respuesta al desafío de ofrecer una sistema que capture el concepto de verdad mediante (una versión de) los bicondicionales $\mathbf{T}$ y que a la vez logre lidiar con los problemas planteados por las paradojas. De lo que se trata es de superar el escollo planteado por Alfred Tarski, el cual ha sido presentado en la introducción y el primer capítulo de este libro: el agregado de un predicado veritativo que respete el esquema $\mathbf{T}$ a un lenguaje con capacidad autoreferencial conduce inevitablemente a paradojas e inconsistencia, si la lógica subyacente es la clásica.

Como suele suceder frente a un problema filosófico existen al menos tres alternativas: adoptar un punto de vista escéptico respecto de la posibilidad de solucionarlo, intentar disolverlo o encomendarse a la tarea de encontrar una solución positiva. Como ya se ha señalado en la introducción a este libro, es posible concebir el argumento asociado con la

oración del mentiroso como evidencia de que el concepto de verdad es inconsistente. Esta *primera* alternativa teórica tiene sus costos dado el rol que juega la noción de verdad en la filosofía. Retomando el ejemplo tematizado en primer capítulo, la noción de verdad resulta central en la explicación de la relación de consecuencia lógica.

La *segunda* opción suele sugerir una reinterpretación de los conceptos involucrados en el planteo del problema que lo revela como un pseudoproblema. Usualmente se considera que las oraciones problemáticas, o bien no son auténticas oraciones o bien carecen de valor de verdad clásico (verdadero o falso). En el primer caso, las oraciones problemáticas son consideradas sin sentidos que no expresan proposición alguna, no pudiendo ser entonces verdaderas o falsas y se disuelve así la aparente contradicción a la cual conducen. Esta es la vía tarskiana expuesta en el primer capítulo de este libro. En el segundo caso, en cambio, se adscribe carácter infundado a dichas oraciones, razón por la cual carecen de valor de verdad clásico. Así, se disuelve el problema a partir de la disolución de las paradojas. La teoría de puntos fijos kripkeanos puede ser interpretada como un intento disolutorio en este segundo sentido.¹

Como se ha notado en el capítulo anterior, el costo que han de pagar teorías como las de Saul Kripke es sacrificar algunas expresiones de la lógica clásica. Lejos de constituir una alteración menor, dicho sacrificio supone una reformulación del problema original. Por ejemplo, resignar la el carácter exhaustivo (o exclusivo) de la negación permite evadir la paradoja.² Sin embargo, ello implica un recorte de los recursos expresivos de dicho lenguaje. Cabe sospechar entonces que no se ha solucionado el problema original que surgía en el contexto del lenguaje enriquecido, sino que se ha disuelto.

La teoría de la revisión se enmarca en la *tercera* alternativa. Una adecuada comprensión del concepto de verdad, de su naturaleza, revela que el problema de la verdad y las paradojas se vuelve resoluble. Son necesarias algunas precisiones respecto del problema específico que la teoría de la revisión se propone resolver y ello será abordado en la siguiente sección. A modo de anticipo, no se intentará bloquear paradoja alguna sino modelarlas. Se procurará ofrecer una teoría de la verdad que las explique. Esta teoría no resignará la autoreferencialidad y evitará jerarquías, a la vez que mantendrá la clasicidad de la lógica y, con ella, aquellos elementos involucrados en las paradojas semánticas, expresiones lógicas tales como las negaciones antes mencionadas. ¿Cómo ha de lograr todo esto? Poniendo en evidencia el carácter circular del concepto de verdad. La estrategia es entonces ofrecer una teoría que se adapte a dicha naturaleza, razón por la cual se requiere una teoría de las definiciones circulares. La teoría de la verdad surge como corolario de tal teoría y el concepto de verdad transparente adquiere nuevo sentido en dicho contexto.

¹Al menos de lo que en la próxima sección llamaremos ‘problema descriptivo’.

²Una negación exhaustiva es aquella que aplicada a cualquier oración no verdadera, esto es o bien falsa o bien carente de valor de verdad clásico, da lugar a una oración verdadera. Una negación exclusiva, en cambio, es aquella que no permite que una oración y su negación posean el mismo valor de verdad.

1.2 El problema a resolver

Se requieren algunas especificaciones respecto al modo en que los defensores de la teoría de la revisión entienden el problema que las paradojas semánticas plantean, pues es a la luz de ello que su propuesta ha de ser evaluada. A diferencia de lo que ocurría en capítulos anteriores—y de lo que se sugerirá en los siguientes—el problema específico al que la teoría de la revisión pretende hacer frente no es el de evitar o eliminar las paradojas, sino el de explicar cómo es que el lenguaje puede convivir con ellas sin tornarse inconsistente o inutilizable. En palabras de Herzberger:³

En lugar de intentar eliminar aquellas paradojas, quisiera considerar el experimento de alentar positivamente a que ocurran y observar cómo encuentran su propia salida. Este abordaje, que denomino semántica intuitiva, es un ejercicio deliberadamente dirigido. La idea es dar un paso atrás y permitir que las paradojas revelen su funcionamiento internos.

Desde una perspectiva similar, Gupta (1982) y Gupta & Belnap (1993) reconocen que las paradojas conllevan dos problemas, uno normativo y otro descriptivo. Así los estudios sobre paradojas caen en dos categorías. Desde un enfoque *normativo*, el objetivo es bloquear las paradojas y descubrir los cambios (si los hubiere) que las paradojas imponen sobre nuestra concepción y uso de ‘verdadero’. En esta línea se enmarcan, entre otras, las teorías abordadas en los capítulos anteriores y posteriores de este libro. El proyecto de los teóricos de la revisión, al contrario, es resolver el problema *descriptivo* propuesto por las paradojas. La explicación del concepto de verdad pretende preservar todas las características de nuestra noción ordinaria de verdad y dar cuenta del significado y el uso de las oraciones de nuestro lenguaje ordinario que contienen la palabra ‘verdadero’.

Conviven así dos orientaciones metodológicas. Por un lado, tal como destaca Herzberger, la ‘inestabilidad semántica’ que exhiben las paradojas sugiere un elemento que resultará central en el modo de concebir la evaluación semántica, como un proceso que va de etapa en etapa de modo progresivo. Pero por otro, Gupta & Belnap proponen poner el foco, no tanto en las paradojas sino en los usos ordinarios, no problemáticos, del concepto de verdad. Hay allí un fenómeno que demanda una explicación sistemática. Más aun, es la comprensión de los principios que subyacen a los usos ordinarios y no problemáticos del concepto verdad la que ha de echar luz sobre el fenómeno de las paradojas semánticas.

Tal aproximación al problema y la consecuente orientación metodológica adoptada delinean el marco en el cual se ofrece una teoría de la verdad. Esta teoría debe explicar tanto las oraciones problemáticas como las no problemáticas. Desde ya que las paradojas han de recibir una explicación satisfactoria pero también han de recibirla nuestros usos ordinarios de ‘verdadero’. Además, es menester que la teoría dé cuenta de la distinción entre ambas. En concordancia con la tesis de superveniencia semántica,⁴ la verdad o falsedad de las oraciones no problemáticas debe quedar completamente determinada

³Herzberger (1982a), p. 479. La traducción es nuestra.

⁴Véase la introducción, p. 3.

por hechos no semánticos. La teoría tiene que explicar entonces de manera sistemática cómo los valores de verdad de tales oraciones no problemáticas dependen de hechos no semánticos. Más aun, si bien los hechos no semánticos no necesariamente fijan un valor de verdad definitivo para todas las oraciones del lenguaje, sí determinan su carácter patológico y la teoría deberá explicar cómo esto se lleva a cabo. En resumen, la teoría de la verdad debe explicar nuestros juicios intuitivos concernientes a la verdad, falsedad y también a la patologicidad.

1.3 Una intuición fundamental

Gupta & Belnap (1993) reconocen la existencia de una intuición fundamental, la cual —naturalmente—habrá de ser resguardada: los bicondicionales T fijan el significado del concepto de verdad y son, por tanto, analíticos. Básicamente porque en conjunto fijan las condiciones bajo las cuales ‘verdadero’ se aplica a cualquier oración del lenguaje. Por tanto, las instancias del esquema T pueden ser consideradas definiciones parciales del concepto de verdad.

Como ya ha señalado Eduardo Barrio en el primer capítulo, cuando esta intuición se codifica por medio de la convención T y los bicondicionales involucrados son clásicos, la intuición fundamental conduce a paradojas. De acuerdo con Gupta & Belnap, ello no sugiere que la intuición sea incorrecta sino que debe ser reconstruida de modo diferente: “Un problema central en la teoría de la verdad es resolver la paradoja sin dañar la intuición fundamental de ningún modo esencial.”⁵ Las diferencias girarán en torno a lo que se entiende por ‘bicondicionales T’, ‘fijar el significado’ y ‘verdad’. La (re)construcción de la teoría de verdad nos conducirá en primer lugar a una interpretación adecuada de los bicondicionales T que desarrollaremos en la segunda sección: se ofrecerá una elucidación diferente del ‘si y sólo si’ involucrado en dichos bicondicionales. En segundo lugar, explicitaremos qué entender por ‘fijar el significado’ en 1.4 y en 1.5 mencionaremos algunos supuestos en torno a la noción de verdad.

1.4 La tesis de la significación

Tal como ha quedado estipulado en el primer capítulo, la convención T no sólo supone que los bicondicionales T allí involucrados son clásicos, sino también que el modo en que fijan el significado del predicado veritativo es determinando su extensión. La manera de conceptualizar dicha fijación es mediante la condición de que la definición de verdad ha de implicar los mencionados bicondicionales.

Como contrapunto, la lectura que Gupta & Belnap hacen de la intuición fundamental consiste en que los bicondicionales T constituyen ellos mismos una definición de la noción de verdad y, por tanto, fijan el significado de la verdad en todo mundo.⁶ En esto consiste la tesis de significación, la cual pretende capturar la mencionada intuición fundamental.

⁵Gupta & Belnap (1993), p. 6. La traducción es nuestra.

⁶Mundos que como veremos serán representados por lo que denominaremos ‘modelos base’.

Ahora bien, Gupta & Belnap argumentarán que dicho significado no es él mismo una regla de aplicación—como lo sería para cualquier concepto ordinario—sino una regla de revisión, pues la verdad no es un concepto ordinario sino circular. Frecuentemente se ha considerado que el carácter circular de un concepto lo torna ilegítimo. Gupta & Belnap, en cambio, argumentarán a favor de la legitimidad de las definiciones circulares.

1.5 Algunos *desiderata*

Ya hemos mencionado cuál es la motivación que subyace al proyecto. Ello constituye sin duda un punto de partida, como así también algunos supuestos adicionales que mencionaremos a continuación.

En primer lugar, la teoría de la verdad propuesta se restringe a lenguajes exentos de complicaciones tales como ambigüedad, vaguedad, indécicos y construcciones intensionales. Es concebida únicamente para lenguajes formalizados. A este supuesto subyace la sospecha de que esos factores no serán decisivos en la respuesta al problema tratado. El lenguaje estudiado puede ser visto como un lenguaje cuantificacional de primer orden clásico al que se le agregará un predicado de verdad con una semántica específica. Así el fragmento libre del predicado de verdad permanece completamente clásico.

Otros supuestos son tomas de posición respecto al concepto de verdad que se desea analizar. Se trata de ofrecer una definición de verdad que capture la intuición fundamental precisada a través de la tesis de significación, de modo que no podrá (ni se propone) capturar todos los usos de ‘verdadero’, sino sólo algunos de ellos. El primero de estos supuestos atañe a la gramática lógica del predicado veritativo. Siguiendo la perspectiva tarskiana, se asume que la verdad es un predicado que se aplica a oraciones. Adicionalmente, la tesis de significación no es una condición de adecuación apropiada para todas las nociones de verdad aplicables a oraciones, sino que concierne sólo a una noción lógica, débil y absoluta de verdad. En lo que sigue explicaremos brevemente estas tres notas.

Dado que queremos preservar la analiticidad del esquema T, todas sus instancias deberán ser verdaderas. Así, por ejemplo, el bicondicional “‘La nieve es blanca’ es verdadera si y sólo si la nieve es blanca” deberá ser necesariamente verdadero, *i.e.* deberá ser verdadero en todo mundo posible. Sin embargo, si permitiéramos que las interpretaciones de los términos no lógicos variaran en cada mundo considerado, bien podría ocurrir que en algún mundo—idéntico al nuestro por todo lo demás—‘nieve’ significara lechuga. En tal mundo el bicondicional no resultaría verdadero. Estamos entonces frente a dos nociones de verdad. Una noción no lógica según la cual determinamos si una oración es verdadera en un mundo atendiendo al significado que los términos no lógicos involucrados en la oración tienen en ese mundo. Y otra noción lógica según la cual determinamos si una oración es verdadera en un mundo considerando el significado que los términos no lógicos involucrados en esta oración tienen en el mundo actual. Esta última es la noción en la cual estamos interesados porque es aquella que permite sostener la analiticidad del esquema T. Es también la noción de verdad útil para la lógica y la filosofía del lenguaje.

Otra clasificación posible para las diversas nociones de verdad distingue entre una

noción débil y otra fuerte. De acuerdo con la primera, el estatus semántico de “ A es verdadera” hereda el estatus semántico de A y resulta por tanto una noción de verdad transparente. Por el contrario, desde una concepción fuerte de la verdad el estatus semántico de ambas puede diferir. Consideremos la siguiente afirmación:

El emperador del mundo juega a la ruleta. (1.1)

Frecuentemente se piensa que carece de valor de verdad clásico. Analicemos ahora esta otra oración:

(1.1) es verdadera. (1.2)

Desde una lectura fuerte, a diferencia de (1.1), (1.2) no carece de valor veritativo sino que es simplemente falsa, pues (1.2) no es verdadera. Como ilustra el ejemplo, la noción fuerte de verdad permite que el valor semántico de una oración y su predicación de verdad difieran. Por esa razón la tesis de significación resulta adecuada sólo para usos de verdad en que la noción expresada es débil, pues dicha tesis nos compromete con una noción transparente de verdad.

Por último, sabemos que la condición que se impone sobre la definición de verdad, lejos de mencionar relativización alguna a un modelo en particular, supone que los bicondicionales T han de ser verdaderos en todo mundo posible. Si ése es el caso, podrá sospecharse que tal condición sólo es apropiada para el caso en que la noción de verdad definida sea absoluta o *simpliciter*, como contrapuesta a la de verdad en un modelo que ya ha sido desarrollada y problematizada en el primer capítulo.

1.6 Verdad circular

Antes de pasar a la presentación formal de la teoría, ofreceremos una primera aproximación al tratamiento que la teoría de la revisión ofrece para la noción de verdad. En 1.4, anticipamos que la definición de la noción de verdad dará lugar a una regla de revisión y no de aplicación, dada la naturaleza circular de la verdad. A continuación desarrollaremos este punto.

Toda definición de un predicado monádico G tiene la siguiente forma:⁷

$$x \text{ es } G =_{df} x \text{ es } A \quad (1.3)$$

donde lo que se encuentra a la izquierda de la identidad es el *definiendum* y a la derecha el *definiens*.⁸ De acuerdo con (1.3) un objeto será un G si y sólo si es un A . De este modo la definición (1.3) daría lugar a una regla de aplicación, siempre que fuéramos

⁷Por cuestiones de simplicidad nos limitaremos aquí a predicados monádicos, pero el enfoque puede extenderse a predicados de cualquier aridad.

⁸Como veremos en el ejemplo que sigue, A puede ser lógicamente complejo.

capaces de determinar si un objeto es un A previa e independientemente de que sea un G . Así por ejemplo, dada la oración:

$$\text{Sócrates es un ser humano.} \quad (1.4)$$

y la definición de ser humano como animal racional:

$$x \text{ es un ser humano} =_{df} x \text{ es animal y } x \text{ es racional} \quad (1.5)$$

podemos afirmar que Sócrates es un ser humano, sabiendo previamente que es animal y es racional. La definición (1.5) puede utilizarse de modo similar para cualesquiera objetos. Sin embargo, esto será posible sólo en la medida en que podamos evaluar que algo es un animal y que es racional independientemente de si es un ser humano o no, lo cual parece plausible. En términos generales, para que una definición de la forma (1.3) funcione como regla de aplicación (como en (1.5)) el *definiendum* no debe aparecer de modo ineliminable en el *definiens*. Ése no siempre es el caso y prueba de ello son las definiciones circulares. Tales definiciones suelen ser excluidas como ilegítimas y los conceptos definidos por ellas considerados inexistentes. Gupta & Belnap, en cambio, consideran que sí es posible ofrecer una teoría consistente que incluya tales definiciones (y nociones). De hecho, la verdad se encuentra entre ellas.

De acuerdo con la tesis de significación el esquema T constituye una definición para la verdad. En términos de Gupta & Belnap éste resulta entonces en las instancias de:

$$'A' \text{ es verdadera} =_{df} A \quad (1.6)$$

dadas por cada oración A de un lenguaje determinado, donde ' A ' es un nombre de A . El bicondicional involucrado en el esquema T, en lugar de ser el bicondicional clásico, es pues el bicondicional distintivo utilizado para definir un concepto. Su semántica será por tanto diferente. Cada instancia de este esquema ofrece una definición parcial de la verdad y algunas de ellas resultarán circulares. Consideremos la siguiente oración:

$$\text{"La nieve es blanca" es verdadera} \text{ es verdadera.} \quad (1.7)$$

y su correspondiente instancia de (1.6):

$$\text{"La nieve es blanca" es verdadera} =_{df} \text{"La nieve es blanca" es verdadera} \quad (1.8)$$

De acuerdo con nuestra presentación preliminar, éste es un caso de circularidad, pues el *definiendum* aparece en el *definiens*. No obstante, la circularidad es eliminable, pues recurriendo a esta nueva instancia de (1.6):

$$\text{'La nieve es blanca' es verdadera} =_{df} \text{La nieve es blanca} \quad (1.9)$$

es posible evaluar el valor de verdad de “La nieve es blanca’ es verdadera” y, a partir de allí, evaluar el de (1.7). De este modo, sabemos si “La nieve es blanca’ es verdadera” cae o no bajo la extensión del predicado veritativo.

Sin embargo, existen oraciones donde este movimiento no es posible. Retomemos la oración del mentiroso introducida en capítulos anteriores:

$$(M) \text{ no es verdadera.} \quad (M)$$

Su instancia correspondiente de (1.6), el esquema T, parece tener un carácter circular no eliminable:

$$(M) \text{ es verdadera} =_{df} (M) \text{ no es verdadera} \quad (1.10)$$

En esta ocasión no existe un modo de obrar paso a paso para dar con el valor de verdad de ‘(M) es verdadera’; al menos si no disponemos de información adicional. De este modo, el conjunto de los bicondicionales que constituyen la definición de verdad no resulta suficiente para dar lugar a una regla de aplicación del predicado de verdad. No resulta suficiente para ofrecernos su extensión, pues tanto el concepto definido como su definición son circulares.

No obstante, no todo está perdido. Para determinar si (M) cae o no bajo el concepto de verdad, resulta suficiente suplementar los bicondicionales T con extensiones hipotéticas del concepto de verdad. Estas extensiones se irán revisando mediante el esquema T dado por (1.6), el cual funcionará entonces, no ya como regla de aplicación sino como una regla de revisión. Veamos entonces cómo funciona. Tomemos por caso las oraciones (M), y la siguiente:

$$(H) \text{ es verdadera.} \quad (H)$$

Como se ha mencionado en capítulos anteriores, (H) es conocida como la ‘oración del honesto’, pues dice de sí misma que es verdadera. Su instancia correspondiente del esquema T está dada por:

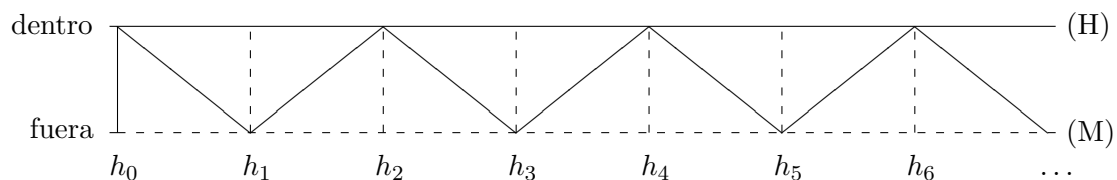
$$(H) \text{ es verdadera} =_{df} (H) \text{ es verdadera} \quad (1.11)$$

Atendamos al razonamiento informal que normalmente asociamos a (M) y (H). Para saber si ambas caen bajo la extensión del concepto de verdad, debemos evaluar el *definiens* de (1.10) y (1.11). Esta evaluación nos fuerza a determinar previamente si ellas mismas están o no en la extensión del predicado veritativo. Dado que no disponemos de esta información debemos hipotetizarla.

Se presentan cuatro opciones diferentes: o bien ambas pertenecen a la extensión, o bien ninguna, o bien sólo una de ellas. Llamemos a la primera ‘hipótesis 0’ y razonemos de acuerdo con ella. Dado que hemos supuesto que (M) es verdadera, el *definiens* en

(1.10) resulta falso y, por ende, no pertenece a la (nueva) extensión del predicado de verdad. En cuanto al honesto, como es verdadera, el *definiens* en (1.11) también lo es. En consecuencia, resulta verdadera nuevamente. Así, la hipótesis 0 da lugar a una nueva extensión hipotética del predicado de verdad que denominaremos ‘hipótesis 1’, la cual contiene a (H) pero ya no a (M). Esta hipótesis puede revisarse también. Ahora el *definiens* en (1.10) será verdadero, al igual que el *definiens* en (1.11), con lo cual ambas oraciones quedarán dentro de la extensión del concepto de verdad en la hipótesis 2, finalizando una nueva etapa de revisión.

Normalmente el argumento asociado a la oración del mentiroso termina aquí, precisamente con la conclusión de que resulta ser verdadera y falsa. No obstante, es evidente que la hipótesis 2 es idéntica a la hipótesis 0, la cual al ser revisada nos conduciría nuevamente a la hipótesis 1 y así sucesivamente, como se muestra en el siguiente gráfico de revisión:



La revisión, lejos de asentarnos en un valor, nos conduce a una oscilación entre la verdad y la falsedad de (M). Se observa así, no una inconsistencia en la noción de verdad, sino el fenómeno de “inestabilidad semántica” antes mencionado. La divergencia entre la semántica de ‘ $=_{df}$ ’ —profundamente intensional pues da lugar a una definición y fija por tanto el significado— y el bicondicional material bloquea el argumento del mentiroso hacia una contradicción y, al mismo tiempo, da cuenta del carácter patológico de esta oración.

El punto de partida ha sido sólo una de la hipótesis posibles. ¿Qué tanto podemos concluir a partir de allí? ¿No introduce ello arbitrariedad? Si prestamos atención únicamente a la oración del mentiroso, no parece ser así. Si hubiéramos comenzado con la hipótesis 1 en lugar de con la hipótesis 0, podríamos observar también la alternación entre ambas hipótesis, tan sólo que la secuencia de revisión hubiera comenzado en un punto diferente. Pero la situación no es la misma si atendemos a la oración del honesto. En este caso, la hipótesis con que partimos determina crucialmente el resultado del proceso de revisión. Si consideramos ahora otra hipótesis inicial que denominaremos ‘hipótesis 0’ en la cual (H) se encuentre fuera de la extensión del predicado de verdad, en una etapa de revisión sucesiva 1’ el *definiens* en (1.11) resultaría falso, permaneciendo así fuera de la extensión del predicado veritativo en la hipótesis 1’. Tal como podrá comprobar el lector, (H) queda fuera de la extensión de dicho predicado a lo largo de las sucesivas iteraciones del proceso de revisión también.

Vemos entonces que el proceso de revisión depende de modo crucial de la hipótesis con la que se decida comenzar. Por otra parte, podemos observar que el proceso de revisión puede concluir de distintas maneras. O bien nunca estabilizarse, como en el

caso de la oración del mentiroso que oscila entre verdadero y falso, o bien eventualmente hacerlo, ya sea en verdadero o en falso, como vimos en el caso del honesto, o bien siempre hacerlo en una y la misma de estas categorías, como es el caso de (1.7), cuya comprobación quedará a cargo del lector. Si al finalizar del proceso de revisión basado en cada hipótesis alternativa una oración se estabilizó siempre dentro de la extensión del predicado veritativo, diremos que es categóricamente verdadera. Si, en cambio, una oración se estabilizó siempre fuera de la extensión del predicado de verdad, diremos que es categóricamente falsa. Finalmente, si una oración no se comporta de ninguno de esos modos, diremos que es patológica.

A modo de resumen de esta primera sección, si los bicondicionales T son considerados definiciones parciales, la verdad debe ser un concepto circular. Lejos de tornarse ilegítimo o inexistente a causa de su circularidad, buena parte del comportamiento del concepto de verdad, tanto en casos ordinarios como en casos patológicos, puede ser explicado gracias a esta característica.

Ejercicios para la subsección 1.6

1. Considere la siguiente oración de Curry:

Si (C) es verdadera entonces los Reyes Magos existen. (C)

- Sabiendo que los Reyes Magos no existen, evalúe cómo resultará un proceso de revisión para (C) comenzando con cada una de las extensiones hipotéticas posibles para la noción de verdad. (Esto es, una a la cual (C) pertenece y otra a la que no.)
2. Evalúe cómo resultará un proceso de revisión de la hipótesis vacía (ninguna oración pertenece inicialmente a la extensión de la verdad) para las oraciones '(M) es verdadera', '(H) es verdadera', '(M) es verdadera o (H) es verdadera', '(M) es verdadera y (H) es verdadera' y 'Alguna de las oraciones a considerar en este ejercicio no es verdadera'. Confeccione un gráfico de revisión (como el que aparece en esta sección) para cada oración.
 3. Evalúe el comportamiento de las oraciones (M) y (C) en el proceso de revisión de la hipótesis según la cual ambas oraciones se encuentran dentro de la extensión del concepto de verdad inicialmente. Confeccione un gráfico de revisión para cada oración.
 4. Evalúe el comportamiento de la oración (1.7) en el proceso de revisión de una hipótesis que la contenga y de otra que no. Confeccione un gráfico de revisión en cada caso.

2 Una presentación formal de la teoría revisionista de la verdad

2.1 La teoría de las definiciones de Gupta & Belnap

Tal como anticipamos, la teoría revisionista de la verdad es el resultado de aplicar la teoría de la definición desarrollada por Gupta & Belnap (1993)⁹ a la noción de verdad. A continuación daremos brevemente una idea de cómo esta teoría funciona, para luego enfocar nuestra atención en su aplicación al concepto de verdad.

Sea L un lenguaje de primer orden que llamaremos ‘lenguaje base’ y G un predicado monádico que no le pertenece, introducido para expresar la nueva noción que se quiere definir.¹⁰ L^+ es el lenguaje que resulta de incorporar G a L . Introducimos entonces el nuevo concepto mediante una definición (o conjunto de definiciones parciales) para el predicado G que lo expresa de la siguiente forma:

$$Gx =_{df} A(x, G) \quad (2.1)$$

donde $A(x, G)$ es una fórmula de L^+ con a lo sumo una variable libre x y con posibles ocurrencias de G . Luego, la teoría permite definiciones circulares, lo cual la pone en condiciones de lidiar con la noción de verdad.

Puesto que (2.1) constituye una definición para G , debe fijar el significado de este predicado en todo modelo. Sea $\mathcal{M}$ un modelo clásico cualquiera para L .¹¹ Lo denominamos ‘modelo base’. Todo modelo base puede ser visto como una interpretación parcial para L^+ en la cual el único símbolo no lógico que carece de interpretación es G , el predicado a definir. El objetivo es dar con un conjunto de oraciones correctamente asertables de L^+ bajo la interpretación de los símbolos no lógicos de L dada por $\mathcal{M}$, en el cual G se comporte de acuerdo con la definición dada por (2.1).

Usualmente llevaríamos a cabo esta tarea buscando una expansión $(\mathcal{M}, S)$ de $\mathcal{M}$ para todo el lenguaje L^+ que asigne la extensión S a G respetando la definición dada por (2.1), de modo tal que los miembros de S sean aquellos elementos de $|\mathcal{M}|$, el dominio de $\mathcal{M}$, que satisfagan $A(x, G)$ de acuerdo con este modelo. Sin embargo, como hemos notado en la primera sección, $A(x, G)$ podría contener ella misma ocurrencias de G , y, por tanto, podría no bastar en cada caso con la información provista por el modelo base para decidir la extensión de G sino que sería preciso saber previamente algo de esta misma extensión.

Tal como hemos sugerido en la primera sección, debemos comenzar por una extensión hipotética arbitraria S_0 para G y luego revisarla de acuerdo con (2.1), obteniendo una

⁹Véase Gupta & Belnap (1993), cap. 5.

¹⁰De acuerdo con la teoría original, podríamos contar en lugar de G con cualquier cantidad finita de predicados no necesariamente monádicos a definir. Aquí optamos por un único predicado de aridad 1 por cuestiones de simplicidad relacionadas con nuestro proyecto general.

¹¹De acuerdo con la teoría original, $\mathcal{M}$ podría no ser clásico sino asignar tres o más valores de verdad a las oraciones de L . Aquí nos limitaremos a interpretaciones bivalentes para mostrar que es posible un lenguaje bivalente que contiene su propio predicado de verdad.

nueva extensión hipotética S_1 para G : pertenecerán a S_1 todos los miembros de $|\mathcal{M}|$ que satisfagan la fórmula $A(x, G)$, no de acuerdo a $\mathcal{M}$ sino a su expansión $(\mathcal{M}, S_0)$. Luego, para cada modelo base $\mathcal{M}$, (2.1) nos sirve de guía para la construcción de una *regla de revisión* $\delta_{\mathcal{M}}$ que asigna nuevas extensiones hipotéticas a extensiones S hipotetizadas previamente del siguiente modo:

$$\delta_{\mathcal{M}}(S) = \{d \in |\mathcal{M}| : (\mathcal{M}, S) \models A(\bar{d}, G)\}^{12} \quad (2.2)$$

donde $\bar{d}$ denota d de acuerdo con $\mathcal{M}$.

Mediante sucesivas aplicaciones de $\delta_{\mathcal{M}}$ mejoraremos nuestra hipótesis original, obteniendo mayor conocimiento acerca de la extensión del concepto definido por (2.1). Este proceso se denomina ‘proceso de revisión’ y la secuencia de modelos extendidos ‘secuencia de revisión’. Si a lo largo del proceso algunos elementos de $|\mathcal{M}|$ se estabilizan dentro de la extensión asignada a G , diremos que caen bajo el nuevo concepto en esa secuencia de hipótesis, mientras que de los que se estabilicen fuera diremos que no son G s en ese contexto.

Como anticipamos, es posible que estas estabilidades dependan en buena medida de la extensión de G hipotetizada al comienzo del proceso, y no sólo de $\mathcal{M}$ y (2.1). Para garantizar que exclusivamente (2.1) fije el significado de G en cada $\mathcal{M}$, debemos llevar a cabo un proceso de revisión con cada hipótesis inicial posible, esto es, con cada subconjunto de $|\mathcal{M}|$, y extraer consecuencias de todas las secuencias obtenidas simultáneamente. Como veremos más adelante, hay diferentes modos de extraer consecuencias a partir de allí.

A continuación haremos una presentación formal de la aplicación de la teoría de las definiciones recién esbozada a la noción de verdad.

Ejercicios para la subsección 2.1

1. Sea L un lenguaje de primer orden cuyos símbolos no lógicos son F y H y sea L^+ el resultado de incorporar a L un nuevo predicado G . Sea $\mathcal{M}$ un modelo de L tal que $|\mathcal{M}| = \{a, b, c\}$, $F^{\mathcal{M}} = \{a\}$ y $H^{\mathcal{M}} = \{a, c\}$. Considere la siguiente definición en L^+ .¹³

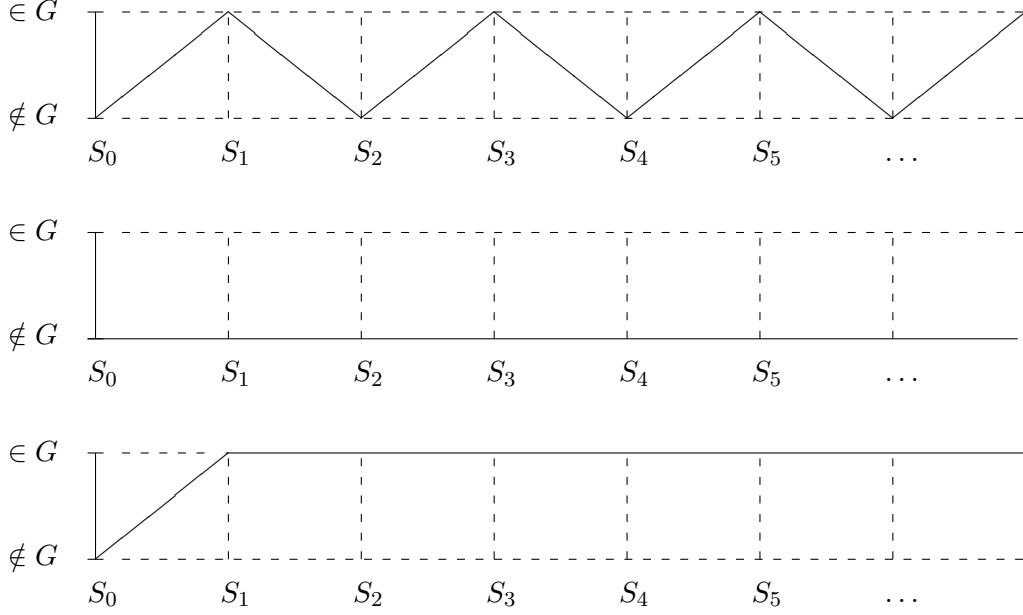
$$Gx =_{df} Fx \vee (Hx \wedge \neg Gx) \quad (2.3)$$

- (a) Escriba los respectivos primeros pasos de las secuencias generadas por el proceso de revisión de las hipótesis $S_0 = \emptyset$ y $S'_0 = \{a, b, c\}$.

¹²A diferencia de otros capítulos, en lugar de escribir $V_{\mathcal{M}}(A) = 1$ para indicar que A es verdadera en un modelo $\mathcal{M}$, emplearemos la siguiente notación por cuestiones de legibilidad: $\mathcal{M} \models A$.

¹³Este ejercicio ha sido extraído de un ejemplo ofrecido por Gupta & Belnap (1993), cap. 4.

- (b) Los gráficos siguientes representan la evolución de la pertenencia o no a la extensión de G de cada miembro de $|\mathcal{M}|$ a lo largo de las primeras etapas del proceso de revisión de S_0 . Indique a qué elemento corresponde cada uno:



2. Sea L un lenguaje de primer orden que contiene a '0' como constante de individuo y a '-' como símbolo diádico de función; y L^+ el lenguaje que resulta de incorporar a L un nuevo predicado monádico N . Sea $\mathcal{M}$ un modelo de L tal que $|\mathcal{M}| = \mathbb{Z}$, el conjunto de los números enteros, '0' es interpretado con el número 0 y '-' es interpretado como la resta. Considere la siguiente definición en L^+ :

$$Nx =_{df} x = 0 \vee N(x - 1) \quad (2.4)$$

- (a) ¿Cuál es la regla de revisión a la cual (2.4) da lugar?
- (b) Tras un proceso de revisión de la hipótesis vacía, ¿qué objetos del dominio de $\mathcal{M}$ resultan dentro de la extensión de N ?
- (c) Tras un proceso de revisión de la hipótesis $S = \mathbb{Z}$, ¿qué objetos del $|\mathcal{M}|$ resultan ser N ?
- (d) ¿Existe una hipótesis S que dé lugar a una extensión diferente de aquella que se obtiene en los incisos anteriores?

2.2 Una definición circular de 'verdad'

Gupta & Belnap (1993) aplican su teoría de la definición al concepto de verdad para cualquier lenguaje de primer orden que contenga nombres o nombres de códigos para

sus propias expresiones y sea capaz de expresar su propia sintaxis, dado que los objetos susceptibles de estar en la extensión del predicado veritativo son de índole lingüística. No obstante, aquí nos limitaremos al lenguaje L de la aritmética de primer orden y utilizaremos números de Gödel como códigos para las expresiones del lenguaje mismo, como es habitual. Las razones para hacerlo de este modo yacen principalmente en preservar la legibilidad del capítulo y en la costumbre de tomar la aritmética como base de cualquier teoría semántica a modo de ejemplo. De este modo, lo presentado aquí permitirá comparar la teoría revisionista de la verdad con muchas otras teorías semánticas y axiomáticas que aparecen en la literatura y, particularmente, con aquellas presentadas en este libro.

2.2.1 El lenguaje base y sus interpretaciones

Sea el lenguaje base L un lenguaje de predicados de primer orden con identidad con una única constante de individuo ‘0’, un único predicado diádico ‘<’ y una cantidad finita de símbolos de función para expresar las funciones recursivas que necesitaremos más adelante, incluyendo ‘ s ’, para la función monádica sucesor y ‘+’ y ‘ $\times$ ’ para las funciones diádicas suma y multiplicación, respectivamente. L^+ es el resultado de incorporar el predicado monádico Tr a L , predicado que expresará al finalizar el proceso la noción de verdad que queremos definir.

Sea $\mathcal{M}$ un modelo base, esto es, una interpretación de L . Escribimos ‘ $\bar{n}$ ’ como una abreviatura del término compuesto por 0 precedido por n ocurrencias de s . Lo denominamos ‘numeral de n ’, del $n \in |\mathcal{M}|$ que $\mathcal{M}$ asigna a $\bar{n}$ como su interpretación. Cuando no dé lugar a confusión omitiremos el guión sobre la n .

Al igual que en otros capítulos de este libro, emplearemos letras de imprenta mayúsculas A, B , etc. para nombrar o esquematizar fórmulas de L^+ , letras griegas minúsculas α, β , etc. para nombrar o esquematizar números ordinales y letras cursivas mayúsculas $\mathcal{M}, \mathcal{N}$, etc. para nombrar o esquematizar interpretaciones de L .

En cada modelo emplearemos miembros del dominio a los cuales les corresponde un numeral como códigos para las oraciones, fórmulas, términos y símbolos de L^+ , de modo tal que en toda interpretación el número de Gödel de cada uno de estos objetos lingüísticos A es el miembro del dominio que esta interpretación asigna a un numeral fijo ‘ A ’ del lenguaje. Por ejemplo, podemos decir que $\bar{m}$ es un nombre para A en L^+ , independientemente del modelo base $\mathcal{M}$ que tomemos, pues en cada caso seleccionamos el $m \in |\mathcal{M}|$ que $\mathcal{M}$ asigna a $\bar{m}$ como su interpretación.

2.2.2 Una Regla de Revisión

Puesto que deseamos que Tr sea un predicado de verdad, debemos proveer una definición o conjunto de definiciones parciales para la noción de verdad. Como dijimos en la introducción, el esquema T de Tarski constituye una definición de esta noción de acuerdo con Gupta & Belnap. En consecuencia, consideraremos las siguientes definiciones parciales:

$$Tr('A') =_{df} A \quad (2.5)$$

donde A es cualquier oración de L^+ . Dado un modelo base $\mathcal{M}$, el objetivo ahora es dar con el conjunto de oraciones de L^+ correctamente afirmables bajo la interpretación que ofrece $\mathcal{M}$ de los símbolos no lógicos de L y teniendo en cuenta que Tr , al cumplir con (2.5), debe comportarse como un predicado de verdad para todo el lenguaje L^+ . Idealmente, llevaríamos esto a cabo dando con una expansión $(\mathcal{M}, S)$ de $\mathcal{M}$ a todo L^+ en la cual S es la extensión asignada a Tr . Para ello, es preciso que en S estén todos los códigos de oraciones de L^+ que son verdaderas en $(\mathcal{M}, S)$, pero esto último depende a su vez de qué objetos de $|\mathcal{M}|$ pertenezcan a S . Esto se debe, como hemos visto, al carácter circular de (2.5).

Al igual que en el caso de (2.1), (2.5) fija el significado de Tr en cada modelo base $\mathcal{M}$ proveyendo una regla de revisión $\tau_{\mathcal{M}}$ que asigna a cada interpretación hipotética de Tr una nueva interpretación hipotética, a la cual pertenecen todos los códigos de oraciones que resultaban verdaderas en la interpretación anterior. En lugar de dar con una única expansión $(\mathcal{M}, S)$ para todo L^+ que haga de Tr un predicado veritativo, tenemos más bien una secuencia de expansiones, cada una de las cuales constituye una revisión de las anteriores.

DEFINICIÓN 3.1 ($\tau_{\mathcal{M}}$) $\tau_{\mathcal{M}}$ es una *regla de revisión* basada en $\mathcal{M}$ tal que $\tau_{\mathcal{M}} : \mathcal{P}(|\mathcal{M}|) \rightarrow \mathcal{P}(|\mathcal{M}|)$ ¹⁴ y

$$\tau_{\mathcal{M}}(S) = \{d : (\mathcal{M}, S) \models A \text{ y } d \text{ es el número de Gödel de } A\}$$

Notemos que el proceso de revisión deja fuera de la extensión de Tr a todos los miembros de $|\mathcal{M}|$ que no codifiquen una oración de L^+ , ya que son descartados por una sola aplicación de $\tau_{\mathcal{M}}$. Quisiéramos afirmar que en cada nueva aplicación de la regla obtenemos una extensión para el predicado veritativo más adecuada que la anterior, pero como veremos más adelante, éste no es siempre el caso.

2.3 Secuencias de Revisión

El proceso de revisión *generado por* una hipótesis inicial $S_0 \subseteq |\mathcal{M}|$ consiste en la aplicación sucesiva de la regla de revisión $\tau_{\mathcal{M}}$ a S_0 , y da lugar a una secuencia de revisión, esto es, cierta secuencia de extensiones posibles para Tr .

EJEMPLO 3.2 Tomemos la interpretación estándar de L —el modelo $\mathcal{N}$ en el cual $|\mathcal{N}| = \mathbb{N}$ y $0, <, ', +, \times$ y todos los símbolos de función restantes reciben su significado usual—como modelo base. Sea cual fuere la hipótesis inicial por la cual optemos, (el

¹⁴ $\mathcal{P}(|\mathcal{M}|)$ denota el conjunto potencia de $|\mathcal{M}|$.

código de)¹⁵ $0=0$ pertenecerá a $\tau_{\mathcal{N}}(S_0)$, la nueva extensión de Tr , por la definición 3.1, pues $(\mathcal{N}, S_0) \models 0 = 0$ ya que la oración pertenece a L y es verdadera en $\mathcal{N}$. Como esto último vale para cualquier expansión $(\mathcal{N}, S)$ de $\mathcal{N}$, $0 = 0$ permanecerá en la interpretación de Tr en cada paso ulterior del proceso de revisión.

$Tr('0 = 0')$ también permanecerá en la interpretación de Tr a lo largo de la secuencia de revisión, pero sólo a partir del segundo paso si $0 = 0 \notin S_0$. Análogamente, $Tr('Tr('0 = 0)')$ se estabilizará con certeza dentro de la extensión del predicado veritativo en la secuencia del tercer paso en adelante.

Sea $f : \mathbb{N}^2 \rightarrow \mathbb{N}$ tal que $f(x, 0) = x$ y $f(x, y + 1)$ es el código de $Tr(f(x, y))$. f es evidentemente una función recursiva. Sea $\dot{f}$ el símbolo de función diádico que la expresa en L y, por ende, en L^+ . $\dot{f}('A', 0) = 'A'$, $\dot{f}('A', 1) = 'Tr('A')'$, $\dot{f}('A', 2) = 'Tr('Tr('A')')'$, etc. son identidades verdaderas en $\mathcal{N}$. Por cuestiones de legibilidad, frecuentemente escribiremos $Tr^0('A')$ en lugar de A y $Tr^n('A')$ en lugar de $Tr(\dot{f}('A', n - 1))$ si $n > 0$.

Sabemos que $Tr^n('0 = 0')$ se estabilizará dentro de la extensión de Tr a lo sumo en el paso $n + 1$ del proceso de revisión. Al contrario, $0 = 1$ no pertenecerá a $\tau_{\mathcal{N}}(S_0)$ aún perteneciendo a S_0 , por la definición 3.1 nuevamente. Lo mismo podemos decir de $Tr^n('0 = 1')$ para todo $n \in \mathbb{N}$.

Queda claro a la luz de este ejemplo que las secuencias de revisión no deben comprender sólo un número finito m de pasos. Si culminamos el proceso de revisión en el paso m , no tenemos garantías de que todas las oraciones de la forma $Tr^n('0 = 0')$ estén dentro de la extensión del predicado veritativo al finalizar el proceso, sino sólo aquellas en las cuales $n < m$. Sin embargo, si nos comprometemos con $Tr^{m-1}('0 = 0')$ parece que lo mismo deberíamos hacer con $Tr^m('0 = 0')$, obedeciendo la definición del predicado de verdad. Es preciso que el proceso de revisión continúe al menos hasta alcanzar un ordinal límite. Veremos que ω —el primer ordinal límite—tampoco es suficiente.

EJEMPLO 3.3 Sea el modelo base como en el ejemplo 3.2 y consideremos la oración $\forall x Tr^{x+1}('0 = 0')$, que afirma que todas las iteraciones de Tr sobre $0 = 0$ son verdaderas. Si bien es intuitivamente verdadera, dado lo dicho en el párrafo anterior, sabemos que no contamos necesariamente con esta generalización dentro de la extensión de Tr en ningún estado finito de una secuencia de revisión basada en $\mathcal{N}$. No obstante, como cada una de sus instancias permanece allí a partir de algún paso finito, sabemos que en ω todas pertenecerán a la extensión de Tr y, por tanto, nuestra generalización será verdadera allí. Por ende, recién en $\omega + 1$ contaremos con $\forall x Tr^{x+1}('0 = 0')$ dentro de la extensión del predicado veritativo.

La siguiente tabla representa mediante los valores **t** y **f** el valor semántico de $0 = 0$, de cada $Tr^{n+1}('0 = 0')$ para todo $n \in \mathbb{N}$, $\forall x Tr^{x+1}('0 = 0')$ y $Tr(' \forall x Tr^{x+1}('0 = 0)')$ en

¹⁵Frecuentemente identificaremos las fórmulas de L^+ con sus respectivos códigos por cuestiones de legibilidad, si bien debe entenderse que sólo los códigos de oraciones pueden pertenecer a la extensión del predicado veritativo, no las oraciones mismas.

cada estado del proceso de revisión generado por la hipótesis $\emptyset$:¹⁶

	0	1	2	3	...	n	...	ω	$\omega + 1$	...
$0 = 0$	t	t	t	t	t	t	t	t	t	t
$Tr^1('0 = 0')$	f	t	t	t	t	t	t	t	t	t
$Tr^2('0 = 0')$	f	f	t	t	t	t	t	t	t	t
...										
$Tr^n('0 = 0')$	f	f	f	f	...	t	t	t	t	t
...										
$\forall x Tr^{x+1}('0 = 0')$	f	f	f	f	f	f	...	t	t	t
$Tr(\forall x Tr^{x+1}('0 = 0'))$	f	f	f	f	f	f	f	f	t	t

Consideremos ahora el proceso de revisión generado por $S = \{Tr(Tr('0 = 0'))\}$. En la etapa 1, $0 = 0$ ingresará a la extensión del predicado de verdad. En cambio, $Tr('0 = 0')$ saldrá de la extensión, pues $(\mathcal{N}, S) \not\models Tr('0 = 0')$, siendo que $0 = 0 \notin S$. Mientras que el ingreso de $0 = 0$ representa una mejoría de la extensión de Tr , el egreso de $Tr('0 = 0')$ implica claramente un desmejoramiento de la extensión de Tr . Este hecho no señala un defecto del proceso de revisión, pues este error se corregirá en una etapa posterior, sino que no siempre cada etapa sucesora en el proceso provee un mejor candidato para la extensión del concepto de verdad que su predecesora.¹⁷

Es preciso entonces que continuemos el proceso de revisión en niveles transfinitos. Por cuestiones de simplicidad, lidiaremos únicamente con secuencias de longitud On , *i.e.* la longitud de todos los ordinales. Esto nos permitirá ver de un modo unificado las diversas teorías de la verdad que Gupta & Belnap (1993) presentan.

Dar la extensión del predicado de verdad para pasos límite no es siempre sencillo. Como hemos visto, los bicondicionales de Tarski sirven muy bien para la formulación de una regla que indique cómo pasar de una extensión hipotética a su sucesora, pero no dicen demasiado acerca de qué debemos hacer en los límites, como ω . Es natural pensar que debemos acumular los datos obtenidos en todos los pasos anteriores de la secuencia: las oraciones que se han estabilizado dentro de la extensión de Tr deben permanecer allí en el paso límite, y las que se han estabilizado fuera deben permanecer fuera. Sin embargo, no toda oración del lenguaje cae siempre bajo una u otra de estas categorías, como vimos en la primera sección.

EJEMPLO 3.4 (El mentiroso) Tomemos nuevamente $\mathcal{N}$ como modelo base. Puesto que el lema de diagonalización es verdadero en esta interpretación sabemos que para

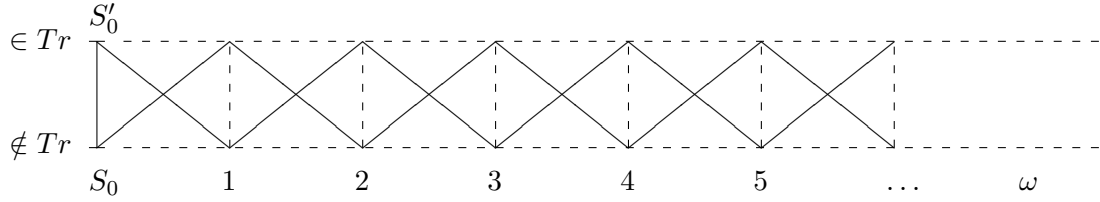
¹⁶Recuérdese que $Tr^{n+1}('0 = 0')$ y $\forall x Tr^{x+1}('0 = 0')$ son licencias notacionales para $Tr(\dot{f}('0 = 0', \bar{n}))$ y $\forall x Tr(\dot{f}('0 = 0', x))$, respectivamente.

¹⁷Un ejemplo similar es empleado por Belnap (1982) precisamente para notar que Gupta (1982) se equivoca cuando afirma que cada etapa provee un mejor candidato para la extensión de la verdad que la etapa anterior.

todo $S \subseteq |\mathcal{N}|$ existe una oración M del mentiroso en L^+ , equivalente a una oración que dice que M no es verdadera, esto es:

$$(\mathcal{N}, S) \models M \equiv \neg Tr('M') \quad (2.6)$$

Al igual que en la contraparte informal desarrollada en 1.6, si comenzamos el proceso de revisión con una hipótesis S_0 tal que $M \notin S_0$, $(\mathcal{N}, S_0) \models \neg Tr('M')$ y, por (2.6), $(\mathcal{N}, S_0) \models M$. Por la definición 3.1, $M \in \tau_{\mathcal{N}}(S_0)$. Pero entonces sabemos que $(\mathcal{N}, \tau_{\mathcal{N}}(S_0)) \models Tr('M')$ y, en virtud de (2.6), $(\mathcal{N}, \tau_{\mathcal{N}}(S_0)) \not\models M$, con lo cual $M \notin \tau_{\mathcal{N}}(\tau_{\mathcal{N}}(S_0))$ y volvemos a donde empezamos con respecto a esta oración. M ingresa a la extensión de Tr en los pasos impares y egresa en los pares. Si, por el contrario, hubiéramos empezado con una hipótesis S'_0 tal que $M \in S'_0$, el resultado habría sido simétricamente opuesto: M estaría en la extensión de Tr en los estados pares y fuera en los impares. El siguiente gráfico representa la evolución de la pertenencia de M a la extensión de Tr en cada estado del proceso de revisión generado por las hipótesis S_0 y S'_0 :



Oraciones como M no se estabilizarán ni dentro ni fuera de la extensión de Tr en ningún punto de la secuencia previo a cada ordinal límite. ¿Qué hacemos con ellas allí? En la literatura encontramos diversas propuestas rivales, llamadas ‘políticas de bootstrapping’,¹⁸ las cuales discutiremos más adelante. Por el momento nos detendremos en la política elegida por Gupta & Belnap (1993), que consiste en dejar total libertad a estas oraciones en pasos límites: se las puede colocar dentro o fuera de la extensión de Tr , según se desee. Luego, no basta con la indicar un modelo base y una hipótesis inicial para determinar todos los estados de una secuencia de revisión, sino que es preciso adicionalmente saber qué se hará con cada una de las oraciones inestables en cada estado límite. Esto último estará dado en cada secuencia por una función $\mathfrak{B}$ que asigna a cada ordinal límite λ un conjunto $\mathfrak{B}(\lambda) \subseteq |\mathcal{M}|$ llamado ‘bootstrapper’, donde podrán estar o no las oraciones que no se han estabilizado antes de λ .

DEFINICIÓN 3.5 (Secuencias de revisión) Si $\mathcal{M}$ es un modelo base y $\mathfrak{B}$ una función que asigna un conjunto $\mathfrak{B}(\lambda) \subseteq |\mathcal{M}|$ a cada ordinal límite λ , $[\mathcal{M}, S, \mathfrak{B}]$ es una *secuencia de revisión* basada en $\mathcal{M}$ si y sólo si asigna a cada ordinal α un conjunto $S_\alpha \subseteq |\mathcal{M}|$ tal que:

1. si $\alpha = 0$, $S_\alpha = S$;

¹⁸La introducción del término en la discusión corresponde a Belnap (1982). Hemos decidido no traducirlo ni italicarlo por tratarse de un término técnico.

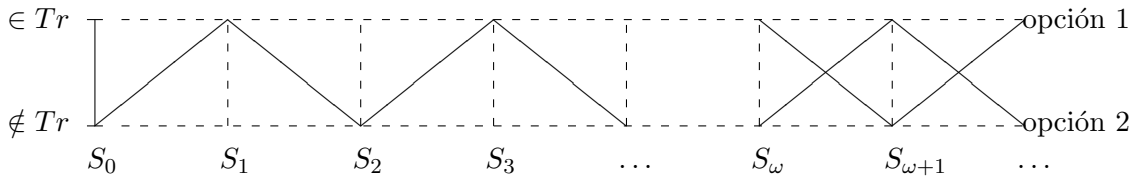
2. si $\alpha = \beta + 1$ para algún β , $S_\alpha = \tau_{\mathcal{M}}(S_\beta)$;
3. si α es un ordinal límite, $S_\alpha = (S_\alpha^\top \cup \mathfrak{B}(\alpha)) - S_\alpha^\perp$, donde

$$S_\alpha^\top = \{d \in |\mathcal{M}| : \exists \beta < \alpha \forall \gamma (\beta \leq \gamma < \alpha \Rightarrow d \in S_\gamma)\}$$

$$S_\alpha^\perp = \{d \in |\mathcal{M}| : \exists \beta < \alpha \forall \gamma (\beta \leq \gamma < \alpha \Rightarrow d \notin S_\gamma)\}$$

La cláusula 1 de esta definición nos indica que en una secuencia de revisión $[\mathcal{M}, S, \mathfrak{B}]$ S debe ser la hipótesis con la cual comencemos el proceso de revisión. La segunda cláusula nos señala cómo obtener la extensión de Tr en pasos sucesores a partir del paso anterior: empleando la regla de revisión $\tau_{\mathcal{M}}$. La cláusula 3 nos muestra cómo obtener la extensión de Tr en estados límites: ésta consiste de todas las oraciones de L^+ que se estabilizaron previamente dentro de la extensión de Tr (los miembros de $S_\alpha^\top$), más las oraciones que el bootstraper provisto por $\mathfrak{B}$ agregue en este caso (los miembros de $\mathfrak{B}(\alpha)$), menos aquellas oraciones que se estabilizaron previamente fuera de la extensión de Tr (los miembros de $S_\alpha^\perp$). Luego, ante la presencia de oraciones inestables como M , aún especificando un modelo base $\mathcal{M}$ y una hipótesis inicial S , no existe una única secuencia de revisión basada en $\mathcal{M}$ y generada por S sino varias, según las elecciones que tomemos para aquellas oraciones en cada ordinal límite. Sin embargo, por cuestiones de legibilidad, en aquellos casos en los cuales no sea relevante esta elección omitiremos $\mathfrak{B}$ y hablaremos de las secuencias mediante la especificación del modelo base y la hipótesis inicial únicamente, esto es, utilizaremos la notación $[\mathcal{M}, S]$.

EJEMPLO 3.6 Retomando el ejemplo 3.4, de acuerdo con la definición 3.5, podemos optar por incorporar M a $\mathfrak{B}(\omega)$ o no hacerlo. Puesto que M no se estabiliza ni dentro ni fuera de la extensión de Tr en un estado previo a ω , siguiendo el primer camino M quedará dentro de S_ω , mientras que en el segundo quedará fuera. En cambio, como $0 = 1$ se estabiliza fuera de la extensión de Tr en un estado previo a ω , $0 = 1 \in S_\omega^\perp$, con lo cual aún incorporándola a $\mathfrak{B}(\omega)$ tendremos que $0 = 1 \notin S_\omega$. El siguiente gráfico representa las dos elecciones posibles para M en el estado ω , dando lugar a dos secuencias diferentes basadas en $\mathcal{N}$ y generadas por $\emptyset$:



Ejercicios para la subsección 2.3

1. Sea M como en (2.6) y sea C una oración de Curry tal que, para todo $S \subseteq |\mathcal{N}|$:

$$(\mathcal{N}, S) \models C \equiv (Tr('C') \supset M) \quad (2.7)$$

- (a) Indique una secuencia de revisión $[\mathcal{N}, S, \mathfrak{B}]$ (una hipótesis inicial S y $\mathfrak{B}(\lambda)$ para cada ordinal límite λ) en la cual M y C no estén nunca ambas dentro o ambas fuera de la extensión del predicado de verdad hasta ω y que, de allí en más, suceda lo contrario.
 - (b) Indique otra secuencia de revisión $[\mathcal{N}, S', \mathfrak{B}]$ en la cual la oración $M \equiv C$ sea verdadera en cada paso $((\mathcal{M}, S_\alpha) \models M \equiv C$ para todo ordinal α).
 - (c) ¿Cuántas secuencias de revisión diferentes en cuanto a M y C generadas por $S = \emptyset$ existen?
2. Considere las oraciones M , $Tr('M')$, $Tr('Tr('M')')$, etc., esto es, todas las instancias de $Tr^n('M')$ con $n \in \mathbb{N}$. Evalúe su comportamiento a lo largo de una secuencia de revisión basada en $\mathcal{N}$ en la cual la hipótesis inicial esté dada por $\{Tr^n('M') : n \in \mathbb{N}\}$.

2.4 Las teorías $\mathbf{T}^*$, $\mathbf{T}^\#$ y $\mathbf{T}^c$

Ante la presencia de oraciones como M las secuencias de revisión no convergen hacia una extensión definitiva del predicado veritativo sino que continúan oscilando a lo largo de todos los niveles de ordinales.¹⁹ Sin embargo, esto no nos impide obtener resultados categóricos para ciertas oraciones, *i.e.* aquellas que logran estabilizarse de algún modo a lo largo del proceso. Es importante notar que para extraer este tipo de conclusiones no basta con considerar una única secuencia de revisión para cada modelo base.

EJEMPLO 3.7 (El honesto) Sea $\mathcal{N}$ nuestro modelo base. En virtud del lema de diagonalización sabemos que para todo $S \subseteq |\mathcal{N}|$ existe una oración H del honesto en L^+ , equivalente a una oración que dice que H es verdadera, esto es:

$$(\mathcal{N}, S) \models H \equiv Tr('H') \quad (2.8)$$

Si bien esta oración no es intuitivamente paradójica como sí lo es M , es claramente patológica como notamos en la primera sección, pues su valor de verdad depende exclusivamente de sí misma. Si consideramos que es verdadera entonces será verdadera, y si consideramos que es falsa entonces será falsa; nada en $\mathcal{N}$ nos indica qué camino debemos tomar. Luego, a la luz de la tesis de la superveniencia parecería que no debiéramos tomar ninguno.

Consideremos la secuencia $[\mathcal{N}, \emptyset]$. Luego, $(\mathcal{N}, S_0) \not\models Tr('H')$ y, por (2.8), $(\mathcal{N}, S_0) \not\models H$, con lo cual $H \notin S_1$ por la definición 3.5. Iterando este razonamiento una y otra vez concluimos que $H \notin S_n$ para todo $n \in \mathbb{N}$. Luego, $H \in S_\omega^\perp$ y, por la definición 3.5, $H \notin S_\omega$: el proceso continua del mismo modo, excluyendo a H de la extensión de Tr .

¹⁹En cambio, si el lenguaje no es lo suficientemente expresivo como para permitir autorreferencia, las secuencias de revisión se estabilizarán en un conjunto que será la extensión del predicado de verdad. Para una demostración de este resultado, véase Gupta & Belnap (1993), cap. 6, sección B).

En general, $H \notin S_\alpha$ para cualquier α . Por tanto, la oración $\neg H$ es verdadera a lo largo de toda la secuencia.

Consideremos ahora la secuencia $[\mathcal{N}, \{H\}]$. Empleando un razonamiento análogo al del párrafo anterior, sabemos que H es verdadera en todos los estados de esta secuencia, mientras que $\neg H$ es falsa.

Sólo considerando ambas secuencias presentadas en este ejemplo podemos advertir que no es correcto concluir ni H ni $\neg H$. Juntas muestran que H es una oración infundada, pues no es exclusivamente el modelo base $\mathcal{N}$ el que determina el valor de verdad de H sino que éste depende de la extensión hipotetizada inicialmente para Tr . Es preciso atender a todas las secuencias de revisión basadas en un modelo para obtener conclusiones definitivas adecuadas acerca del concepto de verdad basado en ese modelo. Gupta & Belnap (1993) presentan tres alternativas para extraer este tipo de conclusiones, cada una de las cuales da lugar a una teoría de la verdad diferente: $\mathbf{T}^*$, $\mathbf{T}^\#$ y $\mathbf{T}^c$.²⁰

2.4.1 La teoría $\mathbf{T}^*$

Este sistema es quizás el más natural de los tres. El concepto central en $\mathbf{T}^*$ es el de *estabilidad*.

DEFINICIÓN 3.8 (Estabilidad) Si $[\mathcal{M}, S]$ es una secuencia de revisión y A es una oración de L^+ :

- A es *establemente verdadera* en $[\mathcal{M}, S]$ si y sólo si existe un ordinal α tal que, para todo $\beta \geq \alpha$, $(\mathcal{M}, S_\beta) \models A$;
- A es *establemente falsa* en $[\mathcal{M}, S]$ si y sólo si existe un ordinal α tal que, para todo $\beta \geq \alpha$, $(\mathcal{M}, S_\beta) \not\models A$;
- A es *estable* en $[\mathcal{M}, S]$ si y sólo si A es establemente verdadera o establemente falsa en $[\mathcal{M}, S]$.

DEFINICIÓN 3.9 (Validez en $\mathbf{T}^*$) Si A es una oración de L^+ y $\mathcal{M}$ un modelo base:

- A es *válida en $\mathbf{T}^*$ en $\mathcal{M}$* ($\mathcal{M} \models^* A$) si y sólo si A es establemente verdadera en toda secuencia de revisión basada en $\mathcal{M}$;
- A es *válida en $\mathbf{T}^*$* ($\models^* A$) si y sólo si, para todo $\mathcal{M}$, A es válida en $\mathbf{T}^*$ en $\mathcal{M}$.

Sea $\mathbf{V}_{\mathcal{M}}^*$ el conjunto de todas las oraciones válidas en $\mathbf{T}^*$ en $\mathcal{M}$. $\mathbf{V}_{\mathcal{M}}^*$ contiene aquellas oraciones de L^+ que en cada secuencia de revisión basada en $\mathcal{M}$ logran estabilizarse dentro de la extensión del predicado veritativo a partir de algún punto.

²⁰ $\mathbf{T}^*$ fue introducida por Belnap (1982), mientras que $\mathbf{T}^c$ pertenece originalmente a Gupta (1987). $\mathbf{T}^\#$, en cambio, aparece por primera vez en Gupta & Belnap (1993).

Cada uno de los tres sistemas desarrollados por Gupta & Belnap constituye una alternativa para la extracción de resultados categóricos del conjunto de posibles procesos de revisión generados por diversas hipótesis sobre un modelo dado. Dado un modelo base, cada teoría induce una clasificación de las oraciones de L^+ en siete categorías exhaustivas y mutuamente excluyentes.

DEFINICIÓN 3.10 (Categorías semánticas en $\mathbf{T}^*$) Si A es una oración de L^+ y $\mathcal{M}$ un modelo base:

- A es *categoricamente verdadera en $\mathbf{T}^*$ en $\mathcal{M}$* si y sólo si A es establemente verdadera en toda secuencia de revisión basada en $\mathcal{M}$;
- A es *categoricamente falsa en $\mathbf{T}^*$ en $\mathcal{M}$* si y sólo si A es establemente falsa en toda secuencia de revisión basada en $\mathcal{M}$;
- A es *paradójica en $\mathbf{T}^*$ en $\mathcal{M}$* si y sólo si A es inestable en toda secuencia de revisión basada en $\mathcal{M}$;
- A es *tf-caprichosa en $\mathbf{T}^*$ en $\mathcal{M}$* si y sólo si A es estable en toda secuencia de revisión basada en $\mathcal{M}$, pero establemente verdadera en al menos una y establemente falsa en al menos otra;
- A es *tp-caprichosa en $\mathbf{T}^*$ en $\mathcal{M}$* si y sólo si A no es establemente falsa en ninguna secuencia de revisión basada en $\mathcal{M}$, pero establemente verdadera en al menos una e inestable en al menos otra;
- A es *fp-caprichosa en $\mathbf{T}^*$ en $\mathcal{M}$* si y sólo si A no es establemente verdadera en ninguna en secuencia de revisión basada en $\mathcal{M}$, pero establemente falsa en al menos una e inestable en al menos otra;
- A es *tfp-caprichosa en $\mathbf{T}^*$ en $\mathcal{M}$* si y sólo si A es establemente verdadera en al menos una secuencia de revisión basada en $\mathcal{M}$, establemente falsa en al menos otra e inestable en al menos una tercera. $\square$

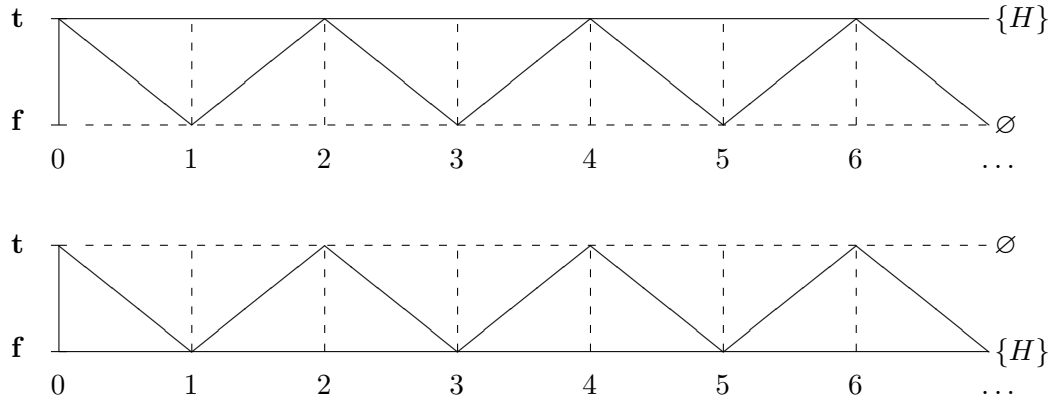
Llamamos caprichosas a aquellas oraciones que se exhiben comportamientos diferentes según la hipótesis inicial que se elija para generar la secuencia. Las primeras dos categorías semánticas incluyen, de acuerdo con Gupta & Belnap, aquellas oraciones categóricas, esto es, no patológicas, mientras que las cinco restantes constituyen una clasificación de las oraciones patológicas según el tipo de patologicidad que exhiben. Mediante esta subdivisión la teoría revisionista de la verdad logra dar una explicación adecuada de oraciones como el mentiroso y el honesto, en lugar de meramente incluirlas dentro de la categoría de lo problemático, como la teoría de punto fijo de Kripke.²¹

²¹Véase Kripke (1975) y el segundo capítulo de este libro para una explicación detallada.

EJEMPLO 3.11 (Un honesto alternativo) Sea $\mathcal{N}$ el modelo base. Consideremos las oraciones $0 = 0$, $0 = 1$, M , H , $M \vee H$, $M \wedge H$. Como es sabido, la clase de puntos fijos en $\mathcal{N}$ de un predicado monádico de L^+ como Tr no es recursiva. Por lo tanto, tenemos certeza de la existencia de una nueva oración H' del honesto independiente de (o, lo que es lo mismo, no equivalente a) H tal que para todo $S \subseteq |\mathcal{N}|$:

$$(\mathcal{N}, S) \models H' \equiv Tr('H')$$
(2.9)

$0 = 0$ es un ejemplo de oración categóricamente verdadera en $\mathcal{N}$, pues es establemente verdadera en toda secuencia basada en este modelo. $0 = 1$ es, en cambio, categóricamente falsa por ser establemente falsa en toda secuencia. M resulta paradójica, ya que oscila en cada secuencia de revisión, como muestra el gráfico del ejemplo 3.6. H , por su parte, pertenece a la categoría de las oraciones tf-caprichosas, pues es establemente verdadera en algunas secuencias y establemente falsa en las restantes, como señalamos en el ejemplo 3.7. Los siguientes gráficos representan la evolución de los valores semánticos de $M \vee H$ y $M \wedge H$, respectivamente, a lo largo de las secuencias de revisión generadas por las hipótesis $\emptyset$ y $\{H\}$:



Es sencillo ver que en toda secuencia $M \vee H$ o bien oscilará o bien será establemente verdadera, mientras que $M \wedge H$ oscilará o será establemente falsa. Luego, la primera es tp-caprichosa y la segunda fp-caprichosa de acuerdo con la definición 3.10. $H' \vee (H \wedge M)$, en cambio, es ttfp-caprichosa, como podrá comprobar el lector al realizar los ejercicios para esta subsección.

2.4.2 La teoría $\mathbf{T}^\#$

El concepto central en $\mathbf{T}^\#$ es el de *estabilidad cercana*, en lugar del de estabilidad *simpliciter*. Como veremos más adelante, si bien no resulta tan natural como $\mathbf{T}^*$ y tiene algunas consecuencias indeseables, este sistema valida muchos principios deseables para el predicado de verdad que los otros sistemas no validan.

DEFINICIÓN 3.12 (Estabilidad cercana) Si $[\mathcal{M}, S]$ es una secuencia de revisión y A es una oración de L^+ :

- A es *cercanamente establemente verdadera* en $[\mathcal{M}, S]$ si y sólo si existe un ordinal α tal que, para todo $\beta \geq \alpha$ hay un número natural m tal que, para todo $n \geq m$, $(\mathcal{M}, S_{\beta+n}) \models A$;
- A es *cercanamente establemente falsa* en $[\mathcal{M}, S]$ si y sólo si existe un ordinal α tal que, para todo $\beta \geq \alpha$ hay un número natural m tal que, para todo $n \geq m$, $(\mathcal{M}, S_{\beta+n}) \not\models A$;
- A es *cercanamente estable* en $[\mathcal{M}, S]$ si y sólo si A es *cercanamente establemente verdadera* o *cercanamente establemente falsa* en $[\mathcal{M}, S]$.

DEFINICIÓN 3.13 (Validez en $\mathbf{T}^\#$) Si A es una oración de L^+ y $\mathcal{M}$ un modelo base:

- A es *válida en $\mathbf{T}^\#$ en $\mathcal{M}$* ($\mathcal{M} \models^\# A$) si y sólo si A es *cercanamente establemente verdadera* en toda secuencia de revisión basada en $\mathcal{M}$;
- A es *válida en $\mathbf{T}^\#$* ($\models^\# A$) si y sólo si, para todo $\mathcal{M}$, A es *válida en $\mathbf{T}^\#$ en $\mathcal{M}$* .

Sea $\mathbf{V}_{\mathcal{M}}^\#$ el conjunto de todas las oraciones válidas en $\mathbf{T}^\#$ en $\mathcal{M}$. $\mathbf{V}_{\mathcal{M}}^\#$ contiene aquellas oraciones de L^+ que o bien son establemente verdaderas en cada secuencia de revisión o bien oscilan a lo largo de todo el proceso pero únicamente un número finito de pasos después de cada ordinal límite, para luego estabilizarse nuevamente. Puesto que toda oración establemente verdadera es también *cercanamente establemente verdadera*, obtenemos la siguiente relación entre los dos sistemas presentados hasta ahora.

TEOREMA 3.14 (Relaciones entre $\mathbf{T}^*$ y $\mathbf{T}^\#$) $\mathbf{V}_{\mathcal{M}}^* \subseteq \mathbf{V}_{\mathcal{M}}^\#$ para todo modelo base $\mathcal{M}$.

Prueba: Si $A \in \mathbf{V}_{\mathcal{M}}^*$, por la definición 3.9, A es establemente verdadera en toda secuencia de revisión $[\mathcal{M}, S]$. Y por la definición 3.8, en cada una de estas secuencias existe un α tal que, para todo $\beta \geq \alpha$, $(\mathcal{M}, S_\beta) \models A$. Luego, para todo $\beta \geq \alpha$ y para todo natural $n \geq 0$, $(\mathcal{M}, S_{\beta+n}) \models A$, con lo cual para todo $\beta \geq \alpha$ existe un número natural m tal que, para todo natural $n \geq m$, $(\mathcal{M}, S_{\beta+n}) \models A$. Por la definición 3.12, A es *cercanamente establemente verdadera* en cada $[\mathcal{M}, S]$. Y por la definición 3.13, $A \in \mathbf{V}_{\mathcal{M}}^\#$. ■

Dado un modelo base, $\mathbf{T}^\#$ también induce una clasificación de las oraciones de L^+ en siete categorías. Éstas son análogas a las presentadas en la definición 3.10, excepto que en lugar de hablar de oraciones establemente verdaderas o falsas define las categorías en función de la noción de estabilidad cercana. Como veremos más adelante, esta diferencia se ve reflejada en una categorización diferente de las oraciones de L^+ .

2.4.3 La teoría $\mathbf{T}^c$

Al igual que $\mathbf{T}^*$, $\mathbf{T}^c$ opta por un concepto de validez basado en la noción de estabilidad; pero, a diferencia de $\mathbf{T}^*$, no considera todas las secuencias de revisión al evaluar la estabilidad de una oración sino sólo aquellas que son *máximamente consistentes*. Como veremos más adelante, si bien no validará tantos principios deseables como $\mathbf{T}^\#$, $\mathbf{T}^c$ valida unos cuantos principios más para Tr que $\mathbf{T}^*$.

DEFINICIÓN 3.15 (Máxima consistencia) Una secuencia de revisión $[\mathcal{M}, S]$ es *máximamente consistente* si y sólo si, para todo α , S_α es máximamente consistente.²²

Una secuencia de revisión es máximamente consistente cuando la extensión que en cada paso asigna al predicado veritativo lo es. En estados sucesores $\alpha + 1$ esta condición es satisfecha automáticamente, pues $S_{\alpha+1}$ contiene a todas las oraciones verdaderas en el modelo $(\mathcal{M}, S_\alpha)$ generado por la extensión asignada a Tr en el paso anterior. Sin embargo, la condición puede fallar en estados iniciales y estados límites, en los cuales es posible tanto que una oración y su negación pertenezcan a la extensión del predicado de verdad como que ninguna lo haga. Por ejemplo, existen secuencias de revisión cuyo estado inicial es el conjunto vacío. Asimismo, en un estado límite es posible incluir en la extensión de Tr una oración inestable y su negación. El sistema $\mathbf{T}^c$ deja de lado estas posibles secuencias de revisión en las cuales las extensiones asignadas a Tr en ciertos estados no son máximamente consistentes. Al lidiar con oraciones del mentiroso como M en (2.6), en los estados límites caben sólo dos posibilidades: o bien incluimos M y excluimos $\neg M$ de la extensión de Tr o bien, a la inversa, excluimos M e incluimos su negación.

DEFINICIÓN 3.16 (Validez en $\mathbf{T}^c$) Si A es una oración de L^+ y $\mathcal{M}$ un modelo base:

- A es *válida en $\mathbf{T}^c$ en $\mathcal{M}$* ($\mathcal{M} \models^c A$) si y sólo si A es establemente verdadera en toda secuencia de revisión máximamente consistente basada en $\mathcal{M}$;
- A es *válida en $\mathbf{T}^c$* ($\models^c A$) si y sólo si, para todo $\mathcal{M}$, A es válida en $\mathbf{T}^c$ en $\mathcal{M}$.

Sea $\mathbf{V}_{\mathcal{M}}^c$ el conjunto de todas las oraciones válidas en $\mathbf{T}^c$ en $\mathcal{M}$. Sólo pertenecen a $\mathbf{V}_{\mathcal{M}}^c$ aquellas oraciones establemente verdaderas en toda secuencia de revisión máximamente consistente. Puesto que las secuencias consideradas por esta teoría son un subconjunto de aquellas consideradas por $\mathbf{T}^*$, obtenemos el siguiente resultado.

TEOREMA 3.17 (Relaciones entre $\mathbf{T}^*$ y $\mathbf{T}^c$) $\mathbf{V}_{\mathcal{M}}^* \subseteq \mathbf{V}_{\mathcal{M}}^c$ para todo modelo base $\mathcal{M}$.

Prueba: A cargo del lector, en los ejercicios para esta subsección. ■

²²Como es usual, dado un lenguaje llamamos *máximamente consistentes* a aquellos conjuntos consistentes de oraciones de este lenguaje tales que si se les incorpora una nueva oración se tornan inconsistentes.

Al igual que las otras teorías, dado un modelo base $\mathbf{T}^c$ también da lugar a una clasificación de las oraciones de L^+ en siete categorías semánticas análogas a las presentadas en la definición 3.10. Como es de esperar, la diferencia está dada por el conjunto de secuencias en las cuales se evalúa el comportamiento de cada oración: no son todas, sino sólo las máximamente consistentes.

Ejercicios para la subsección 2.4

1. Confeccione un gráfico o una tabla en la cual quede claro que la oración $H' \vee (H \wedge M)$ es tfp-caprichosa en $\mathbf{T}^*$ en $\mathcal{N}$.
2. Ofrezca una prueba del teorema 3.17.
3. Por el lema de diagonalización, recurriendo a la función f del ejemplo 3.2, podemos formular la *oración de McGee*,²³ equivalente a una fórmula que dice que alguna iteración de Tr sobre la oración de McGee no es verdadera en todo $S \subseteq |\mathcal{N}|$:

$$(\mathcal{N}, S) \models E \equiv \neg \forall x Tr(\dot{f}('E', x)) \quad (2.10)$$

- (a) Indique el progreso de E a lo largo de las secuencias de la forma $[\mathcal{N}, \emptyset, \mathfrak{B}]$, para cualquier $\mathfrak{B}$. Indique también qué sucede especialmente en aquel subconjunto de secuencias máximamente consistentes.
- (b) ¿ $E \in \mathbf{V}_{\mathcal{N}}^*$? ¿ $E \in \mathbf{V}_{\mathcal{N}}^\#$? ¿ $E \in \mathbf{V}_{\mathcal{N}}^c$?

2.5 Algunos resultados sobre los tres sistemas

Mencionamos a continuación algunos resultados relevantes sobre los tres sistemas presentados.

TEOREMA 3.18 (Consecuencia lógica clásica) Si A es una oración de L^+ y $\mathcal{M}$ es un modelo base y $\mathcal{M} \models A$, entonces $A \in \mathbf{V}_{\mathcal{M}}^*$, $A \in \mathbf{V}_{\mathcal{M}}^\#$ y $A \in \mathbf{V}_{\mathcal{M}}^c$.

Prueba: Si $\mathcal{M} \models A$ entonces $(\mathcal{M}, S) \models A$ para todo $S \subseteq |\mathcal{M}|$. Luego, para toda secuencia de revisión $[\mathcal{M}, S]$ existe un α (i.e., 0) tal que, para todo $\beta \geq \alpha$, $(\mathcal{M}, S_\beta) \models A$. Por la definición 3.8, A es establemente verdadera en toda secuencia. Y, por la definición 3.9, $A \in \mathbf{V}_{\mathcal{M}}^*$. Finalmente, por los teoremas 3.14 y 3.17, $A \in \mathbf{V}_{\mathcal{M}}^\#$ y $A \in \mathbf{V}_{\mathcal{M}}^c$. ■

De acuerdo con este resultado, los tres sistemas desarrollados por Gupta & Belnap preservan la lógica clásica, i.e. están cerrados bajo la noción de consecuencia lógica, como se quería originalmente. En lo que resta del capítulo haremos uso de este resultado sin mencionarlo explícitamente, ya que su uso será más que frecuente.

²³Originalmente esta oración fue construida por Vann McGee (1985) en el marco de la obtención de un resultado para teorías de la verdad, del cual el teorema 3.27 es un caso particular.

TEOREMA 3.19 (Esquema T de Tarski I) Si A es una oración de L^+ y t es un término del lenguaje que denota el código de A en $\mathcal{M}$:

- $A \in \mathbf{V}_{\mathcal{M}}^*$ si y sólo si $Tr(t) \in \mathbf{V}_{\mathcal{M}}^*$;
- $A \in \mathbf{V}_{\mathcal{M}}^\#$ si y sólo si $Tr(t) \in \mathbf{V}_{\mathcal{M}}^\#$;
- $A \in \mathbf{V}_{\mathcal{M}}^c$ si y sólo si $Tr(t) \in \mathbf{V}_{\mathcal{M}}^c$.

Prueba: Si $A \in \mathbf{V}_{\mathcal{M}}^*$, por la definición 3.9, A es establemente verdadera en toda secuencia de revisión. Sea $[\mathcal{M}, S]$ una secuencia de revisión cualquiera. Por la definición 3.8, existe un α tal que, para todo $\beta \geq \alpha$, $(\mathcal{M}, S_\beta) \models A$. Por la definición 3.5, existe un α tal que, para todo $\beta \geq \alpha$, $(\mathcal{M}, S_{\beta+1}) \models Tr('A')$. Por tanto, el código de A pertenece a S_γ para todo $\gamma \geq \beta + 1$. Luego, si t denota el código de A , existe un α (*i.e.* $\beta + 1$) tal que para todo $\gamma \geq \alpha$, $(\mathcal{M}, S_\gamma) \models Tr(t)$. Por la definición 3.8, $Tr(t)$ es establemente verdadera en $[\mathcal{M}, S]$. Puesto que $[\mathcal{M}, S]$ era una secuencia arbitraria, por la definición 3.9 sabemos que $Tr(t) \in \mathbf{V}_{\mathcal{M}}^*$. La prueba de la conversas y de los casos para las otras teorías son semejantes y quedarán a cargo del lector en los ejercicios para esta subsección. ■

TEOREMA 3.20 (Esquema T de Tarski II) Si A es una oración de L^+ y t es un término del lenguaje que denota el código de A en $\mathcal{M}$:

- $\neg A \in \mathbf{V}_{\mathcal{M}}^*$ si y sólo si $\neg Tr(t) \in \mathbf{V}_{\mathcal{M}}^*$;
- $\neg A \in \mathbf{V}_{\mathcal{M}}^\#$ si y sólo si $\neg Tr(t) \in \mathbf{V}_{\mathcal{M}}^\#$;
- $\neg A \in \mathbf{V}_{\mathcal{M}}^c$ si y sólo si $\neg Tr(t) \in \mathbf{V}_{\mathcal{M}}^c$.

Prueba: Si $\neg A \in \mathbf{V}_{\mathcal{M}}^\#$, por la definición 3.13, $\neg A$ es cercanamente establemente verdadera en toda secuencia de revisión. Sea $[\mathcal{M}, S]$ una secuencia de revisión cualquiera. Por la definición 3.12, existe allí un α tal que, para todo $\beta \geq \alpha$, existe un número natural m tal que, para todo $n \geq m$, $(\mathcal{M}, S_{\beta+n}) \models \neg A$, esto es, $(\mathcal{M}, S_{\beta+n}) \not\models A$. Por la definición 3.5, para todo $n \geq m$, $(\mathcal{M}, S_{\beta+n+1}) \not\models Tr('A')$ y, por tanto, $(\mathcal{M}, S_{\beta+n+1}) \not\models Tr(t)$ si $(\mathcal{M}, S_{\beta+1}) \models t = 'A'$. Con lo cual, en $[\mathcal{M}, S]$ existe un α tal que, para todo $\beta \geq \alpha$, existe un número natural m' (*i.e.* $m + 1$) tal que, para todo $n \geq m'$, $(\mathcal{M}, S_{\beta+n}) \models \neg Tr(t)$. Por la definición 3.12, $\neg Tr(t)$ es cercanamente establemente verdadera y, por la definición 3.13, $\neg Tr(t) \in \mathbf{V}_{\mathcal{M}}^\#$. La prueba de la conversas y de los casos para las otras teorías son semejantes y quedarán a cargo del lector en los ejercicios para esta subsección. ■

Un caso particular de estos dos últimos resultados es aquél en el cual t es el numeral del código de A . De acuerdo con el teorema 3.19, A es una oración categóricamente verdadera en cada sistema siempre y cuando $Tr('A')$ también lo sea. Y, de acuerdo con el teorema 3.20, A es una oración categóricamente falsa en cada sistema siempre y cuando $Tr('A')$ también lo sea. En consecuencia, si A es una oración categórica, en

cada una de las tres teorías tenemos que $A \equiv Tr('A')$ es categóricamente verdadera. La teoría revisionista de la verdad mantiene la validez del Esquema T de Tarski en su versión extensional para todas las oraciones no patológicas.

TEOREMA 3.21 (Consistencia) $\mathbf{V}_{\mathcal{M}}^*$, $\mathbf{V}_{\mathcal{M}}^\#$ y $\mathbf{V}_{\mathcal{M}}^c$ son conjuntos consistentes.

Prueba: Supongamos que $\mathbf{V}_{\mathcal{M}}^\#$ es inconsistente para algún modelo base $\mathcal{M}$. Luego, existe una fórmula A tal que $A \in \mathbf{V}_{\mathcal{M}}^\#$ y $\neg A \in \mathbf{V}_{\mathcal{M}}^\#$. Por la definición 3.13, A y $\neg A$ son cercanamente establemente verdaderas en toda secuencia basada en $\mathcal{M}$. Sea $[\mathcal{M}, S]$ una secuencia arbitraria tal. Por la definición 3.12, existe un α tal que para todo $\beta \geq \alpha$ existe un natural m tal que, para todo $n \geq m$, $(\mathcal{M}, S_{\beta+n}) \models A$, y existe un α' tal que para todo $\beta' \geq \alpha'$ existe un natural m' tal que, para todo $n' \geq m'$, $(\mathcal{M}, S_{\beta'+n'}) \models \neg A$. Sea $\max(x, y)$ el mayor de los ordinales x e y . Luego, para todo $\beta \geq \max(\alpha, \alpha')$ existe un natural $\max(m, m')$ tal que, para todo $n \geq \max(m, m')$, $(\mathcal{M}, S_{\beta+n}) \models A \wedge \neg A$, lo cual es un absurdo. Por el teorema 3.14, $\mathbf{V}_{\mathcal{M}}^*$ también es consistente. La prueba para $\mathbf{V}_{\mathcal{M}}^c$ es análoga a la de $\mathbf{V}_{\mathcal{M}}^\#$ y quedará a cargo del lector en los ejercicios para esta subsección. ■

Ejercicios para la subsección 2.5

1. Pruebe que:

- (a) $A \in \mathbf{V}_{\mathcal{M}}^*$ sólo si $Tr(t) \in \mathbf{V}_{\mathcal{M}}^*$;
- (b) $A \in \mathbf{V}_{\mathcal{M}}^\#$ si y sólo si $Tr(t) \in \mathbf{V}_{\mathcal{M}}^\#$;
- (c) $A \in \mathbf{V}_{\mathcal{M}}^c$ si y sólo si $Tr(t) \in \mathbf{V}_{\mathcal{M}}^c$.

2. Pruebe que:

- (a) $\neg A \in \mathbf{V}_{\mathcal{M}}^*$ si y sólo si $\neg Tr(t) \in \mathbf{V}_{\mathcal{M}}^*$;
- (b) $\neg A \in \mathbf{V}_{\mathcal{M}}^\#$ sólo si $\neg Tr(t) \in \mathbf{V}_{\mathcal{M}}^\#$;
- (c) $\neg A \in \mathbf{V}_{\mathcal{M}}^c$ si y sólo si $\neg Tr(t) \in \mathbf{V}_{\mathcal{M}}^c$.

3. Pruebe que $\mathbf{V}_{\mathcal{M}}^c$ es un conjunto consistente.

2.6 Verdad en la aritmética

En lo que resta de esta sección trabajaremos exclusivamente con $\mathcal{N}$ como modelo base para el proceso de revisión de hipótesis. Buscamos entonces el conjunto de oraciones correctamente asertables de L^+ bajo la interpretación pretendida de los símbolos no lógicos y en la cual Tr se comporte como un predicado veritativo para L^+ , *i.e.* respete la definición dada en (2.5). Esto nos permitirá comparar las bondades de la teoría revisionista de la verdad con las de otras teorías semánticas como la teoría de punto fijo

de Kripke, la teoría paracompleta de Field y la teoría de la verdad transparente de Beall, usualmente desarrolladas sobre $\mathcal{N}$. Todas ellas serán discutidas en este libro.

En subsección 2.2 hemos indicado que L contiene símbolos de función para expresar ciertas funciones recursivas que nos serán útiles:

- $\dot{\neg}x$ expresa la función monádica que aplicada al código x de una fórmula A da el código de $\neg A$.
- $x \dot{\wedge} y$ y $x \dot{\vee} y$ expresan funciones diádicas análogas para la conjunción y la disyunción, respectivamente.
- $\dot{\forall}vx$ y $\dot{\exists}vx$ expresan las funciones diádicas que asignan al código x de una fórmula A y el código v de una variable de individuo z el código de $\forall zA$ y $\exists zA$, respectivamente.
- $x(y/v)$ expresa la función triádica que asigna a los códigos x de una fórmula A , y de un término t y v de una variable z el código de la fórmula que resulta de sustituir en A todas las apariciones libres de z por t .
- $\dot{x}$ expresa la función monádica que asigna a cada número x el código de su numeral.

Puesto que el conjunto de códigos de oraciones de L^+ es recursivo, existe una fórmula $Sent_{L^+}(x)$ que lo expresa. Del mismo modo, $At(x)$ expresa el conjunto de códigos de oraciones atómicas de L , $Ver(x)$ el de códigos de oraciones atómicas verdaderas de ese lenguaje, $Fal(x)$ de las falsas y $var(x)$ expresa el conjunto de códigos de variables de individuo de L^+ .

DEFINICIÓN 3.22 (Principios semánticos) Sea A una oración de L^+ :

(Tr-step) $Tr^n('A') \equiv Tr^m('A')$, para todo m y todo $n \in \mathbb{N}$

(Tr¬) $\forall x(Sent_{L^+}(x) \supset (Tr(\dot{\neg}x) \equiv \neg Tr(x)))$

(Tr∧) $\forall x\forall y((Sent_{L^+}(x) \wedge Sent_{L^+}(y)) \supset (Tr(x \dot{\wedge} y) \equiv (Tr(x) \wedge Tr(y))))$

(Tr∨) $\forall x\forall y((Sent_{L^+}(x) \wedge Sent_{L^+}(y)) \supset (Tr(x \dot{\vee} y) \equiv (Tr(x) \vee Tr(y))))$

(Tr∀) $\forall x\forall v((var(v) \wedge Sent_{L^+}(x(\dot{0}/v))) \supset (Tr(\dot{\forall}vx) \equiv \forall yTr(x(\dot{y}/v))))$

(Tr∃) $\forall x\forall v((var(v) \wedge Sent_{L^+}(x(\dot{0}/v))) \supset (Tr(\dot{\exists}vx) \equiv \exists yTr(x(\dot{y}/v))))$

(LEM) $\forall x(Sent_{L^+}(x) \supset (Tr(x) \vee Tr(\dot{\neg}x)))$

(LNC) $\forall x(Sent_{L^+}(x) \supset \neg(Tr(x) \wedge Tr(\dot{\neg}x)))$

(*Tr*-step) es un esquema. En el caso en que $m = 1$ y $n = 0$ obtenemos el esquema T de Tarski en su versión extensional para cada oración A de L^+ . Los principios (*Tr*¬), (*Tr*∧) y (*Tr*∨) expresan la conmutatividad del predicado veritativo con las conectivas lógicas correspondientes y contribuyen a expresar el carácter composicional de la verdad. Los principios (*Tr*∀) y (*Tr*∃) culminan la tarea, expresando que una oración universalmente cuantificada es verdadera siempre y cuando todas sus instancias numéricas lo sean, y que una oración existencialmente cuantificada es verdadera siempre y cuando alguna de sus instancias numéricas lo sea. Eso es correctamente afirmable gracias al hecho de que todos los elementos de $|\mathcal{N}|$ tienen un numeral en L^+ .²⁴ Los principios (LEM)²⁵ o ley de tercero excluido y (LNC)²⁶ o ley de no contradicción expresan que *Tr* es un predicado veritativo para un lenguaje clásico, lo cual es deseable pues L^+ lo es.

TEOREMA 3.23 (*Tr*-step) Para todo m y todo $n \in \mathbb{N}$, si $m \neq n$, existe una oración D de L^+ que genera una instancia inválida de (*Tr*-step) en $\mathbf{T}_{\mathcal{N}}^*$, $\mathbf{T}_{\mathcal{N}}^\#$ y $\mathbf{T}_{\mathcal{N}}^c$.

Prueba: Supongamos sin pérdida de generalidad que $n > m$. Por el lema de diagonalización, utilizando la función f del ejemplo 3.2 podemos formular una oración D equivalente a otra que dice que el resultado de iterar $n - (m + 1)$ veces el predicado veritativo a D no es verdadera para todo $S \subseteq |\mathcal{N}|$:

$$(\mathcal{N}, S) \models D \equiv \neg Tr^{n-(m+1)}('D')$$
(2.11)

Luego, en toda secuencia de revisión $[\mathcal{N}, S]$, para todo ordinal α , $(\mathcal{N}, S_\alpha) \models D$ si y sólo si $(\mathcal{N}, S_\alpha) \models \neg Tr^{n-m}('D')$. Por la definición 3.5, $(\mathcal{N}, S_{\alpha+m}) \models Tr^m('D')$ si y sólo si $(\mathcal{N}, S_{\alpha+m}) \models \neg Tr^n('D')$, con lo cual $(\mathcal{N}, S_{\alpha+m}) \not\models Tr^m('D') \equiv Tr^n('D')$. Y por las definiciones 3.8 y 3.12, $Tr^n('D') \equiv Tr^m('D')$ no es establemente ni cercanamente establemente verdadera en ninguna secuencia de revisión. Por las definiciones 3.9, 3.13 y 3.16, $Tr^n('D') \equiv Tr^m('D')$ no es válida en $\mathbf{T}_{\mathcal{N}}^*$, $\mathbf{T}_{\mathcal{N}}^\#$ ni en $\mathbf{T}_{\mathcal{N}}^c$. ■

A la luz de los teoremas 3.19, 3.20 y 3.23, concluimos que $Tr^n('A')$ y $Tr^m('A')$ son intercambiables en contextos categóricos, esto es, en casos en los cuales A es una oración categóricamente verdadera o categóricamente falsa. El Esquema T de Tarski en su versión extensional puede fallar en instancias patológicas.

TEOREMA 3.24 (Principios semánticos I)

1. (*Tr*¬), (*Tr*∧) y (*Tr*∀) pertenecen a $\mathbf{V}_{\mathcal{N}}^\#$.
2. (*Tr*¬) y (*Tr*∧) pertenecen a $\mathbf{V}_{\mathcal{N}}^c$, pero no (*Tr*∀).

²⁴Si el modelo base contara con una cantidad no enumerable de objetos en su dominio, dado que L^+ es un lenguaje contable, (*Tr*∀) y (*Tr*∃) resultarían falsos. Para expresar el carácter composicional de la verdad precisaríamos recurrir en ese caso a un predicado de satisfacción. Para una explicación detallada del problema, véase Volker Halbach (2011), p. 192.

²⁵Sus siglas en inglés están por '*Law of Excluded Middle*'.

²⁶Sus siglas en inglés están por '*Law of Non-Contradiction*'.

3. $(Tr\neg)$, $(Tr\wedge)$ y $(Tr\forall)$ no pertenecen a $\mathbf{V}_{\mathcal{N}}^*$.

Prueba: Sea $[\mathcal{N}, S]$ una secuencia de revisión cualquiera:

1. Sea A una oración cualquiera de L^+ . Para todo β , $(\mathcal{N}, S_\beta) \models A$ si y sólo si $(\mathcal{N}, S_\beta) \not\models \neg A$. Por la definición 3.5, para todo β , $(\mathcal{N}, S_{\beta+1}) \models Tr('A')$ si y sólo si $(\mathcal{N}, S_{\beta+1}) \models \neg Tr(' \neg A')$. Luego, existe un α (i.e. 0) tal que para todo $\beta \geq \alpha$ existe un natural m (i.e. 1) tal que, para todo $n \geq m$, $(\mathcal{N}, S_{\beta+n}) \models Tr(' \neg A') \equiv \neg Tr('A')$. Como A era una oración arbitraria, por las definiciones 3.12 y 3.22, $(Tr\neg)$ es cercanamente establemente verdadera y, por la definición 3.13, pertenece a $\mathbf{V}_{\mathcal{N}}^\#$.

Dejaremos a cargo del lector en los ejercicios para esta sección la prueba de que $(Tr\wedge) \in \mathbf{V}_{\mathcal{N}}^\#$.

Sea A una fórmula arbitraria de L^+ con a lo sumo z como su única variable libre. Para todo β , $(\mathcal{N}, S_\beta) \models \forall z A$ si y sólo si $(\mathcal{N}, S_\beta) \models A(\bar{n})$ para todo $n \in \mathbb{N}$. Por la definición 3.5, para todo β , $(\mathcal{N}, S_{\beta+1}) \models Tr(' \forall z A')$ si y sólo si $(\mathcal{N}, S_{\beta+1}) \models Tr('A(\bar{n})')$ para todo $n \in \mathbb{N}$ o, lo que es lo mismo, $(\mathcal{N}, S_{\beta+1}) \models \forall y Tr('A'(\dot{y}/'z'))$. Luego, existe un α (i.e. 0) tal que para todo $\beta \geq \alpha$ existe un natural m (i.e. 1) tal que, para todo $n \geq m$, $(\mathcal{N}, S_{\beta+n}) \models Tr(' \forall z A') \equiv \forall y Tr('A'(\dot{y}/'z'))$. Como A y z eran arbitrarias, por las definiciones 3.12 y 3.22, $(Tr\forall)$ es cercanamente establemente verdadera y, por la definición 3.13, pertenece a $\mathbf{V}_{\mathcal{N}}^\#$.

2. Sean A y B oraciones cualesquiera de L^+ . Si $[\mathcal{N}, S]$ es una secuencia máximamente consistente, por la definición 3.15, para todo β , $(\mathcal{N}, S_\beta) \models Tr('A')$ si y sólo si $(\mathcal{N}, S_\beta) \models \neg Tr(' \neg A')$ y $(\mathcal{N}, S_\beta) \models Tr('A \wedge B')$ si y sólo si $(\mathcal{N}, S_\beta) \models Tr('A') \wedge Tr('B')$. Como A y B eran arbitrarias, por las definiciones 3.8 y 3.22, $(Tr\neg)$ y $(Tr\wedge)$ son establemente verdaderas y, por la definición 3.16, pertenecen a $\mathbf{V}_{\mathcal{N}}^c$.

Como hemos mencionado, el lema de diagonalización nos permite formular la oración E de McGee mediante la equivalencia obtenida en (2.10) en todo $S \subseteq |\mathcal{N}|$. Supongamos que $(Tr\forall) \in \mathbf{V}_{\mathcal{N}}^c$. Luego, la instancia de $(Tr\forall)$ generada por $Tr(\dot{f}('E', x))$ también pertenece a $\mathbf{V}_{\mathcal{N}}^c$. Por las definiciones 3.8 y 3.16, en cada $[\mathcal{N}, S]$ máximamente consistente existe un α tal que para todo $\beta \geq \alpha$:

$$(\mathcal{N}, S_\beta) \models \forall x Tr('Tr(\dot{f}('E', x))') \equiv Tr(' \forall x Tr(\dot{f}('E', x))') \quad (*)$$

$(\mathcal{N}, S_\alpha) \models E$ o $(\mathcal{N}, S_\alpha) \not\models E$. En el primer caso, por la definición 3.5, para cada $n \in \mathbb{N}$, $Tr(\dot{f}('E', \bar{n})) \in S_{\alpha+m}$ para todo natural $m > n$. En el segundo caso, por (2.10) y la definición 3.5, $(\mathcal{N}, S_{\alpha+1}) \models \neg Tr('E')$, con lo cual $(\mathcal{N}, S_{\alpha+1}) \models \neg \forall x Tr(\dot{f}('E', x))$ y, por (2.10), $(\mathcal{N}, S_{\alpha+1}) \models E$. Por la definición 3.5, nuevamente tenemos que, para cada $n \in \mathbb{N}$, $Tr(\dot{f}('E', \bar{n})) \in S_{\alpha+m}$ para todo natural $m > n + 1$. En cualquier caso, por la definición 3.5:

$$(\mathcal{N}, S_{\alpha+\omega}) \models \forall x Tr(\dot{f}('E', x)) \quad (**)$$

Y, por tanto, $(\mathcal{N}, S_{\alpha+\omega}) \models \forall x Tr(\dot{f}('E', x+1))$ o, lo que es lo mismo:

$$(\mathcal{N}, S_{\alpha+\omega}) \models \forall x Tr(Tr(\dot{f}('E', x))) \quad (***)$$

De (*) y (***), $(\mathcal{N}, S_{\alpha+\omega}) \models Tr(\forall x Tr(\dot{f}('E', x)))$, con lo cual, $\forall x Tr(\dot{f}('E', x)) \in S_{\alpha+\omega}$. Por la definición 3.15, $S_{\alpha+\omega}$ es máximamente consistente y, por (2.10), $E \notin S_{\alpha+\omega}$. Pero al mismo tiempo, por (**), tenemos que $(\mathcal{N}, S_{\alpha+\omega}) \models Tr(\dot{f}('E', 0))$, esto es, $E \in S_{\alpha+\omega}$, lo cual es un absurdo.

3. Dejaremos la prueba de que $(Tr\forall) \notin \mathbf{V}_{\mathcal{N}}^*$ a cargo del lector en los ejercicios para esta subsección.

Puesto que M en (2.6) es inestable a lo largo de toda secuencia de revisión basada en $\mathcal{N}$, por la definición 3.5 existe una secuencia $[\mathcal{N}, S, \mathfrak{B}]$ tal que $\{M, \neg M\} \subseteq \mathfrak{B}(\lambda)$ para todo ordinal límite λ , esto es, $(\mathcal{N}, S_\lambda) \models Tr('M') \wedge Tr('¬M')$ o, lo que es lo mismo, $(\mathcal{N}, S_\lambda) \models \neg Tr('M') \equiv Tr('¬M')$ para todo ordinal límite λ . Por lo tanto, no existe un ordinal α tal que para todo $\beta \geq \alpha$ $(\mathcal{N}, S_\beta) \models \neg Tr('M') \equiv Tr('¬M')$ y, por las definiciones 3.8 y 3.22, $(Tr¬)$ no es establemente verdadera en $[\mathcal{N}, S, \mathfrak{B}]$. Por la definición 3.9, $(Tr¬) \notin \mathbf{V}_{\mathcal{N}}^*$. La prueba para $(Tr\wedge)$ es similar y quedará en manos del lector en los ejercicios al finalizar esta subsección. ■

COROLARIO 3.25 (Principios semánticos II)

1. $(Tr\vee)$, $(Tr\exists)$, (LEM) y (LNC) pertenecen a $\mathbf{V}_{\mathcal{N}}^\#$.
2. $(Tr\vee)$, (LEM) y (LNC) pertenecen a $\mathbf{V}_{\mathcal{N}}^c$, pero no $(Tr\exists)$.
3. $(Tr\vee)$, $(Tr\exists)$, (LEM) y (LNC) no pertenecen a $\mathbf{V}_{\mathcal{N}}^*$.

Prueba: Las pruebas son inmediatas en virtud de la interdefinibilidad de los operadores lógicos. Quedarán a cargo del lector. ■

El teorema 3.24 y el corolario 3.25 muestran que $\mathbf{T}_{\mathcal{N}}^\#$ valida todos los principios composicionales para Tr , esto es, $(Tr¬)$, $(Tr\wedge)$, $(Tr\vee)$, $(Tr\forall)$ y $(Tr\exists)$ y, *a fortiori*, los principios semánticos de Tercero Excluido y No Contradicción. $\mathbf{T}_{\mathcal{N}}^c$ sólo valida los tres primeros y los dos últimos y $\mathbf{T}_{\mathcal{N}}^*$ no valida ninguno. Puesto que los principios en cuestión son *prima facie* deseables e intuitivos, $\mathbf{T}_{\mathcal{N}}^\#$ pareciera ubicarse por encima de sus rivales en cuanto teoría semántica de la verdad para la aritmética. Sin embargo, un teorema de McGee (1985) prueba que cualquier sistema que valide todos los principios en cuestión y contenga $Tr('A')$ siempre que contenga A es ω -inconsistente. $\mathbf{T}_{\mathcal{N}}^\#$ verifica todas las hipótesis de este teorema de McGee.

DEFINICIÓN 3.26 (ω -inconsistencia) Un conjunto Γ de fórmulas de L^+ cerrado bajo consecuencia lógica es ω -inconsistente si y sólo si existe una fórmula A de L^+ con una única variable libre x tal que $A[\bar{n}/x] \in \Gamma$ para todo $n \in \mathbb{N}$ y, a la vez, $\neg\forall x A \in \Gamma$.

TEOREMA 3.27 (ω -inconsistencia) $\mathbf{V}_{\mathcal{N}}^{\#}$ es ω -inconsistente.

Prueba: Sea E como en (2.10). Puesto que $E \equiv \neg \forall x Tr(\dot{f}('E', x)) \in \mathbf{V}_{\mathcal{N}}^{\#}$, $\neg E \supset Tr('E') \in \mathbf{V}_{\mathcal{N}}^{\#}$ (*). Por (2.10) y el punto 3 del teorema 3.24, $Tr('E') \supset \neg \forall x Tr(\dot{f}('E', x+1)) \in \mathbf{V}_{\mathcal{N}}^{\#}$ y, por ende, $Tr('E') \supset \neg \forall x Tr(\dot{f}('E', x)) \in \mathbf{V}_{\mathcal{N}}^{\#}$, con lo cual $Tr('E') \supset E \in \mathbf{V}_{\mathcal{N}}^{\#}$ (**). De (*) y (**), $\neg E \supset E \in \mathbf{V}_{\mathcal{N}}^{\#}$ y, por tanto, $E \in \mathbf{V}_{\mathcal{N}}^{\#}$. Por (2.10) nuevamente, $\neg \forall x Tr(\dot{f}('E', x)) \in \mathbf{V}_{\mathcal{N}}^{\#}$, pero al mismo tiempo, $Tr(\dot{f}('E', \bar{n})) \in \mathbf{V}_{\mathcal{N}}^{\#}$ para cada $n \in \mathbb{N}$, por el teorema 3.19. Finalmente, por la definición 3.26, $\mathbf{V}_{\mathcal{N}}^{\#}$ es ω -inconsistente. ■

Teniendo en mente que el objetivo de esta sección es dar con el conjunto de oraciones correctamente asertables de L^+ bajo la interpretación estándar de los símbolos no lógicos de L , y en la cual Tr se comporte como un predicado de verdad para todo L^+ , el teorema 3.27 indica que $\mathbf{V}_{\mathcal{N}}^{\#}$ no es el conjunto que buscamos. Pues bajo la interpretación estándar del lenguaje de la aritmética de primer orden, los únicos objetos que pertenecen al dominio son los números naturales estándar, tal como los conocemos. Sin embargo, cada una de las instancias de $Tr(\dot{f}('E', \bar{n}))$ dadas por $n \in \mathbb{N}$ pertenecen a $\mathbf{V}_{\mathcal{N}}^{\#}$ y, al mismo tiempo, también pertenece a este conjunto $\exists x \neg Tr(\dot{f}('E', x))$. Luego, de acuerdo con $\mathbf{T}_{\mathcal{N}}^{\#}$, existe al menos un número natural no estándar y, por lo tanto, no es un conjunto de oraciones correctamente asertables de L^+ bajo la interpretación pretendida.²⁷

En consecuencia, si bien $\mathbf{T}_{\mathcal{N}}^{\#}$ prueba ser una mejor teoría semántica por permitir una buena cantidad de principios intuitivos y deseables para Tr que sus rivales no validan, paga el precio de no cumplir nuestro objetivo principal. Gupta & Belnap (1993) no se pronuncian por ninguna de la tres teorías que presentan y desarrollan.

Ejercicios para la subsección 2.6

1. Pruebe que $(Tr \wedge) \in \mathbf{V}_{\mathcal{N}}^{\#}$.
2. Pruebe que $(Tr \forall) \notin \mathbf{V}_{\mathcal{N}}^*$.
3. Pruebe que $(Tr \wedge) \notin \mathbf{V}_{\mathcal{N}}^*$.
4. Pruebe que:
 - (a) $(Tr \vee)$, $(Tr \exists)$, (LEM) y (LNC) pertenecen a $\mathbf{V}_{\mathcal{N}}^{\#}$.
 - (b) $(Tr \vee)$, (LEM) y (LNC) pertenecen a $\mathbf{V}_{\mathcal{N}}^c$, pero no $(Tr \exists)$.
 - (c) $(Tr \vee)$, $(Tr \exists)$, (LEM) y (LNC) no pertenecen a $\mathbf{V}_{\mathcal{N}}^*$.

²⁷Para una discusión acerca de este punto véase Halbach y Leon Horsten (2005), Hannes Leitgeb (2007) y Barrio (2010).

3 Algunos problemas y soluciones alternativas

3.1 Políticas de bootstrapping

Como hemos mencionado ya, las primeras versiones de la teoría revisionista de la verdad fueron dadas independientemente por Herzberger (1982a), (1982b) y Gupta (1982). Estos primeros intentos no fueron exactamente idénticos entre sí. Parte de la diferencia entre ambos yace en la elección de diferentes políticas de bootstrapping, esto es, del modo en que las oraciones inestables son clasificadas dentro o fuera de la extensión del predicado veritativo en los estados límites de una secuencia de revisión.²⁸

En cada secuencia de revisión hay dos tipos de estados: los sucesores y los no sucesores. Para los primeros el modo de proceder es claro y razonable: aplicamos el esquema T de Tarski. Para los estados no sucesores, también hay reglas razonables a seguir, que nos indican qué hacer con las oraciones que se han estabilizado dentro o fuera de la extensión del predicado de verdad. Sin embargo, estas reglas no son suficientes, pues nada dicen sobre las oraciones inestables. Deben ser por tanto suplementadas con una política de bootstrapping. Si bien puede parecer que su elección es completamente arbitraria y poco importante, veremos en esta subsección que diferentes políticas de bootstrapping darán lugar a diferentes clasificaciones de las oraciones de L^+ en las siete categorías semánticas que hemos presentado en la definición 3.10.

3.1.1 El sistema de Herzberger

Herzberger ha elegido quizás la política más simple, que notaremos $\mathfrak{B}_H$.

DEFINICIÓN 3.28 ($\mathfrak{B}_H$) Sea $[\mathcal{M}, S, \mathfrak{B}_H]$ una secuencia de revisión. $\mathfrak{B}_H(\lambda) = \emptyset$ para todo ordinal límite λ .

De acuerdo con Herzberger, las oraciones inestables previamente a cada ordinal límite λ deberán quedar fuera de la extensión del predicado veritativo en la etapa λ del proceso de revisión. Reemplazando $\mathfrak{B}$ en la definición 3.5 por $\mathfrak{B}_H$, obtenemos un conjunto más pequeño de secuencias de revisión con las cuales clasificar las oraciones de L^+ dentro de las siete categorías semánticas. Las llamaremos ‘secuencias de revisión de Herzberger’. En cada estado límite λ de cada secuencia $[\mathcal{M}, S]$, $S_\lambda = S_\lambda^\top$.

Como veremos en los ejemplos subsiguientes, las secuencias de Herzberger constituyen un conjunto demasiado pequeño, permitiendo clasificar oraciones de L^+ de forma intuitivamente incorrecta. El primero tendrá que ver con la interacción de distintas oraciones del mentiroso lógicamente independientes entre sí.

EJEMPLO 3.29 (Bicondicionales de mentirosos)²⁹ Sea M como en (2.6). Puesto que

²⁸Véase la definición 3.5.

²⁹El primero en notar este error de clasificación como un producto de la elección de diferentes políticas de bootstrapping, tanto en el sistema de Herzberger como en el de Gupta, fue Belnap (1982).

la clase de puntos fijos en $\mathcal{N}$ de $\neg Tr$ no es recursiva, existe de una nueva oración M' del mentiroso independiente de (o, lo que es lo mismo, no equivalente a) M tal que para todo $S \subseteq |\mathcal{N}|$:

$$(\mathcal{N}, S) \models M' \equiv \neg Tr('M') \quad (3.1)$$

Al igual que M , M' es inestable en toda secuencia de revisión. Luego, lo es previamente a ω , el primer ordinal límite. Por lo tanto, ni M ni M' caerán dentro de la extensión de Tr en la etapa ω de una secuencia de revisión de Herzberger. A partir de allí, ambas oscilarán juntas, ya sea dentro o fuera de la extensión de este predicado, en todas las etapas de la secuencia. Luego, la siguiente oración es establemente verdadera y, por ende, categóricamente verdadera de acuerdo con el sistema de Herzberger:

$$M \equiv M' \quad (3.2)$$

Aquello constituye un error, en tanto ambas oraciones son independientes.

La segunda clase de errores cometidos por el sistema de Herzberger fue estudiada principalmente por Aladdin Yaqūb (1993) y concierne a aquellos conjuntos de oraciones infundadas que intuitivamente poseen una única asignación aceptable de valores de verdad clásicos. Bautizada posteriormente por André Chapuis (1996) como ‘Gupta puzzles’, Gupta & Belnap reconocen ellos mismos primero la existencia de esta clase de problemas dentro de su teoría.³⁰

EJEMPLO 3.30 (Gupta puzzles I) Con la ayuda del lema de diagonalización, podemos probar la existencia de dos oraciones J_a y J_b tales que, para todo $S \subseteq |\mathcal{N}|$:³¹

$$(\mathcal{N}, S) \models J_a \equiv (\neg Tr('J_a') \vee \neg Tr('J_b')) \quad (3.3)$$

$$(\mathcal{N}, S) \models J_b \equiv (\neg Tr('J_a') \wedge \neg Tr('J_b')) \quad (3.4)$$

J_a afirma que ella o J_b es falsa, mientras que J_b afirma que ambas lo son. Supongamos que J_b es verdadera. Luego, por una aplicación de la versión material y extensional del esquema T de Tarski, tenemos que $\neg Tr('J_a') \wedge \neg Tr('J_b')$ y, con ello, que $\neg Tr('J_b')$, contradiciendo nuestra hipótesis inicial. Diremos entonces que J_b es falsa, esto es, que $\neg Tr('J_b')$. Aplicando la regla de introducción de la disyunción, tenemos que $\neg Tr('J_a') \vee \neg Tr('J_b')$ y, mediante otra aplicación del esquema T extensional podemos concluir que J_a es verdadera. En consecuencia, la única asignación consistente de valores de verdad

³⁰Véase Gupta & Belnap (1993), cap. 6, sec. C.

³¹Una versión más sofisticada del ejemplo corresponde a Roy Cook (2002). Este artículo se enmarca en una discusión sobre la adecuación material de la teoría de la revisión—precisamente el punto que nos concierne aquí—con Michael Kremer (2002), quien ofreció la versión simplificada del ejemplo original, que presentamos aquí.

clásicos a estas oraciones es la que da verdad a J_a y falsedad a J_b . Sin embargo, la siguiente tabla muestra que existe al menos una secuencia de Herzberger en la cual las oraciones en cuestión no se estabilizan en estos valores:

	0	1	2	3	4	5	...	ω	$\omega + 1$	...
J_a	t	f	t	f	t	f		t	f	t
J_b	t	f	t	f	t	f		t	f	t

Por ende, el sistema no representa fielmente nuestro razonamiento alrededor de la verdad, pues no nos permite afirmar que J_a es categóricamente verdadera y que J_b es categóricamente falsa, como es aparentemente correcto.

Podría argumentarse, como ha sugerido Chapuis (1996)³² y desarrollado Kremer (2002), que los pasos que llevan a concluir que J_a ha de ser verdadera mientras que J_b falsa hacen un uso ilícito de la versión extensional del esquema T que la teoría misma señala como incorrecta,³³ a la vez que descartan sin aparente razón aquellas asignaciones inestables de valores de verdad a estas oraciones que la teoría considera legítimas. En efecto, una política que busque minimizar las inestabilidades llevaría a errores, como muestra el siguiente ejemplo de Chapuis.

EJEMPLO 3.31 (Gupta puzzles II) Sea H como en (2.8). Gracias al lema de diagonalización, podemos probar la existencia de una oración I tal que, para todo $S \subseteq |\mathcal{N}|$:

$$(\mathcal{N}, S) \models I \equiv (Tr('H') \vee \neg Tr('I')) \quad (3.5)$$

I afirma que o bien H es verdadera, o bien ella misma es falsa. Mediante un razonamiento sencillo podemos ver que la única asignación estable de valores de verdad a H e I implica que H sea verdadera. Luego, si las asignaciones estables fueran preferibles a las inestables, deberíamos concluir que es preferible que la oración del honesto sea verdadera a que sea falsa, lo cual es incorrecto.³⁴

Asimismo, Kremer (2002) argumenta a favor de las inestabilidades en grupos de oraciones infundadas como son los Gupta puzzles, a pesar de que intuitivamente existe una asignación de valores de verdad estable para ellas. No obstante, Cook (2003) señala un caso en el que las oraciones involucradas son infundadas y, sin embargo, la teoría de la revisión sí da un veredicto categórico—el único posible—, el cual desarrollamos en el siguiente ejemplo.

³²Véase Chapuis (1996), pp. 407-408.

³³Aunque esto podría ser visto como un argumento circular: a favor de la teoría empleando la teoría misma.

³⁴Un modo de ver la incorrección de esta afirmación es considerar la oración I' , equivalente a $(Tr('H') \wedge \neg Tr('I'))$.

EJEMPLO 3.32 (Gupta puzzles III) Una vez más, vía el lema de diagonalización, podemos probar la existencia de dos oraciones K_a y K_b tales que, para todo $S \subseteq |\mathcal{N}|$:

$$(\mathcal{N}, S) \models K_a \equiv \neg(Tr('K_a') \equiv Tr('K_b')) \quad (3.6)$$

$$(\mathcal{N}, S) \models K_b \equiv (Tr('K_a') \equiv Tr('K_b')) \quad (3.7)$$

K_a afirma que su valor de verdad es diferente al de K_b , mientras que K_b afirma que son iguales. Una vez más, un argumento intuitivo nos lleva a concluir que K_a es verdadera y K_b falsa, como podrá comprobar el lector. Y en efecto, el sistema de Herzberger las clasifica de ese modo.

Debiendo abandonar entonces la explicación según la cual las oraciones resultan patológicas de acuerdo con el sistema de Herzberger por ser infundadas, surge una nueva explicación posible: la diferencia entre las oraciones del ejemplo 3.30 y 3.32 yace en que K_a es una tautología y K_b una contradicción, mientras que J_a y J_b no responden a ninguno de estos esquemas. Puesto que las secuencias de Herzberger constituyen un subconjunto de las secuencias de revisión, el teorema 3.18 ofrece una buena razón para diferenciar entre ambos casos. Sin embargo, existe un tercer conjunto de oraciones claramente infundadas consideradas no patológicas por el sistema de Herzberger y, no obstante, ninguna de ellas es una tautología o una contradicción, como ilustra el siguiente ejemplo.

EJEMPLO 3.33 (Gupta puzzles IV) Al igual que en los casos anteriores, el lema de diagonalización garantiza la existencia de dos oraciones P_a y P_b tales que, para todo $S \subseteq |\mathcal{N}|$:

$$(\mathcal{N}, S) \models P_a \equiv (Tr('P_a') \vee Tr('P_b')) \quad (3.8)$$

$$(\mathcal{N}, S) \models P_b \equiv (\neg Tr('P_a') \wedge \neg Tr('P_b')) \quad (3.9)$$

P_a afirma que ella o P_b ha de ser verdadera, mientras que P_b afirma que ambas son falsas. Tanto un razonamiento intuitivo como el sistema de Herzberger llevan a concluir que P_a es verdadera y P_b falsa. Su comprobación quedará a cargo el lector. Gupta & Belnap (1993) mismos presentan una versión de este ejemplo y sostienen que ninguna de las oraciones es una tautología o una contradicción.³⁵

Resta entonces explicar porqué en algunos casos los conjuntos de oraciones infundadas que dan lugar a los Gupta puzzles llegan a estabilizarse en la única asignación de valores de verdad clásicos posible mientras que en otros casos no. Como hemos visto, no es posible hallar la causa en el carácter lógico de las oraciones involucradas.

³⁵Véase Gupta & Belnap (1993), p. 151.

Existen otros Gupta puzzles que representan problemas para la teoría de Herzberger y para otras teorías de las cuales hablaremos a continuación. A la luz de ellos y del tratamiento que estas teorías ofrecen a oraciones como (3.2), éstas parecen ser materialmente inadecuadas en cuanto teorías de la verdad.

3.1.2 El sistema de Gupta

Al igual que Herzberger, Gupta (1982) adopta una política constante de bootstrapping, esto es, la misma para cada ordinal límite.

DEFINICIÓN 3.34 ($\mathfrak{B}_G$) Sea $[\mathcal{M}, S, \mathfrak{B}_G]$ una secuencia de revisión (generada por la hipótesis inicial S). $\mathfrak{B}_G(\lambda) = S$ para todo ordinal límite λ .

De acuerdo con la teoría original de Gupta, si una oración resulta inestable previamente a un estado límite λ en una secuencia $[\mathcal{M}, S]$, pertenecerá a S_λ siempre y cuando pertenezca a S , pues reemplazando $\mathfrak{B}$ en la definición 3.5 por $\mathfrak{B}_G$, $S_\lambda = (S_\lambda^\top \cup S) - S_\lambda^\perp$. Al igual que las secuencias de Herzberger, las secuencias de Gupta también constituyen un subconjunto propio del conjunto de secuencias de revisión dadas por la definición 3.5. Y como veremos a continuación, las secuencias de Gupta también tienen problemas de clasificación de las oraciones de L^+ de acuerdo con las siete categorías que han sido presentadas en la definición 3.10.

EJEMPLO 3.35 (Bicondicionales de mentirosos) Sean M, M' y (3.2) como en el ejemplo 3.29. Como podrá comprobar el lector, a diferencia del sistema de Herzberger, el sistema de Gupta no clasifica esta última oración dentro de la categoría de oraciones categóricamente verdaderas sino tf-caprichosas, esto es, la trata como una oración del honesto. Dado que M y M' son independientes, Yaqūb (1993) identifica esta clasificación con un error, mientras que Belnap (1982) la considera, cuanto menos, extraña. Este último sugiere que la oración debería resultar inestable en al menos una secuencia. El primero, en cambio, afirma que la falta de relación entre M y M' es tal que (3.2) debería ser inestable en toda secuencia de revisión y catalogada, por tanto, como paradójica.

En cuanto a los Gupta puzzles, el sistema de Gupta se comporta exactamente del mismo modo que el de Herzberger con respecto a los conjuntos de oraciones presentados en los ejemplos 3.30, 3.32 y 3.33. La verificación quedará a cargo del lector.

EJEMPLO 3.36 (Ciclos de honestos) Consideremos la oración H del honesto dada por (2.8). Al igual que la teoría de Gupta & Belnap en el ejemplo 3.7, el sistema de Gupta clasifica correctamente esta oración como tf-caprichosa. Sin embargo, con la ayuda del lema de diagonalización, podemos probar la existencia de dos oraciones H_a y H_b tales que, para todo $S \subseteq |\mathcal{N}|$:

$$(\mathcal{N}, S) \models H_a \equiv Tr('H_b') \quad (3.10)$$

$$(\mathcal{N}, S) \models H_b \equiv Tr('H_a') \quad (3.11)$$

que intuitivamente funcionan del mismo modo que H y que, no obstante, no son clasificadas adecuadamente por el sistema de Gupta como tf-caprichosas. Pues si partimos de una hipótesis a la cual pertenece H_a pero no H_b , los valores de verdad de estas oraciones serán inestables a lo largo de toda la secuencia, ya que ningún ordinal límite las juntará dentro o fuera de la extensión del predicado de verdad, sino que copiará la elección original.

3.1.3 El sistema de Belnap

Belnap (1982) considera insatisfactorios tanto el sistema de Gupta como el de Herzberger. Desde su punto de vista, habiendo reconocido que es preciso decidir arbitrariamente la suerte de oraciones inestables previamente a cada ordinal límite en las secuencias de revisión, Gupta y Herzberger borran con el codo lo escrito con la mano, esto es, introducen un grado de arbitrariedad al comienzo de las secuencias—pues permiten cualquier hipótesis inicial—pero no dejan libertad alguna en estados límites.

Belnap señala esta restricción como la causa de los errores de clasificación de bicondicionales de mentirosos como los presentados en los ejemplos 3.29 y 3.34, y de otras oraciones compuestas por oraciones del mentiroso independientes entre sí. A raíz de ello, propone lo siguiente:³⁶

[...] Pido que se considere un esquema en todo similar al de Gupta, excepto en que hay un bootstrapper definido aparte para cada ordinal no sucesor. [...] El propósito de mi sugerencia es tornar máximamente arbitrarias aquellas notas de la Regla de Revisión que requieran cualquier tipo de arbitrariedad. Sólo entonces estaremos seguros de que lo que *sobrevive* esta arbitrariedad realmente depende entera y exclusivamente del modelo con el cual comenzamos junto con la Regla de Revisión, los componentes del proceso de revisión.

Formalmente, la política de bootstrapping propuesta por Belnap (1982) es la que hemos presentado en la definición 3.5 y que ha sido aceptada también por Gupta en Gupta & Belnap (1993). Si bien este sistema clasifica oraciones como (3.2) como tfp-caprichosas, cumpliendo las expectativas de Belnap aunque no las de Yaqūb, lamentablemente adolece de los mismos problemas de clasificación en cuanto a los Gupta puzzles presentados en los ejemplos 3.30, 3.32 y 3.33 que los sistemas de Gupta y Herzberger, y

³⁶Belnap (1982), pp. 106-107. La traducción es nuestra; las mayúsculas y las itálicas corresponden al texto original.

tampoco clasifica correctamente el ciclo de honestos del ejemplo 3.35. No obstante, ‘The Revision Theory of Truth’ se ha convertido en el *locus classicus* con respecto a teorías revisionistas de la verdad.

3.1.4 El sistema de Yaqūb

Como ya hemos mencionado, Yaqūb (1993) no está satisfecho con la clasificación que los tres sistemas presentados hasta aquí hacen de oraciones como (3.2). Adicionalmente, es el primero en señalar las dificultades que tienen los sistemas presentados a la hora de clasificar honestos cíclicos y también insiste sobre los problemas de clasificación a los cuales dan lugar los Gupta puzzles.

De acuerdo con Yaqūb, estos errores de clasificación se producen debido a una arbitrariedad excesiva, tanto en la hipótesis inicial como en los bootstrappers en cada ordinal límite, que la regla de revisión tarskiana no logra corregir en estados sucesores. Pero dado que la arbitrariedad al inicio del proceso no puede advertirse sino a la luz del proceso mismo, Yaqūb concluye, debemos aceptarla. La corrección de aquellos errores deberá venir, por tanto, de la mano de una mayor restricción en la política de bootstrapping para los estados límites. La nueva política deberá lograr deshacerse de aquellas elecciones incorrectas realizadas por ciertas hipótesis iniciales y por estados límites previos que la regla de revisión tarskiana no logra corregir en pasos sucesores.

DEFINICIÓN 3.37 ($\mathfrak{B}_Y$) Sea $[\mathcal{M}, S, \mathfrak{B}_Y]$ una secuencia de revisión y sea κ el primer ordinal no enumerable. Si λ es un ordinal límite:

1. si $\omega \leq \lambda \leq \kappa$, $\mathfrak{B}(\lambda) = S$;
2. si $\lambda \geq \kappa$, $\mathfrak{B}(\lambda) = \mathfrak{B}(\kappa\pi)$, donde π es el único ordinal tal que $\kappa\pi \leq \lambda < \kappa(\pi + 1)$;
3. Para todo $\Gamma_0, \Gamma_1, \Gamma_2 \subseteq |\mathcal{M}|^{37}$ y para todo ordinal β existe un ordinal γ tal que $\kappa\gamma \geq \beta$ y, para todo $\delta \leq \omega$:

$$\mathfrak{B}(\kappa\gamma + \kappa\delta) = \begin{cases} \Gamma_0 & \text{si } \delta \text{ es par} \\ \Gamma_1 & \text{si } \delta \text{ es impar} \\ \Gamma_2 & \text{si } \delta = \omega \end{cases}$$

La condición 1 de esta definición establece que S debe ser el bootstrapper asignado a ω y a todos los ordinales límites que se encuentren entre ω y κ . Por su parte, 2 requiere que una vez que se introduce un bootstrapper en una etapa límite λ , éste sea empleado en todos los estados límites entre λ y $\lambda + \kappa$, *i.e.* durante κ pasos. De este modo, Yaqūb se asegura de que la elección original pueda tener impacto sobre la secuencia, y de que cada elección que se haga posteriormente también lo tenga. La tercera condición asegura

³⁷Esto es, para cualesquiera tres conjuntos de (códigos de) oraciones de L^+ .

que cada tripla de bootstrappers siempre vuelva a aparecer en las secuencias, empleando cada uno de los tres bootstrappers durante κ pasos y alternando estas subsecuencias de longitud $\kappa\omega + 1$ veces. Así, cada bootstrapper y cada combinación de dos o tres bootstrappers tendrá acceso irrestricto e igualitario al proceso.

En suma, las secuencias de Yaqūb están diseñadas para dar lugar y capacidad de impacto a cada uno de los subconjuntos de oraciones de L^+ . Como veremos en los ejemplos que siguen, las secuencias de Yaqūb solucionan los problemas presentados en los ejemplos 3.30, 3.32 y 3.33, que afectan los sistemas de Herzberger, Gupta y Belnap.

EJEMPLO 3.38 (Bicondicionales de mentirosos) Sean $\Gamma_0 = \emptyset$ y $\Gamma_1 = \Gamma_2 = \{M\}$. En virtud de la condición 3 de la definición 3.37, en cada secuencia, para todo ordinal β existe un ordinal γ tal que a partir del paso $\kappa\gamma \geq \beta$ en la secuencia hasta $\kappa\gamma + \kappa\omega$, los bootstrappers Γ_0 y Γ_1 se emplean en segmentos sucesivos de longitud κ . Luego, no existe una etapa en ninguna secuencia a partir de la cual $\emptyset$ y $\{M\}$ no sean utilizados ya. En cada uno de estos períodos de longitud κ , donde se emplean o bien $\emptyset$ o bien $\{M\}$ alternadamente, (3.2) resulta verdadera y falsa, respectivamente. Luego, no se estabiliza dentro ni fuera de la extensión del predicado veritativo, ni siquiera cercanamente, en secuencia alguna. Por ende, la oración es clasificada como paradójica si tomamos en cuenta exclusivamente las secuencias de Yaqūb.

EJEMPLO 3.39 (Gupta puzzles) Sean J_a y J_b como en el ejemplo 3.30. Supongamos que existe una secuencia $[\mathcal{N}, S, \mathfrak{B}_Y]$ en la cual J_a y J_b no se estabilizan dentro y fuera de la extensión del predicado veritativo, respectivamente. En virtud de la condición 3 de la definición 3.37, tomando $\Gamma_0 = \Gamma_1 = \Gamma_2 = \{J_a\}$ y $\beta = 0$, debe existir un γ tal que $\{J_a\}$ sea el bootstrapper utilizado en todos los ordinales límites que se encuentran entre $\kappa\gamma$ y $\kappa\omega$, inclusive. Dado que J_a y J_b son inestables a lo largo de toda la secuencia, la condición 3 de la definición 3.5 indica que $J_a \in S_{\kappa\gamma}$, mientras que $J_b \notin S_{\kappa\gamma}$. Pues $S_{\kappa\gamma} = (S_{\kappa\gamma}^\top \cup \mathfrak{B}_Y(\kappa\gamma)) - S_{\kappa\gamma}^\perp$ y $\mathfrak{B}(\kappa\gamma) = \{J_a\}$, pero ni $J_b \in S_{\kappa\gamma}^\top$, por ser inestable, ni $J_a \in S_{\kappa\gamma}^\perp$, por la misma razón. Luego, como podrá comprobar el lector, a partir de $\kappa\gamma + 1$, J_a permanece dentro de la extensión de Tr y J_b fuera, contradiciendo la suposición original.

A diferencia de los sistemas de Herzberger, Gupta y Belnap, el sistema de Yaqūb clasifica a J_a como categóricamente verdadera y a J_b como categóricamente falsa, esto es, les asigna los únicos valores de verdad clásicos que pueden tomar si razonamos intuitivamente acerca de estas oraciones.

EJEMPLO 3.40 (Ciclos de honestos) Sean H_a y H_b como en el ejemplo 3.36. Dejaremos a cargo del lector la prueba de que ambas oraciones son clasificadas como tñ-caprichosas por el sistema de Yaqūb, como se quería originalmente.

Tras presentar su sistema, Yaqūb (1993) afirma no haber encontrado error de clasificación alguno. Como veremos a continuación, Chapuis (1996) sí lo encontrará.

3.1.5 El sistema de Chapuis

De acuerdo con Chapuis (1996), muchos de los problemas señalados por Yaqūb en las teorías de Herzberger, Gupta y Belnap son serias dificultades que han de ser solucionadas para obtener una teoría de la verdad materialmente adecuada. Sin embargo, el modo en que Yaqūb soluciona este inconveniente resulta insatisfactorio, como muestra el siguiente ejemplo dado por el mismo Chapuis.

EJEMPLO 3.41 Consideremos la siguiente instancia del esquema T extensional dada por (3.2):

$$Tr('M \equiv M') \equiv (M \equiv M') \quad (3.12)$$

Siendo que (3.2) es inestable en toda secuencia de Yaqūb, también lo es (3.12). Consideremos ahora esta otra oración:

$$\exists x(P(x) \wedge Tr(x)) \quad (3.13)$$

donde $P(x)$ es una fórmula de L^+ con una variable libre x que en $\mathcal{N}$ expresa el conjunto que contiene a (3.12) y a todas las iteraciones de Tr sobre esta oración, esto es:

$$P^{\mathcal{N}} = \{Tr^n('Tr('M \equiv M') \equiv (M \equiv M')) : n \in \mathbb{N}\}$$

Luego, (3.13) afirma que hay al menos un número de iteraciones de Tr sobre (3.12) que da lugar a una fórmula verdadera. Puesto que (3.12) es inestable, (3.13) también deberá serlo y, en efecto, es declarada paradójica en el sistema de Yaqūb. Finalmente, consideremos la siguiente y última oración de este ejemplo:

$$\exists x(Q(x) \wedge Tr(x)) \quad (3.14)$$

donde $Q(x)$ es una fórmula de L^+ con una variable libre x que en $\mathcal{N}$ expresa el conjunto que contiene a (3.13) y a todas las iteraciones de Tr sobre esta oración, esto es:

$$Q^{\mathcal{N}} = \{Tr^n(' \exists x(P(x) \wedge Tr(x))') : n \in \mathbb{N}\}$$

Luego, (3.14) afirma que hay al menos un número de iteraciones de Tr sobre (3.13) que resulta en una fórmula verdadera. Una vez más, como (3.13) es inestable, (3.14) también debería serlo. Sin embargo, (3.14) se estabiliza dentro de la extensión del predicado veritativo las secuencias de Yaqūb; no resulta paradójica en ese sistema, contrario a nuestras expectativas.

Puesto que tanto M como M' son inestables, existen dos tipos de bootstrappers diferentes en cuanto ellas: o bien se encuentran ambas dentro o ambas fuera de la extensión de Tr , o bien están desfasadas, esto es, una dentro y una fuera. Dado que de acuerdo con Yaqūb todo bootstrapper debe permanecer en cada secuencia al menos κ pasos, lo mejor

que podemos hacer para desestabilizar a (3.14) es seguir la condición 3 de la definición 3.37 alternando los bootstrappers $\{M\} = \Gamma_0$ y $\emptyset = \Gamma_1 = \Gamma_2$ $\omega + 1$ veces en segmentos de longitud κ . Como veremos a continuación, nuestro mejor intento posible resultará infructuoso.

Comencemos por un segmento de $\kappa\gamma$ a $\kappa\gamma + \kappa$ para algún ordinal γ con el bootstrapper $\{M\}$, donde los mentirosos están fuera de fase y supongamos que previamente no lo estaban, con lo cual (3.2) y todas sus iteraciones de verdad se habían estabilizado localmente como verdaderas.³⁸ Luego, (3.2) se estabilizará como falsa en todo el segmento. $Tr(3.2)$ copia el comportamiento de (3.2) recién un paso más adelante en cada secuencia, en virtud de la definición 3.5. Por tanto, (3.12) será falsa en $\kappa\gamma$ pero se estabilizará como verdadera en los pasos subsiguientes. Puesto que todas las iteraciones del predicado de verdad sobre una oración copian el comportamiento de esta última tantos pasos más adelante como ocurrencias del predicado veritativo contengan, cada $Tr^n(3.12)$ con $n \in \mathbb{N}$ se estabilizará como verdadera en $\kappa\gamma + n + 1$, y siempre habrá una de ellas que sea verdadera en cada paso del segmento en cuestión. Como $P(x)$ es verdadero precisamente de todas estas oraciones, (3.13) será siempre verdadera en este segmento de longitud κ , al igual que todas sus predicaciones de verdad. Y dado que $Q(x)$ es verdadero de todas estas, (3.14) también será verdadera a lo largo del segmento.

De acuerdo con la condición 3 de la definición 3.37, en $\kappa\gamma + \kappa$ debemos emplear $\emptyset$ como bootstrapper, pues es un segmento impar. Ambos mentirosos dejarán pues de estar desfasados. Las elecciones que aquí hagamos para las otras proposiciones son irrelevantes en tanto se han estabilizado previamente. Sin embargo, es visible que este desfasaje en nada cambiará el comportamiento de las otras oraciones excepto (3.2) a lo largo de este nuevo segmento de longitud κ : tanto (3.13) como (3.14) seguirán siendo verdaderas. Y es evidente que lo mismo sucederá en $\kappa\gamma + \kappa 2$, donde se repiten los valores de $\kappa\gamma$.

En $\kappa\gamma + \kappa\omega$ tenemos la última oportunidad de desestabilizar (3.14). Para hacerlo, debemos tener que, para todo $n \in \mathbb{N}$, $Tr^n(3.13)$ es falsa, pero ningún bootstrapper podrá lograr eso, ya que estas oraciones se han estabilizado como verdaderas en los estados previos a $\kappa\gamma + \kappa\omega$ y están por ello fuera de nuestro alcance. Lo mejor que podemos hacer aquí es desestabilizar (3.13), colocando a cada $Tr^n(3.12)$ fuera de la extensión de Tr , lo cual es posible pues estas oraciones fluctuaron en un punto en cada segmento previo de longitud κ : son inestables desde el punto de vista de $\kappa\gamma + \kappa\omega$. Pero esto no es suficiente, pues sólo conduce a que (3.13) sea falsa en un número finito de estados subsiguientes, volviendo a estabilizarse como verdadera después, con lo cual (3.14) será establemente verdadera también en este segmento. La siguiente tabla ilustra el razonamiento que llevamos a cabo en los párrafos anteriores.

³⁸Como veremos inmediatamente, esta suposición no afectará en nada nuestro intento, pues sea cual fuere el estado de estas oraciones en $\kappa\gamma$, tras un número finito de pasos el resultado será el mismo.

	$\kappa\gamma$	$\kappa\gamma + \kappa$	$\kappa\gamma + \kappa^2$	$\dots$	$\kappa\gamma + \kappa\omega$
M	f t f t...	t f t f...	f t f t...		t f t f...
M'	t f t f...	t f t f...	t f t f...		t f t f...
(3.2)	f f f f...	t t t t...	f f f f...		t t t t...
$Tr(3.2)$	t f f f...	f t t t...	t f f f...		f t t t...
(3.12)	f t t t...	f t t t...	f t t t...		f t t t...
$Tr(3.12)$	t f t t...	t f t t...	t f t t...		f f t t...
$Tr^2(3.12)$	t t f t...	t t f t...	t t f t...		f f f t...
(3.13)	t t t t...	t t t t...	t t t t...		f t t t...
$Tr(3.13)$	t t t t...	t t t t...	t t t t...		t f t t...
$Tr^2(3.13)$	t t t t...	t t t t...	t t t t...		t t f t...
(3.14)	t t t t...	t t t t...	t t t t...		t t t t...

La idea que guía este ejemplo está basada en que la desestabilización progresiva que se despliega desde las primeras oraciones— M y M' —no tiene alcance suficiente para desestabilizar a (3.14). Sería preciso que consideráramos segmentos sucesivos más y más largos para evaluar esta oración en el segmento ω ; pero entonces podríamos pensar en otra oración que no llegue a desestabilizarse sino al tener en cuenta segmentos aún más largos, y así sucesivamente.

De acuerdo con Chapuis, las fallas del enfoque de Yaqūb yacen en que los bootstrappers son los responsables de introducir la variedad de opciones necesaria para evitar errores de clasificación, de la cual los sistemas de Herzberger, Gupta y Belnap carecen. Propone, por tanto, un modo alternativo de llevar a cabo esa tarea.

DEFINICIÓN 3.42 (Coherencia) Sea $[\mathcal{M}, S]$ una secuencia de revisión. La hipótesis $S' \subseteq |\mathcal{M}|$ es *coherente* con $[\mathcal{M}, S]$ si y sólo si, para toda fórmula A de L^+ :

- si existe un ordinal α tal que para todo $\beta \geq \alpha$ $(\mathcal{M}, S_\beta) \models A$, $A \in S'$;
- si existe un ordinal α tal que para todo $\beta \geq \alpha$ $(\mathcal{M}, S_\beta) \not\models A$, $A \notin S'$.

Luego, una hipótesis es coherente con una secuencia si contiene todas las oraciones establemente verdaderas en esa secuencia y ninguna de las establemente falsas.

DEFINICIÓN 3.43 (Secuencias completamente variadas) $[\mathcal{M}, S]$ es una *secuencia completamente variada* si y sólo si es una secuencia de revisión y para toda hipótesis S' coherente con $[\mathcal{M}, S]$ y para todo ordinal α existe un $\beta \geq \alpha$ tal que $S_\beta = S'$.

Una secuencia de revisión es pues completamente variada siempre y cuando toda hipótesis coherente con ella esté presente indefinidamente en ella, esto es, no exista una etapa en la secuencia a partir de la cual no podamos encontrarla ya. Estas secuencias fueron originalmente consideradas por Gupta & Belnap (1993) y bautizadas por ellos.

La propuesta de Chapuis no consiste en una nueva política de bootstrapping. La presencia indefinida de hipótesis coherentes con la secuencia puede darse tanto en etapas límites como sucesoras. Además, las secuencias completamente variadas no pueden

construirse en cada etapa sino que son seleccionadas de la clase de secuencias de revisión definidas en 3.5. Las políticas de bootstrapping de Belnap y Yaqūb resultan demasiado permisivas, pues incluyen secuencias en las cuales muchas extensiones coherentes son ignoradas. Así, dan lugar a errores de clasificación. Como veremos, las secuencias de Chapuis evitan todas las dificultades de este tipo que hemos mencionado.

EJEMPLO 3.44 (Bicondicionales de mentirosos) Sea $[\mathcal{N}, S]$ una secuencia completamente variada. Puesto que en cualquier secuencia las oraciones del mentiroso son inestables, sabemos que existen hipótesis S' y S'' coherentes con $[\mathcal{M}, S]$ tales que M pertenece a S pero no M' y ninguna de las dos pertenece a S'' . Por la definición 3.43, S' y S'' estarán presentes a lo largo de toda la secuencia. Dado que en $(\mathcal{N}, S')$ $M \equiv M'$ es falsa, mientras que en $(\mathcal{N}, S'')$ es verdadera, (3.2) es inestable en $[\mathcal{N}, S]$, con lo cual es clasificada como paradójica por el sistema de Chapuis.

EJEMPLO 3.45 (Gupta puzzles) Sean J_a y J_b como en el ejemplo 3.30. Allí hemos visto que, una vez que estas oraciones están fuera de fase, J_a se estabiliza como verdadera y J_b como falsa. Supongamos entonces que existe una secuencia completamente variada $[\mathcal{M}, S]$ en la cual esto no sucede, esto es, tal que para todo α o bien ambas oraciones pertenecen a S_α o bien ninguna lo hace. Como muestra la tabla del ejemplo 3.30, J_a y J_b oscilarán sincronizadamente a lo largo de toda la secuencia. Luego, existe una hipótesis S' coherente con $[\mathcal{M}, S]$ tal que $J_a \in S'$ y $J_b \notin S'$. En virtud de la definición 3.43, debe existir un β tal que $S' = S_\beta$, lo cual es un absurdo, pues dijimos que las oraciones estaban en fase a lo largo de toda la secuencia.

EJEMPLO 3.46 (Ciclos de honestos) Sean H_a y H_b como en el ejemplo 3.36. Dejaremos que el lector compruebe que estas oraciones son clasificadas como tf-caprichosas por el sistema de Chapuis.

Aún no han surgido errores de clasificación en la teoría de la revisión basada en secuencias completamente variadas. No obstante, no tenemos ninguna certeza de que no los haya ni de que no vayamos a encontrarlos en el futuro.

Ejercicios para la subsección 3.1

1. Muestre que mediante un argumento intuitivo que haga uso del esquema T extensional las oraciones K_a y K_b del ejemplo 3.32 resultan verdadera y falsa, respectivamente. Pruebe que el sistema de Herzberger las clasifica como categóricamente verdadera y categóricamente falsa, correspondientemente.
2. Muestre que mediante un argumento intuitivo que haga uso del esquema T extensional las oraciones P_a y P_b del ejemplo 3.33 resultan verdadera y falsa, respectivamente. Pruebe que el sistema de Herzberger las clasifica como categóricamente verdadera y categóricamente falsa, correspondientemente.

3. Pruebe que en el sistema de Gupta la oración (3.2) es clasificada como tf-caprichosa.
4. Pruebe que el sistema de Gupta arroja los mismos resultados que el de Herzberger para los conjuntos de oraciones presentados en los ejemplos 3.30, 3.32 y 3.33.
5. Pruebe que tanto el sistema de Yaqūb como el de Chapuis clasifican oraciones como H_a y H_b en (3.10) como tf-caprichosas.
6. Muestre que no toda secuencia de Herzberger es una secuencia de Gupta ni, viceversa, toda secuencia de Gupta es una secuencia de Herzberger.
7. Pruebe que:
 - (a) El conjunto de oraciones de L^+ que son establemente verdaderas en el sistema de Belnap está incluido en el de aquellas que lo son en el de Gupta, que a su vez está incluido en el conjunto de oraciones establemente verdaderas en el sistema de Herzberger.
 - (b) El conjunto de oraciones de L^+ que son establemente falsas en el sistema de Belnap está incluido en el de aquellas que lo son en el de Gupta, que a su vez está incluido en el conjunto de oraciones establemente falsas en el sistema de Herzberger.
 - (c) El conjunto de oraciones de L^+ que son paradójicas en el sistema de Belnap está incluido en el de aquellas que lo son en el de Gupta, que a su vez está incluido en el conjunto de oraciones paradójicas en el sistema de Herzberger.
8. Indique qué categorías de las siete introducidas en la definición 3.10 le son asignadas a las oraciones $M \wedge M'$ y $M \vee M'$, donde M y M' son como en (2.6) y (3.1), respectivamente, en cada uno de los sistemas presentados en esta subsección.
9. Sea F una oración de L^+ tal que, para todo $S \subseteq |\mathcal{N}|$:

$$(\mathcal{N}, S) \models F \equiv (Tr('F') \supset Tr(Tr('F')))) \quad (3.15)$$

- (a) Desarrolle un argumento intuitivo que muestre que existe una única asignación estable de valores de verdad clásicos para F y señale cuál es esta asignación.
 - (b) Indique qué categoría de las dadas por la definición 3.10 asigna cada uno de los sistemas presentados en esta subsección a F .
 - (c) ¿Debería una teoría de la verdad asignar a F la única asignación estable de valores de verdad clásicos posible o debería permitir que F sea inestable en algunas secuencias?
10. ¿Debería una teoría de la verdad solucionar los Gupta puzzles como quiere Yaqūb o debería dar lugar a las inestabilidades que esos conjuntos de oraciones infundadas presentan en sistemas como el de Belnap? ¿O la respuesta a la pregunta anterior debería variar según el conjunto de oraciones infundadas de que se trate?

3.2 Venganzas

Entendemos que los fenómenos conocidos como ‘venganzas’ o ‘revanchas’ surgen dentro de teorías que pretenden ser semánticamente autosuficientes, esto es, que pretenden expresar correctamente sus propias categorías semánticas, y sin embargo no logran expresar esos mismos conceptos que ellas generan. Bajo esta perspectiva, entendemos que la teoría revisionista de la verdad carece de revanchas, ya que su objetivo es, como hemos señalado en la introducción a este capítulo, dar con una teoría que dé cuenta del uso del predicado de verdad para un lenguaje que contenga el suyo propio, y no simultáneamente de todos los predicados semánticos que puedan existir, inclusive los que ella misma genera.

Si este último fuera el objetivo de la teoría de la revisión, estaría en efecto plagada de venganzas. Gupta & Belnap (1993) mismos señalan este hecho y lo ejemplifican mediante la noción de categoricidad—aplicable a todas las oraciones categóricamente verdaderas y categóricamente falsas—a la cual su teoría da lugar.³⁹ El siguiente teorema les pertenece.

TEOREMA 3.47 (Indefinibilidad de la categoricidad) No existe un predicado Ca de L^+ que sea verdadero en $\mathcal{N}$ de todas las oraciones categóricas en $\mathbf{T}^\#$ en $\mathcal{N}$.⁴⁰

Prueba: Supongamos que Ca expresa en $\mathcal{N}$ la noción de categoricidad en $\mathbf{T}^\#$ en $\mathcal{N}$. Gracias al lema de diagonalización, sabemos de la existencia de una oración Q en L^+ tal que para todo $S \subseteq |\mathcal{N}|$:

$$(\mathcal{N}, S) \models Q \equiv (\neg Ca('Q') \vee \neg Tr('Q')) \quad (3.16)$$

O bien $Q \in \mathbf{V}_\mathcal{N}^\#$ o bien $Q \notin \mathbf{V}_\mathcal{N}^\#$. En el primer caso, también $Tr('Q') \in \mathbf{V}_\mathcal{N}^\#$ y, además, Q será categórica, con lo cual $Ca('Q') \in \mathbf{V}_\mathcal{N}^\#$. Pero por (3.16) esto implica que $Q \notin \mathbf{V}_\mathcal{N}^\#$, lo cual es un absurdo. Luego, $Q \notin \mathbf{V}_\mathcal{N}^\#$, esto es, o bien $\neg Q \in \mathbf{V}_\mathcal{N}^\#$ o bien Q no es categórica. Si lo primero sucede, $\neg Tr('Q') \in \mathbf{V}_\mathcal{N}^\#$. En el segundo caso, $\neg Ca('Q') \in \mathbf{V}_\mathcal{N}^\#$. Por (3.16), en ambos casos tenemos que $Q \in \mathbf{V}_\mathcal{N}^\#$, lo cual es imposible. ■

Gupta & Belnap llaman a Q ‘mentiroso reforzado’, pues su estructura es en todo semejante a la del mentiroso, excepto en que en lugar de utilizar sólo el predicado de verdad emplea también el de categoricidad. Al igual que en el caso del mentiroso tradicional, la reacción de Gupta & Belnap ante el argumento dado en la prueba del teorema 3.47 no consiste en sostener que es imposible tener tal noción para un lenguaje dentro del mismo lenguaje sin efectuarle modificación alguna. En cambio, sostienen que el argumento entiende incorrectamente el bicondicional involucrado en la definición de la noción de categoricidad como si fuera un bicondicional material:

$$Ca('A') \text{ si y sólo si } A \in \mathbf{V}_\mathcal{N}^\# \text{ ó } \neg A \in \mathbf{V}_\mathcal{N}^\#$$

³⁹Véase Gupta & Belnap (1993), cap. 6, sec. D y cap. 7, sec. I.

⁴⁰Mediante pruebas prácticamente idénticas es posible establecer resultados análogos para $\mathbf{T}^*$ y $\mathbf{T}^c$.

donde A es cualquier oración de L^+ . De acuerdo con estos autores, la noción de categoricidad es análoga al concepto de verdad y tan circular como él. Requiere por tanto una definición circular dada por un esquema como el presentado en (2.1).

Sea L^{++} es resultado de incorporar a L^+ un nuevo símbolo de predicado monádico Ca . Consideraremos las siguientes definiciones parciales para Ca :

$$Ca('A') =_{df} A \text{ es categórica en } \mathbf{T}^\# \text{ en } \mathcal{M} \quad (3.17)$$

donde A es una oración cualquiera de L^{++} y $\mathcal{M}$ una interpretación de L . Como la categoricidad es un concepto circular, para evaluar si una oración es o no categórica precisamos una hipótesis que revisar. Basándonos en (3.17), podemos definir una regla de revisión de hipótesis $\psi_{\mathcal{M}}$ que haga esta tarea.

DEFINICIÓN 3.48 ($\psi_{\mathcal{M}}$) $\psi_{\mathcal{M}}$ es una *regla de revisión* basada en $\mathcal{M}$ tal que $\psi_{\mathcal{M}} : \mathcal{P}(|\mathcal{M}|) \rightarrow \mathcal{P}(|\mathcal{M}|)$ y

$$\psi_{\mathcal{M}}(S) = \{d : A \in \mathbf{V}_{(\mathcal{M};S)}^\# \text{ ó } \neg A \in \mathbf{V}_{(\mathcal{M};S)}^\# \text{ y } d \text{ es el número de Gödel de } A\}$$

donde $(\mathcal{M}; S)$ es la expansión de $\mathcal{M}$ a un modelo parcial de L^{++} que asigna S a Ca como su interpretación y deja Tr sin interpretar.⁴¹

Luego, dada una extensión hipotética S para Ca , evaluamos si una oración cae o no bajo la extensión de Ca de acuerdo con esta hipótesis estudiando si resulta categórica en el sistema $\mathbf{T}^\#$ empleando como modelo base a $(\mathcal{M}; S)$. Las definiciones de ‘secuencia de revisión basada en la regla $\psi_{\mathcal{M}}$ ’ y las consecuentes nociones de estabilidad, estabilidad cercana y validez son análogas a las que hemos presentado en la sección 2.

Parece que hemos logrado un lenguaje L^{++} que contiene no sólo su propio predicado veritativo sino también su propio predicado de categoricidad. Esta teoría puede lidiar con oraciones como el mentiroso reforzado, como muestra el siguiente ejemplo.

EJEMPLO 3.49 (El mentiroso reforzado) Mediante el lema de diagonalización, tenemos una oración $M^\star$ en L^{++} tal que para todo $S \subseteq |\mathcal{N}|$:

$$(\mathcal{N}; S) \models M^\star \equiv (\neg Ca('M^\star') \vee \neg Tr('M^\star')) \quad (3.18)$$

$M^\star$ oscila a lo largo de todo el proceso de revisión para Ca , independientemente de la hipótesis inicial con la cual comencemos. Si elegimos una hipótesis S a la cual esta oración pertenezca, $M^\star$ se comportará como una oración del mentiroso en todo el proceso de revisión para Tr basado en $(\mathcal{N}; S)$, pues $\neg Ca('M^\star')$ resultará siempre falsa allí. Luego, quedará afuera en el siguiente paso de revisión. Si en cambio, comenzamos con una

⁴¹Nótese que cuando hablamos de este nuevo tipo de expansiones de un modelo $\mathcal{M}$ para L escribimos ‘;’ en lugar de ‘,’.

hipótesis S' tal que $M^* \notin S'$, $\neg Ca('M^*')$ será verdadera a lo largo de todo el proceso de revisión para Tr basado en $(\mathcal{N}; S')$ y, por ende, lo mismo sucederá con M^* . Luego ingresará en la extensión de Ca en el paso siguiente, repitiéndose el razonamiento que llevamos a cabo primero.

No obstante, esta teoría no logra dar cuenta correctamente de otras oraciones relevantes, como el honesto reforzado, como veremos a continuación.

EJEMPLO 3.50 (El honesto reforzado) Una vez más, el lema de diagonalización garantiza la existencia de una oración H^* en L^{++} tal que para todo $S \subseteq |\mathcal{N}|$:

$$(\mathcal{N}; S) \models H^* \equiv (Ca('H^*') \vee \neg Tr('H^*')) \quad (3.19)$$

Como podrá comprobar el lector en los ejercicios para esta subsección, H^* resulta verdadera si pertenece a la hipótesis inicial del proceso de revisión para Ca y falsa en caso contrario. Luego, no es cierto que $\neg Ca('H^*')$ de acuerdo con la teoría revisionista de la categoricidad. La patologicidad del honesto reforzado no puede ser descripta por ella.

De acuerdo con Gupta & Belnap (1993), lejos de ser una dificultad, esto es una manifestación de la necesidad de nociones de orden superior para dar cuenta de la patologicidad de oraciones que involucran conceptos semánticos.

En el caso de la teoría revisionista de la verdad, no resulta correcto afirmar que el honesto no es verdadera ni falsa para dar cuenta de su carácter patológico, pues $Tr('H') \vee \neg Tr('H')$ es verdadera allí. Más bien necesitamos predicados como 'categórica' para esa tarea. Del mismo modo, el comportamiento patológico de oraciones como H^* , cuya patologicidad se debe a la presencia del predicado de categoricidad, precisa nociones de orden superior para ser explicado, y así sucesivamente. Gupta & Belnap sostienen que estas nociones son semejantes a la verdad: todas ellas son circulares, dan lugar a oraciones patológicas si el lenguaje tiene capacidades autorreferenciales y apuntan hacia la existencia de nociones de orden aún superior para explicarlas.

Si bien la propuesta de Gupta & Belnap (1993) no es capaz de recoger estos conceptos en una misma teoría sino que recurre a jerarquías, no consideramos que se encuentre por ello en desventaja con respecto a otras teorías semánticas. Pues éste nunca fue su objetivo sino dar cuenta del comportamiento de predicados semánticos como la verdad para un lenguaje dentro del lenguaje mismo, y ha sido perfectamente alcanzado. Aún más, Gupta & Belnap han mostrado cómo llevar a cabo esta tarea para conceptos de orden superior, como el de categoricidad y cualquier otro concepto semejante capaz de dar lugar a oraciones patológicas.

Ejercicios para la subsección 3.2

1. Pruebe resultados análogos al teorema 3.47 para las nociones de validez y paradojicidad en $\mathbf{T}^*$ en $\mathcal{N}$.

2. Defina las nociones de paradojicidad y validez y dé una regla de revisión para cada una de estas nociones como lo hemos hecho para el concepto de categoricidad en (3.17) y en la definición 3.48, correspondientemente.
3. Sea H^* como en el ejemplo 3.50. Pruebe que esta oración es tf-caprichosa de acuerdo con la teoría de la categoricidad presentada en esta subsección.

4 Conclusiones

El propósito de la teoría revisionista es lograr describir el uso del concepto de verdad transparente dentro de un lenguaje clásico que contiene su propio predicado veritativo. Tras destacar el carácter circular de la verdad, concluimos la imposibilidad de dar con una extensión definitiva para dicho concepto y la necesidad, en cambio, de considerar secuencias de extensiones o modelos para este lenguaje. Atendiendo a regularidades exhibidas por dichas secuencias, obtuvimos un conjunto de oraciones correctamente asertables.

Más aun, notamos que la consideración de tales secuencias permite una clasificación de todas las oraciones del lenguaje en siete categorías semánticas capaces de distinguir, entre otras cosas, entre oraciones patológicas y no patológicas.

No obstante, vimos que no existe un único modo de llevar adelante esta tarea de clasificación. Presentamos diversas alternativas y los respectivos desafíos que ellas afrontan. Si bien hemos puesto en duda la necesidad de dar una solución a los problemas planteados por los Gupta puzzles, creemos que es preciso clasificar correctamente los bicondicionales de oraciones mentirosas como paradójicos y los ciclos de honestos como tf-caprichosos. Tanto el sistema de Yaqūb como el de Chapuis cumplen este cometido. Sin embargo, como muestra el ejemplo 3.41, el primero clasifica erróneamente otras oraciones, razón por la cual—dejando de lado los Gupta puzzles—la teoría de Chapuis parece ser por el momento la mejor candidata.

A diferencia de lo sucede con otras teorías presentadas en este libro, la teoría revisionista no se propone ser semánticamente autosuficiente y, por ende, las venganzas generadas por oraciones como (3.19) no constituyen una auténtica amenaza para ella.

Finalmente, en el apéndice veremos que el afán de obtener una teoría semánticamente autosuficiente o de privilegiar la certeza empistémica por sobre el poder expresivo puede llevarnos a optar por un abordaje de tipo formal, que intente axiomatizar el conjunto de oraciones correctamente asertables de acuerdo con la teoría de la revisión.

Apéndice: Un punto de vista axiomático

Razones para adoptar un enfoque axiomático

Se llama ‘teorías semánticas’ a aquellas teorías de la verdad que proporcionan una clase de modelos. Todas las teorías presentadas en los cuerpos principales de los capítulos de

este libro son consideradas teorías semánticas en la literatura. Tal vez esta clasificación no sea del todo justa para con la teoría revisionista de la verdad, en tanto su resultado no es una clase de modelos sino un conjunto de oraciones correctamente asertables. No obstante, dado que recurre a clases de modelos para obtener este conjunto de oraciones, puede ser laxamente considerada una teoría semántica, y así lo haremos aquí.

En todas las teorías semánticas de la verdad, la consideración de los modelos para un lenguaje que contiene su propio predicado veritativo se lleva a cabo desde un lenguaje ‘esencialmente más rico’, esto es, con mayor capacidad expresiva que el lenguaje base.⁴² En caso de la teoría revisionista, mientras que el lenguaje base es el de la aritmética junto con un predicado de verdad, el metalenguaje es aún más expresivo que el de la teoría de conjuntos, pues cuantifica sobre secuencias cuyo ordinal es el de todos los conjuntos.

Es el proyecto de muchos teóricos de la verdad actuales dar con un lenguaje semánticamente autosuficiente, evitando todo tipo de revanchas—aunque no el de Gupta & Belnap (1993), como señalamos en la sección 3.2. Un modo de llevar a cabo esta tarea en el marco de la teoría de la revisión consiste en axiomatizar la teoría en el mismo lenguaje para el cual está formulada. De ese modo, no habría una distinción entre lenguaje y metalenguaje. El conjunto de oraciones correctamente asertables de L^+ se obtendría, ya no observando clases de modelos, sino haciendo derivaciones.

La obtención de teorías semánticamente suficientes no es la única razón por la cual un abordaje axiomático de la verdad podría resultar atractivo. Existen factores epistémicos que podrían embarcarnos en el proyecto formal. Mientras que las teorías semánticas de la verdad no proporcionan un método seguro para extraer conclusiones a partir de clases de modelos, los sistemas axiomáticos son candidatos ideales.

Naturalmente, esta seguridad se obtiene a expensas de una gran pérdida de información. Ninguna teoría semántica puede ser completamente capturada por una axiomatización recursiva, *i.e.* para la cual contemos con un algoritmo que nos permita saber si una fórmula dada es o no un axioma de la teoría, que es esencial si lo que buscamos es certeza epistémica. Pues todas parten de un modelo base como $\mathcal{N}$, el cual no es recursivamente axiomatizable. No obstante, existen mejores y peores axiomatizaciones. Lamentablemente, como veremos a continuación, la teoría de la revisión no admite versiones formales tan buenas como, por ejemplo, la teoría de punto fijo de Kripke, presentada en el capítulo anterior.

Una versión axiomática de $V_{\mathcal{N}}^{\#}$

Harvey Friedman & Michael Sheard (1987) han presentado una teoría axiomática de la verdad para un lenguaje con su propio predicado veritativo, hoy conocida como *FS*. La presentación que hacemos a continuación es ligeramente diferente pero equivalente a la

⁴²Véase el capítulo primero de este libro.

original.⁴³

DEFINICIÓN 3.51 (*FS*) *FS* es una teoría de primer orden formulada en L^+ cuyos axiomas son los de la aritmética de Peano de primer orden con inducción sobre todas las fórmulas de L^+ , $(Tr\neg)$, $(Tr\wedge)$, $(Tr\forall)$,

$$(\mathbf{At}^+) \quad \forall x(At(x) \supset (Tr(x) \equiv Ver(x)))$$

y cuyas reglas de inferencia son las de necesidad y co-necesidad. Si A es una oración de L^+ :

(**NEC**) De una prueba de A inferir $Tr(A)$;

(**CONEC**) De una prueba de $Tr(A)$ inferir A .

Visiblemente, todos los axiomas de *PA* pertenecen a $\mathbf{V}_{\mathcal{N}}^{\#}$, pues son verdaderos en todos los pasos de todas las secuencias basadas en $\mathcal{N}$. El teorema 3.24 establece que $(Tr\neg)$, $(Tr\wedge)$ y $(Tr\forall)$ también pertenecen a $\mathbf{V}_{\mathcal{N}}^{\#}$. Y el teorema 3.10 muestra que este conjunto está cerrado bajo (**NEC**) y (**CONEC**). Finalmente, es inmediato—y podrá comprobar el lector—que $(\mathbf{At}^+)$, que afirma que todas las oraciones atómicas de L verdaderas en $\mathcal{N}$ son verdaderas de acuerdo con la teoría, también pertenece a $\mathbf{V}_{\mathcal{N}}^{\#}$. Consecuentemente, el teorema 3.18 nos asegura que *FS* es cercanamente establemente verdadera y, por ende, un buen candidato para la axiomatización de $\mathbf{V}_{\mathcal{N}}^{\#}$.

Al igual que $\mathbf{V}_{\mathcal{N}}^{\#}$, *FS* es ω -inconsistente.⁴⁴ Luego, todos los problemas que esta característica trae a $\mathbf{T}^{\#}$ son heredados por *FS*.

Una versión axiomática de $\mathbf{V}_{\mathcal{N}}^*$

Puesto que los axiomas composicionales no son establemente verdaderos, como indican el teorema 3.24 y el corolario 3.25, *FS* no sirve como axiomatización de $\mathbf{V}_{\mathcal{N}}^*$ y encontrar un buen candidato parece complicado. Horsten, Graham Leigh, Leitgeb & Philip Welch (2012) han propuesto la siguiente axiomatización, no de $\mathbf{V}_{\mathcal{N}}^*$, sino del sistema de verdad estable de Herzberger, donde las secuencias consideradas aplican únicamente el bootstrapper vacío y, por ende, tornan verdaderos principios como $(Tr\wedge)$ y $(Tr\forall)$.⁴⁵

DEFINICIÓN 3.52 (*FSPos*) *FSPos* es una teoría de primer orden formulada en L^+ cuyos axiomas son los de la aritmética de Peano de primer orden con inducción sobre todas las fórmulas de L^+ , $(\mathbf{At}^+)$, $(Tr\wedge)$, $(Tr\forall)$, (**LNC**),

$$(\mathbf{At}^-) \quad \forall x(At(x) \supset (Tr(\neg x) \equiv Fal(x)))$$

⁴³Véase Halbach (1994) para una prueba.

⁴⁴La prueba es análoga a la del teorema 3.27.

⁴⁵Véase la subsección 3.1.1 de este capítulo.

(LIC) $\forall x \forall y (Sent_{L^+}(x) \supset ((Tr(\dot{\neg}x) \vee Tr(\dot{\neg}y)) \supset Tr(\dot{\neg}(x \wedge y))))$

(PMP) $\forall x \forall y (Sent_{L^+}(x) \supset ((Tr(x) \wedge Tr(\dot{\neg}(x \wedge \dot{\neg}y)) \supset Tr(y)))$

(LIQ) $\forall x \forall v ((var(v) \wedge Sent_{L^+}(x(\dot{0}/v))) \supset (\exists y Tr(\dot{\neg}x(\dot{y}/v)) \supset Tr(\dot{\neg}\dot{\forall}vx)))$

y cuyas reglas de inferencia son (NEC) y (CONCEC).

(At^-) expresa que la negación de una fórmula atómica de L será verdadera cuando esta fórmula sea falsa. (LIC)⁴⁶ expresa que si algunos de los conjuntos es falso será falsa la conjunción, mientras que (PMP)⁴⁷ expresa la validez de un principio semejante al Modus Ponens, pero anidado dentro del predicado veritativo. Finalmente, (LIQ)⁴⁸ expresa la interacción entre los cuantificadores.

Las conversas de (LIC) y (LIQ) no han sido incluidas pues no son establemente verdaderas en toda secuencia de Herzberger. El teorema 3.19 establece la validez de (NEC) y (CONCEC) en ese sistema. La prueba de que los axiomas de $FSPos$ para Tr son establemente verdaderos quedará a cargo del lector.

$FSPos$ es un subsistema de FS en el cual sólo axiomas positivos han sido incluidos, esto es, se ha prescindido de ($Tr\neg$), que nos permitía incluir una oración dentro de la extensión del predicado veritativo cada vez que supiéramos que su negación estaba fuera. Para compensar el poder de prueba perdido por el abandono de este principio, se han incorporado otros que nos indican cómo lidiar con fórmulas negadas en algunos casos particulares.

Resta proporcionar una axiomatización de $\mathbf{V}_{\mathcal{N}}^c$. Aquí no es posible recurrir a secuencias de Herzberger, pues éstas no son máximamente consistentes. Sin embargo, la tarea es muy sencilla a la luz de los teoremas 3.19 y 3.24. La dejaremos por tanto a cargo del lector.

Tanto FS como $FSPos$ son axiomatizaciones muy incompletas de sus respectivos conjuntos. Un modo de verlo es notar que, por ejemplo, la oración $Tr^n('0 = 0')$, que es establemente verdadera y, por ende, también cercanamente establemente verdadera, no es un teorema de FS ni, *a fortiori*, de $FSPos$. En general, ninguna de las dos axiomatizaciones que hemos presentado aquí es capaz de probar iteraciones transfinitas del predicado veritativo,⁴⁹ mientras que la teoría revisionista incluye a todas las iteraciones transfinitas correctas en el conjunto de oraciones asertables. En cambio, la axiomatización de la teoría de punto fijo de Kripke, KF , sí es capaz de probar iteraciones transfinitas del predicado de verdad y, aún más, todos los modelos de esta teoría son puntos fijos en el sistema kripkeano, como se ha mencionado en el segundo capítulo de este libro. No existe ningún resultado análogo para la teoría de la revisión.

⁴⁶Sus siglas en inglés están por '*Law of Interdefinition of Connectives*'.

⁴⁷Sus siglas en inglés están por '*Pseudo Modus Ponens*'.

⁴⁸Sus siglas en inglés están por '*Law of Interdefinition of Quantifiers*'.

⁴⁹Véase Horsten (2011), prop. 48 para una prueba

Ejercicios para el apéndice

1. Pruebe que (At^+) pertenece a $\mathbf{V}_{\mathcal{N}}^{\#}$.
2. Pruebe que (At^+) , (At^-) , $(Tr\wedge)$, (LIC), (PMP), (LNC), $(Tr\forall)$ y (LIQ) son establemente verdaderos en toda secuencia de Herzberger basada en $\mathcal{N}$.
3. Pruebe que las conversas de (LIC) y (LIQ) no son establemente verdaderas en toda secuencia de Herzberger basada en $\mathcal{N}$.
4. Proporcione una axiomatización razonable para el conjunto $\mathbf{V}_{\mathcal{N}}^c$.

Bibliografía

- Barrio, E.: “Theories of Truth without Standard Models and Yablo’s Sequences”, *Studia Logica*, vol. 96, 2010, pp. 375-391.
- Belnap, N. D.: “Gupta’s Rule of Revision Theory of Truth”, *Journal of Philosophical Logic*, vol. 11, 1982, pp. 113-116.
- Chapuis, A.: “Alternative Revision Theories of Truth”, *Journal of Philosophical Logic*, vol. 25, 1996, pp. 399-423.
- Cook, R. T.: “Counterintuitive Consequences of the Revision Theory of Truth”, *Analysis*, vol. 62, 2002, pp. 16-22.
- Cook, R. T.: “Still Counterintuitive: A Reply to Kremer”, *Analysis*, vol. 63, 2003, pp. 257-261.
- Friedman, H. & Sheard, M.: “An Axiomatic Approach to Self-Referential Truth”, *Annals of Pure and Applied Logic*, vol. 33, 1987, pp. 1-21.
- Gupta, A.: “Truth and Paradox”, *Journal of Philosophical Logic*, vol. 11, 1982, pp. 1-60.
- Gupta, A. & Belnap, N. D.: *The Revision Theory of Truth*, Cambridge, MIT Press, 1993.
- Herzberger, H. G.: “Notes on Naive Semantics”, *Journal of Philosophical Logic*, vol. 11, 1982, pp. 61-102.
- Herzberger, H. G.: “Naive Semantics and the Liar Paradox”, *The Journal of Philosophy*, vol. 79, 1982, pp. 479-497.
- Halbach, V.: “A System of Complete and Consistent Truth”, *Notre Dame Journal of Formal Logic*, vol. 35, 1994, pp. 311-327.
- Halbach, V.: *Axiomatic Theories of Truth*, Cambridge, Cambridge University Press, 2011.
- Halbach, V. & Horsten, L.: “The Deflationist’s Axioms for Truth”, en *Deflationism and Paradox*, editado por Armour-Garb, B. & Beall, J. C., Oxford, Oxford University

Press, 2005, pp. 203-217.

Horsten, L.: *The Tarskian Turn. Deflationism and Axiomatic Truth*, Cambridge, MIT Press, 2011.

Horsten, L., Leigh, G., Leitgeb, H. & Welch, P.: "Revision Revisited", *Review of Symbolic Logic*, vol. 5, 2012, pp. 642-664.

Kremer, M.: "Intuitive Consequences of the Revision Theory of Truth", *Analysis*, vol. 62, 2002, pp. 330-336.

Kripke, S.: "Outline of a Theory of Truth", *The Journal of Philosophy*, vol. 72, 1975, pp. 690-716.

Leitgeb, H.: "What Theories of Truth Should Be Like (But Cannot Be)", *Philosophy Compass*, vol. 2, 2007, pp. 276-290.

McGee, V.: "How Truth-Like can a Predicate Be? A Negative Result", *Journal of Philosophical Logic*, vol. 14, pp. 399-410.

Yaqūb, A. M.: *The Liar Speaks the Truth. A Defense of the Revision Theory of Truth*, Nueva York, Oxford University Press, 1993.