Supervaluar o Revisar

Paula Teijeiro

Sociedad Argentina de Analisis Filosofico – IIF
Bulnes 642 Buenos Aires 1176
Argentina
T: 011 4867-2556
paulateijeiro@gmail.com

Article info
CDD: 160.1
Received: 23.03.2016; Revised: 07.06.2016; Accepted: 17.08.2016
DOI: http://dx.doi.org/10.1590/0100-6045.2016.V39N3.PT

Keywords/Palabras Clave:
Vaguedad/Vagueness
Sorites/Sorites
Supervaluacionismo Supervaluationism
Teoría de la Revisión/Revision Theory
Paradojas/Paradoxes

ABSTRACT

Los intentos de solucionar la paradoja de Sorites han generado una multiplicidad muy grande de propuestas. Tantas, que es importante detenerse a reflexionar acerca de qué criterios deberíamos usar para compararlas.

En el presente trabajo voy a, en primer lugar, ofrecer una taxonomía gruesa de las teorías de la vaguedad que clasifica a las propuestas en dos grandes grupos: aquellas que consideran que Sorites se trata de un problema lógico y aquellas que no. Luego, voy a proponer un criterio general de comparación entre las teorías que pertenecen al primer grupo. Por último, voy a aplicar ese criterio a la comparación de dos casos particulares: una versión del Supervaluacionismo y una de la Teoría de la revisión aplicada a términos vagos.

There are already too many attempts to solve the Sorites paradox. So many, that it becomes crucial to stop and reflect on which criteria should we use to compare them.

On the present paper, I will offer, on the first place, a coarse taxonomy of theories of vagueness, that classifies them into two big groups: those which consider Sorites to be a logical problem, and those which do not. Then, I will propose a general criterion for comparing theories of the first type. Lastly, I will show how it applies to two particular cases: one version of Supervaluationism and a Revision Theory for vague terms.

1. Paradojas y soluciones

Supongamos que F es un predicado vago y que a_n es el nombre de un objeto que se encuentra en una serie que va desde objetos que son claramente F (sea a_1) hasta objetos que son claramente no-F (sea a_{100}), en la cual cada objeto difiere sólo marginalmente de aquellos adyacentes en la serie (respecto de las propiedades que determinan la aplicación de F). El siguiente argumento, clásicamente válido, nos permite inferir una contradicción:

(S1)

$$\begin{array}{l} \forall x(Fx_n \rightarrow Fx_{n+1}) \\ Fa_1 \\ \neg Fa_{100} \\ \hline \bot \end{array}$$

La verdad de la premisa mayor (llamada a veces *principio inductivo*) se funda en la presunta naturaleza tolerante de los predicados vagos respecto de cambios pequeños, mientras que las otras premisas pueden fundarse, por ejemplo, en la observación directa (como puede ser para el predicado *Rojo*).

Las soluciones más populares a la paradoja de Sorites suelen basarse en algún tipo de teoría paracompleta (que abandona alguna forma de tercero excluido). Sin embargo, si miramos S1, dichas soluciones no invalidan el argumento, sino que rechazan el principio inductivo. Dado que la revisión de la lógica parece más costosa que el abandono de otras creencias, y el rechazo del principio inductivo es suficiente ¿por qué no permanecer en la lógica clásica?

1.1. Una taxonomía básica

Paradoja no es un concepto puramente lógico, en el sentido de que no hay nada en el argumento como entidad lingüística ni en la lógica que rija el lenguaje en que se formule que permita determinar si un argumento dado es o no una paradoja. Como sucede con el concepto de falacia, para ofrecer una definición se necesitan tanto conceptos lógicos como epistémicos.

Por ejemplo, podemos pensar en una paradoja como un triplo $\langle S, \vDash_C, A\rangle$, donde S es un punto de referencia, $\vDash_C$ es una relación de consecuencia *prima facie* plausible para S (definida sobre una clase de modelos C) y A es un conjunto de argumentos $\vDash_C$-válidos con premisas *prima facie* correctas para S y una conclusión *prima facie* inaceptable para S (relativamente a algún modelo pretendido en C). S podría ser simplemente un sujeto individual, una comunidad, una posición filosófica, etc.

Dada esta aproximación, una taxonomía gruesa de posibles modos de entender a las paradojas y sus soluciones es la siguiente:

Teoría lógica de una paradoja
S entiende a la paradoja $\langle S, \vDash_C, A\rangle$ como un problema lógico si y sólo si $\langle S, \vDash_{C^*}, A\rangle$ no es una paradoja, donde $\vDash_{C^*}$ invalida algún elemento de A, y $\vDash_{C^*}$ es preferible para S a $\vDash_C$.

Teoría no-lógica de una paradoja
S entiende a la paradoja $\langle S, \vDash_C, A\rangle$ como un problema no-lógico si y sólo si $\langle S, \vDash_C, A^*\rangle$ no es una paradoja, donde A^* tiene alguna de las premisas α de A negada, y $\neg\alpha$ es preferible para S a α.

En ambos casos, el nuevo triplo deja de ser una paradoja porque algún argumento es considerado inválido, pero mientras que en el primer grupo esto sucede porque modificamos la lógica, en el segundo sucede porque negamos alguna premisa.

Nótese que al cambiar la relación de consecuencia, puede cambiar la clase de modelos disponibles para interpretar el lenguaje, y en particular el modelo pretendido que interprete a las oraciones en A. Con lo cual también podría suceder en el primer grupo que alguna de las premisas de alguno de los elementos de A ya no sea considerada verdadera. Esto es de hecho lo que sucede con el principio inductivo para K3. Pero es fundamental recordar que si la relación de consecuencia cambia, el objetivo tiene que ser invalidar algún argumento.

Y por ello es que es iluminador pensar en un conjunto A de argumentos y no en uno único. Además de que una paradoja se puede presentar con diversas formulaciones equivalentes, es preciso incluir también las especies de “subparadojas” que se asocian al rechazo de las premisas del argumento principal. Supongamos que $\sim\varphi$ significa que rechazo φ. Si la lógica es clásica, lo único que puede cumplir el rol de ~ es ¬ (es decir, rechazar es afirmar la falsedad). Y por ende, el siguiente argumento es válido:

(S2)
$$\frac{\sim\forall x(Fx_n \rightarrow Fx_{n+1})}{\exists x(Fx_n \wedge \neg Fx_{n+1})}$$

Sin embargo, la conclusión, dado que expresa que existe un corte preciso entre lo *F* y lo no-*F*, no es aceptable. En el marco de K3 tenemos la posibilidad de interpretar el operador ~ mediante la función $\{\langle 1,0\rangle, \langle 0,1\rangle, \langle 1/2, 1\rangle\}$ (donde rechazar sería afirmar la no-verdad) y de ese modo la inferencia ya no resulta válida[1]. S2 debe entonces ser parte de *A* para poder entender por qué una salida de este tipo pertenece al grupo de teorías lógicas de las paradojas.

¿Colapsa la distinción con aquella entre lógica clásica y no clásica? Por un lado, es cierto que un lógico clásico, en términos generales, es aquel que ha ofrecido diagnósticos no lógicos de todas las paradojas, dado que no es habitual que una lógica valide argumentos que resulten inválidos en lógica ásica[2]. Lo que no es cierto es que el lógico no clásico es aquel que ofrece teorías lógicas para las paradojas.

En primer lugar, podría ser que la razón para adoptar la lógica no clásica fuera independiente de cualquier antinomia, como podría ser un *S* que adopta K3 para lidiar con futuros contingentes. En ese caso, si *M* es el Mentiroso, $\langle S, \vDash_{CL}, M\rangle$ no será una paradoja, puesto que $\vDash_{CL}$ no es una relación de consecuencia *prima facie* plausible para *S*. Por otro lado, si M_2 es el Mentiroso Reforzado[3], $\langle S, \vDash_{K3}, M_2\rangle$ sí es una paradoja, pero el lógico en cuestión podría resolverla conservando K3 y ofreciendo una respuesta no lógica.

[1] Por supuesto, la inferencia en donde la conclusión también está formulada con ~ en lugar de ¬ sigue siendo válida, lo cual genera las paradojas de "revancha" de las lógicas trivalentes, que no pueden incorporar la vaguedad de orden superior.

[2] La lógica conexiva es un ejemplo de ello, pero no hay, que yo conozca, paradojas que surjan en una lógica conexiva y se resuelvan adoptando lógica clásica.

[3] La paradoja del Mentiroso Reforzado para K3 es la contradicción que se deriva de asumir que la oración "Yo soy falsa o indeterminada" es verdadera, falsa o indeterminada.

En segundo lugar, hay casos de adopción de lógicas no clásicas por motivos independientes pero vinculados a las paradojas, que aun así no constituyen diagnósticos lógicos de ellas. Por ejemplo, Priest (2003) propone una relación de identidad difusa entre los valores de verdad de una lógica difusa, para explicar por qué la existencia de cortes precisos resulta anti intuitiva, pero no para disolver la paradoja. La paradoja se disuelve rechazando el principio inductivo. De hecho, sugiere que para el caso cualquier lógica difusa podría ser adoptada.

1.2. Cómo decidir entre dos propuestas alternativas

Una primera gran decisión a la hora de abordar una paradoja se ubica entonces en la elección de alguno de los dos caminos que indicamos en la sección anterior. Una vez hecho esto, los *desiderata* de cada uno de estos grupos de teorías son distintos. Una respuesta no lógica debería ofrecer una justificación de la falsedad de los principios involucrados, y quizás una explicación de su inicial y engañoso atractivo. Por ejemplo, teorías contextualistas explican que creemos en la verdad del principio inductivo porque al buscar la instancia falsa, se produce un cambio de contexto, mientras que teorías epistemicistas argumentan que esa instancia falsa es incognoscible.

Por el otro lado, las teorías lógicas –que son aquellas sobre las que vamos a detenernos en este artículo- deberían ofrecernos un sistema de lógica alternativo y mejor que el que dio origen a la paradoja. Lo primero y fundamental es que sea extensionalmente correcto. Es decir, suponemos que hay alguna (no necesariamente única) relación inferencial pre teórica entre las afirmaciones de los lenguajes que usamos, y que el objetivo de la teoría es capturarla.

Esto no quiere decir que aquel que ofrezca un diagnóstico lógico acerca de la paradoja no puede, además, pronunciarse acerca de otras cuestiones, y ofrecer explicaciones epistémicas, psicológicas, sociológicas, o del tipo que sea, para dar cuenta, por ejemplo, de por qué creemos que la lógica correcta es una y no otra, etc. Lo que quiere decir es que la lógica no es más que una teoría de la relación de consecuencia, y como tal, es suficiente con que sea adecuada a los datos (inferencias válidas) que pretendemos capturar.

Un primer problema, claro está, es cómo se determina esa relación de consecuencia pre teórica con la que las lógicas en disputa deberían coincidir. Pero incluso suponiendo que eso está resuelto, resta la cuestión de cómo elegir entre dos sistemas que poseen el mismo conjunto de inferencias, valideces y metainferencias. Y ello depende de cuáles consideremos que son las virtudes metateóricas relevantes. A pesar de haya aquí amplio espacio para debate, un principio que cuenta con suficientes credenciales es el de parsimonia:

(NO) Si dos teorías tienen el mismo conjunto de valideces, inferencias y metainferencias, la más simple debe ser preferida sobre la otra.

No pretendo dar una caracterización exhaustiva de lo que significa "simple" en este contexto, sino que voy a llamar la atención sobre un tipo de simpleza relevante para comparar las teorías que presentaremos a continuación: el "tamaño" de la clase de modelos. Como veremos en la sección final, si bien es complejo definir una noción de tamaño para clases de modelos, hay algún sentido en el que es posible que la cantidad de modelos sobre los que se define la noción de consecuencia varíe sin que ello impacte sobre la validez, con lo cual la teoría nos estaría pidiendo que consideremos modelos superfluos.

2. TRV

La teoría revisionista de la vaguedad es una propuesta de Asmus (2013) que consiste en aplicar una adaptación de la semántica revisionista al tratamiento de predicados vagos. El punto de partida es un lenguaje de base *L* con un correspondiente modelo *M*, formado por un dominio no vacío *D* y una función de interpretación *I* que asigna, en primer lugar, objetos de *D* o funciones sobre *D* a los términos, y en segundo lugar, funciones de $D^n \rightarrow \{0,1\}$ a cada predicado n-ario[4]. Ese lenguaje se enriquece a un lenguaje L^+ mediante la incorporación de los predicados que recibirán un tratamiento

[4] Asumimos que los modelos son sustitucionales, para evitar complejidades innecesarias para el problema que nos ocupa.

revisionista. En el caso de Gupta y Belnap (1993), los predicados se introducen mediante definiciones en la metateoría de la forma[5]:

$$F(x) =_{\mathrm{def}} \varphi(x)$$

El modelo de base no nos dice nada respecto de cómo interpretar el nuevo vocabulario. Para hacerlo, en lugar de simplemente extender la función de interpretación (lo cual puede no ser posible para ciertos predicados circularmente definidos, como el predicado de verdad), es preciso considerar secuencias de interpretaciones.

DEFINICIÓN (*Hipótesis)*
Una hipótesis σ_{n} es una interpretación del nuevo vocabulario de L^{+}.

Los modelos extendidos mediante hipótesis nos permiten definir una función completa de valuación $\mathrm{Val}_{\mathrm{M}\sigma\mathrm{n}}$ del modo habitual. Pero a diferencia de lo que sucede con *I*, la interpretación de los nuevos predicados que ofrece σ_{n} no necesariamente es definitiva. Las definiciones de los predicados limitan sus posibles interpretaciones, con lo cual una vez planteada una hipótesis, ellas funcionan como reglas que nos obligan a ajustar la interpretación. Esto es, proponemos una hipótesis y la regla de revisión "mira" la definición del predicado y nos ofrece una nueva hipótesis que esté de acuerdo con ella.

DEFINICIÓN *(Regla de revisión)*
Una regla de revisión ϱ es una operación de hipótesis en hipótesis tal que $\varrho(\sigma_{\mathrm{n+1}})(F)(\mathrm{d})=\mathrm{Val}_{\mathrm{M}\sigma\mathrm{n}}(F(d))$[6]

DEFINICIÓN *(Secuencia de revisión)*[7]

[5] Para simplificar la notación, tomaremos como ejemplo lenguajes que sean extendidos sólo mediante predicados monádicos.

[6] *d* es el término cuya interpretación es el objeto d.

[7] En su versión original, las secuencias de revisión tienen longitud ON.

Una secuencia de revisión σ es una secuencia de longitud ω donde cada elemento σ_{n+1} es el resultado de aplicar la regla ρ a la hipótesis σ_n.

Las estructuras que Asmus propone para evaluar las oraciones de L^+ carecen de reglas de revisión y definiciones, y en su lugar poseen conjuntos de "irrevisables" y "equivalencias"[8]:

DEFINICIÓN (*R-Modelo* $\langle M, E, IR, S \rangle$)
Un R-modelo es una estructura $R = \langle M, E, IR, S \rangle$ tal que

(i) M es el modelo de base.
(ii) E es el conjunto de equivalencias en el que hay, para cada nuevo predicado F, un conjunto E_F de pares ordenados de objetos del dominio.
(iii) IR es un conjunto de irrevisables en el que hay, para cada nuevo predicado F, un par $\langle A_{IRF}, B_{IRF} \rangle$ que son, respectivamente, los objetos irrevisablemente F y los irrevisablemente no-F.
(iv) S es un conjunto de secuencias de revisión.

Intuitivamente, los casos paradigmáticos de F son los miembros de A_{IRF}, los casos paradigmáticos de no-F son los miembros de B_{IRF}, y el nivel de tolerancia del predicado está representado abstractamente por su conjunto de equivalentes E_F. Vale la pena aclarar que los conjuntos de equivalentes no son clases de equivalencia, sino simplemente pares ordenados que rescatan las relaciones de cercanía entre los objetos respecto de F. Esto quiere decir que no se exige que la relación E_F sea, por ejemplo, transitiva, sino sólo que:

$$\langle a,b \rangle \in E_F \Rightarrow (a \in A_{IRF} \Leftrightarrow b \notin B_{IRF}) \& (a \in B_{IRF} \Leftrightarrow b \notin A_{IRF})$$

Dada una determinada extensión hipotética, el hecho de que los predicados son tolerantes nos obliga siempre a ajustarla, pero por el mismo motivo no estamos obligados a ajustarla en ninguna dirección en particular.

[8] La presentación de TRV que se encuentra a continuación ha sido ligeramente modificada a efectos expositivos.

Es por ello que Asmus habla de "restricciones de revisión" en lugar de reglas.

(i) $a \in A_{IRF} \Rightarrow \sigma_n(F)(a)=1$
(ii) $a \in B_{IRF} \Rightarrow \sigma_n(F)(a)=0$
(iii) $\langle a,b \rangle \in E_F \Rightarrow (\sigma_n(F)(a)=1 \Rightarrow$ o bien $\sigma_{n+1}(F)(b)=1$ o bien $\sigma_{n+1}(F)(a)=0)$

Las dos primeras restricciones indican que los irrevisables deben mantenerse fijos en la extensión y la antiextensión del predicado –según corresponda- en toda etapa de revisión. La tercera restricción dice que si dos objetos son equivalentes respecto de un predicado vago, entonces si en una etapa de revisión uno de ellos se encuentra en la extensión, en la siguiente debe o bien entrar el otro o bien salir el primero.

Como puede verse, la tercera restricción habilita que las secuencias puedan comportarse de un modo muy distinto, aunque compartan los conjuntos de irrevisables y equivalencias. Una manera posible de darle una interpretación filosófica a la semántica es pensar la secuencia como un proceso de ajuste que realiza un sujeto cuando se enfrenta a una serie sorítica: llega hasta determinado punto, cambia su juicio de positivo a negativo, lo cual genera que retracte su juicio respecto de algunos de los objetos anteriores en la serie, y así sucesivamente. Podría suceder que eventualmente se llegue a un *loop*, pero ello no es necesario, el proceso no tiene que respetar ningún patrón en particular.

Distintas secuencias podrían representar el comportamiento de distintos sujetos. Las oraciones que en algún punto un sujeto decide que son verdaderas se denominan "verdades estables":

DEFINICIÓN *(Verdad estable)* $R, \sigma \vDash^{VE} \varphi$
φ es *establemente verdadera* en un modelo R de acuerdo con σ si y sólo si hay un n tal que $Val_{M\sigma m}(\varphi) = 1$ en todo σ_m para $m>n$.

Las verdades categóricas, por su parte, son aquellas que todos los sujetos deciden que son verdaderas:

DEFINICIÓN *(Verdad Categórica)* $R \vDash^{VC} \varphi$
φ es *categóricamente verdadera* en R si y sólo si en toda secuencia σ, φ es establemente verdadera en σ.

De manera análoga, los casos limítrofes para un sujeto y para todos ellos se definen respectivamente como:

DEFINICIÓN *(Oración Patológica)*
φ es *patológica* en un modelo R si y sólo si hay alguna secuencia σ en la que no es estable.

DEFINICIÓN *(Oración Paradójica)*
φ es *paradójica* en un modelo R si y sólo si en toda secuencia σ, φ es inestable.

La relación de consecuencia es preservación de verdad categórica:

DEFINICIÓN *(Consecuencia lógica)* $\Gamma \vDash_{TRV} \varphi$
Para todo modelo R, si hay un número n a partir del cual todas las premisas son establemente verdaderas en toda secuencia, la conclusión es categóricamente verdadera.

Por otro lado, Asmus menciona, sin definir estrictamente, este otro sentido de verdad:

DEFINICIÓN *(Verdad irrevisable)* $R \vDash^{VI} \varphi$
Verdad determinada por los irrevisables.

Parece haber dos formas posibles de interpretar esto. Una es considerar que se está refiriendo al conjunto de fórmulas atómicas Fa tales que $a \in A_{IRF}$. Esta interpretación parece innecesariamente restringida. Otra posibilidad sería cerrar ese conjunto bajo consecuencia lógica. Correspondientemente, la falsedad irrevisable sería el conjunto de oraciones Fa tales que $a \in B_{IRF}$, cerrado a su vez bajo consecuencia inversa (todas las oraciones que las implican).

Finalmente, en lo que respecta a Sorites, la respuesta que da la TRV a la paradoja es que hay una instancia de la premisa universal que no es ni (categóricamente) verdadera ni (categóricamente) falsa, puesto que su antecedente y consecuente oscilan en al menos una de las secuencias. La oración existencial que afirma que existe un corte preciso es (categóricamente) verdadera, ya que en toda etapa de toda secuencia existe algún punto de corte, aunque ese punto vaya variando a lo largo de la revisión. Y por lo tanto, el principio de inducción es (categóricamente) falso.

3. SPV

El supervaluacionismo puede presentarse de muchas maneras distintas. La idea general consiste en entender las expresiones vagas como teniendo un significado intensional, compuesto por una extensión actual imprecisa (con casos indeterminados) y muchas extensiones "potenciales" precisas. Una oración sería entonces verdadera siempre y cuando resulte verdadera sin importar el modo en que precisemos el significado de los términos.

De modo más formal:

DEFINICIÓN *(Modelo Parcial)*
Un modelo parcial M es una estructura ⟨D,I⟩, donde D es un conjunto no vacío e I es una función que asigna a cada predicado n-ario funciones de $D^n \rightarrow \{1, \frac{1}{2}, 0\}$, y a los términos, objetos de D o funciones sobre D.

Dados estos modelos, verdad en ellos puede definirse del modo habitual, mediante una función de valuación Val_M que respete algún esquema trivalente como Kleene fuerte o débil. La elección del esquema a utilizar es en principio irrelevante –mientras que sea normal- puesto que la noción de consecuencia va a estar definida sobre los modelos clásicos. Que una oración recibe valor 1 en un modelo lo escribiremos $M \vDash^V \varphi$

Los distintos modelos pueden ordenarse parcialmente según la relación de extensión:

DEFINICIÓN *(Extensión)*[9]
Un modelo $M = \langle D,I \rangle$ extiende a otro $M' = \langle D',I' \rangle$ si y sólo si D=D', $I(a_i)=I'(a_i)$ y $I(F)(a_1...a_n) \leq I'(F)(a_1...a_n)$ para todo F y todo a_i

Luego, consideramos las estructuras que pueden armarse con los modelos parciales:

DEFINICIÓN *(Espacio de precisificaciones)*
Un espacio de precisificaciones es una estructura $P = \langle M_B, P, \subseteq \rangle$ tal que $\subseteq$ es la relación de extensión entre modelos definida en P×P, P es un conjunto de modelos parciales, con un elemento mínimo M_B (respecto del orden parcial inducido por $\subseteq$), y al menos un elemento M_C que sea un modelo clásico[10] y tal que $M_B \subseteq M_C$.

Finalmente, dada la estructura, podemos definir el concepto de superverdad:

DEFINICIÓN *(Superverdad en M)* $M \vDash^{SV} \varphi$
φ es superverdadera en M si y sólo si para todo M_C tal que $M \subseteq M_C$, $M_C \vDash \varphi$

DEFINICIÓN *(Superverdad en P)* $P \vDash^{SV} \varphi$
φ es superverdadera en P si y sólo si es superverdadera en M_B.

Este concepto de verdad no composicional permite que las verdades lógicas sean superverdaderas, aunque sus componentes no lo sean. Además, podemos imponer ciertas restricciones sobre P, de modo tal que no toda extensión de M_B sea una precisificación admisible. Esto nos permite rescatar

[9] El orden de los valores de verdad al que hace referencia "≤" es el orden parcial según el cual ½ es menor que ambos valores clásicos, y estos son incomparables entre sí.

[10] No es necesario que el espacio de precisificaciones contenga toda precisificación posible (y de hecho, podría no ser deseable), pero sí es necesario que al menos alguno de los modelos sea clásico, puesto que de lo contrario, todas las oraciones serían (trivialmente) superverdaderas en el modelo de base, haciendo que $\varphi \wedge \neg \varphi$ tenga modelo.

principios generales verdaderos (aunque no necesariamente válidos) que involucren predicados vagos. Por ejemplo, suponiendo que tanto "hombre" como "animal" tengan casos limítrofes, podemos aun así afirmar que "Todos los hombres son animales" si restringimos las precisificaciones admisibles a aquellas en las cuales la interpretación de "hombres" está incluida en la de "animales". En la literatura, esto se conoce como "conexiones penumbrales".

Dados estos dos conceptos de verdad, vale la pena preguntarse cuál es la "verdadera verdad", de modo de poder elegirla como valor designado. Existen muchas definiciones alternativas de consecuencia lógica (ver Varzi (2007)), pero vamos a considerar sólo las dos más habituales:

DEFINICIÓN *(Consecuencia lógica global)* $\Phi \vDash_{SVG} \Psi$
$\Phi \vDash_{SVG} \Psi$ si y sólo si para todo espacio P, si toda φ es superverdadera en M_B, alguna ψ es superverdadera en M_B.

DEFINICIÓN *(Consecuencia lógica local)* $\Phi \vDash_{SVL} \Psi$
$\Phi \vDash_{SVL} \Psi$ si y sólo si para todo espacio P y para toda modelo clásico en P, si toda φ es verdadera en M_C, alguna ψ es verdadera en M_C.

4. TRV vs SPV

El supervaluacionismo conserva las mismas valideces e inferencias que la lógica clásica, lo cual podría hacernos pensar que quien adopta esta técnica para interpretar al vocabulario vago, no considerará que la paradoja de Sorites represente un problema lógico. Sin embargo, si adoptamos la definición global de consecuencia, las siguientes inferencias con múltiples conclusiones resultan inválidas:

- $\forall x\varphi \vDash \forall x\psi, \neg\forall x(\varphi \to \psi)$
- $\varphi \vDash \psi, \neg(\varphi \to \psi)$
- $\varphi \vee \neg\varphi \vDash \varphi, \neg\varphi$

¿Deberíamos tener esto en cuenta a la hora de pensar qué tipo de teoría de Sorites es SPV, o se trata de una divergencia sin importancia? La respuesta es que esta característica de SPV -a veces señalada como una falencia- es crucial para la solución de la paradoja. La diferencia entre una solución clásica y una supervaluacionista radica en que, si bien en ambas (S1) resulta ser un razonamiento incorrecto con una premisa universal que debe ser rechazada, sólo en SPV la siguiente subparadoja resulta inválida:

(S3)
$$\frac{\exists x(Fx_n \wedge \neg Fx_{n+1})}{Fa_1 \wedge \neg Fx_2,\ Fa_2 \wedge \neg Fx_3,\ \ldots,\ Fa_{99} \wedge \neg Fx_{100}}$$

De ese modo, el rechazo de la premisa universal resulta "menos grave" que para el lógico clásico, puesto que no me obliga a aceptar ninguna instancia. Por ende, no es la negación del principio de inducción lo que desarticula la paradoja, sino la invalidez de al menos uno de los argumentos que la componen.

Por su parte, TRV también tiene el mismo conjunto de valideces e inferencias con conclusiones únicas que la lógica clásica, pero, al igual que SPV, invalida las inferencias con conclusiones múltiples mencionadas. Esto significa que, si uno propone la adopción de alguno de estos sistemas *como salida a la paradoja de Sorites*, es porque considera que estamos frente a un problema lógico.

Vamos entonces, en primer lugar, a mostrar que las nociones de consecuencia entre ambas coinciden también para el caso general de conclusiones múltiples, para luego, en la sección final, argumentar por qué el supervaluacionismo resulta preferible[11].

TEOREMA. $\Gamma \vDash_{TRV} \Phi \Rightarrow \Gamma \vDash_{SVG} \Phi$

Dado un contramodelo supervaluacionista P para $\Gamma \vDash \Phi$, construimos un modelo revisionista R tomando como A_{IRF} a $\{a : I(F)(a)=1 \text{ en } M_B\}$ y como B_{IRF} a $\{a : I(F)(a)=0 \text{ en } M_B\}$. El conjunto de equivalencias se deja vacío y el

[11] Una comparación exhaustiva requeriría analizar también las metainferencias que ambas validan, pero no hay, hasta donde llega mi conocimiento, un cálculo de secuentes para SPV (ni para TRV), así que nos restringiremos al resultado parcial.

conjunto de secuencias puede formarse tomando, para cada modelo clásico M_C del espacio, una secuencia constante $\langle M_C, M_C, M_C,\ldots\rangle$.

Dado que toda $\gamma \in \Gamma$ es superverdadera en *P*, es verdadera en todo M_C, por ende establemente verdadera en toda σ de *R* y por lo tanto, categóricamente verdadera en *R*. Por otra parte, ninguna $\varphi \in \Phi$ es superverdadera en *P*, con lo cual cada φ es falsa en algún M_C, con lo cual cada φ es establemente falsa en alguna σ, y por lo tanto ninguna es categóricamente verdadera. Esto establece que *R* es un contramodelo de $\Gamma \vDash \Phi$. □

Las características particulares del contramodelo obtenido no son relevantes, mientras se trate efectivamente de un contramodelo (por ejemplo, no importa el hecho de que en *R* no hay oraciones inestables). Por el contrario, para la dirección inversa del condicional –la que resta para probar que ambas nociones de consecuencia coinciden- vamos a construir un modelo con ciertos requisitos adicionales, de modo que también nos permita dar una respuesta al siguiente desafío de Asmus:

> Supertruth cannot coincide with both categorical truth and unrevisable truth (truth determined by unrevisables), as categorical truth and unrevisable truth differ. (...) Supertruth is like unrevisable truth: if a formula is unrevisably true, then it is true on all the hypotheses in a sequence. Supertruth is like categorical truth: they are designated values. (…) Straightforward mappings from the revision approach to supervaluationism will collapse these distinctions. Any attempt to show that the revision approach is disguised supervaluationism will have to find a subtle way of dealing with this.

Dado un modelo revisionista R, construimos su correspondiente espacio de precisificaciones $P = \langle \mathrm{P}, M_B, \subseteq\rangle$ tomando, en primer lugar, como base al modelo M_B tal que $\mathrm{I}(F)(\mathrm{a})=1$ para todo a en $\mathrm{A}_{\mathrm{IRF}}$, $\mathrm{I}(F)(\mathrm{a})=0$ para todo a en $\mathrm{B}_{\mathrm{IRF}}$ y $\mathrm{I}(F)(\mathrm{a})= ½$ para el resto de los elementos del dominio. En principio, P será el conjunto de todas sus extensiones.

Como podrá intuirse por el modo en que seleccionamos el modelo de base, el concepto de verdad irrevisable puede definirse en función de él:

DEFINICIÓN *(Verdad Irrevisable en SPV)* $P \vDash^{VI} \varphi$
Las oraciones irrevisablemente verdaderas (falsas) son las oraciones atómicas verdaderas (falsas) en M_B, cerradas bajo consecuencia (inversa).

Esta definición corresponde a la interpretación "amplia" del concepto. Si elegimos la interpretación "estrecha", deberemos reemplazarla por verdad en M_B, restringido a las oraciones atómicas[12]. En segundo lugar, dado que verdad categórica es el valor designado en TRV, su contraparte supervaluacionista es sencilla de identificar:

DEFINICIÓN *(Verdad categórica en SPV)* $M_B \vDash^{SV} \varphi$
Una oración es categóricamente verdadera (falsa) si y sólo si es Superverdadera (Superfalsa) en M_B.

Dado el espacio que tenemos hasta el momento, verdad categórica y verdad irrevisable (amplia) de hecho colapsan: lo único que puede ser superverdadero sin ser verdadero en M_B son las verdades lógicas, y ellas resultan irrevisables por ser consecuencia del conjunto vacío.

Si queremos incorporar nuevas verdades como categóricas, será necesario que restrinjamos las precisificaciones admisibles. Como dijimos antes, esto suele hacerse para hacer superverdaderas a aquellas oraciones que involucran a las conexiones penumbrales, esto es, para que resulten superverdaderas oraciones complejas y contingentes que no son verdaderas en el modelo de base. En términos de TRV, lo que queremos es incorporar las oraciones categóricas que son tales, no por ser irrevisables, sino porque el proceso de revisión de ellas se estabiliza en algún punto[13].

---

[12] Hay también otro concepto intermedio -verdad en M_B- que sí depende de la elección del sistema de valuación.

[13] Cabe aclarar que estas oraciones categóricas no irrevisables podrían ser atómicas o complejas. Nuevas oraciones atómicas superverdaderas no suele ser lo que se busca cuando se restringen las precisificaciones admisibles, dado que suele considerarse al supervaluacionismo como una técnica no composicional para

El concepto de verdad estable, al igual que el de Superverdad, no es veritativo funcional, y por eso, podría no haber ningún modelo en P que haga verdaderas (falsas) a todas y sólo las oraciones establemente verdaderas (falsas) de una secuencia. Lo que sí encontraremos es lo siguiente:

DEFINICIÓN *(Modelo H)*
M_H es un modelo H si y sólo si existe un σ en S tal que para toda oración φ, $M_H \vDash^{SV} \varphi$ si y sólo si φ es establemente verdadera en σ y $M_H \vDash^{SF} \varphi$ si y sólo si φ es establemente falsa en σ.

Sabemos que los M_H existen, dado que P está formado por todas las extensiones de M_B y las oraciones establemente verdaderas en una secuencia son una extensión consistente (aunque no necesariamente completa) de las irrevisables.

DEFINICIÓN *(Verdad estable en SPV)* $M_H \vDash^{SV} \varphi$
Una oración es establemente verdadera (falsa) en un M_H si y sólo si es Superverdadera (Superfalsa) en M_H.

Dado el espacio P, construimos P' quitando de *P* todos los modelos que no sean M_B, un M_H o un M_C que extienda a algún M_H. Naturalmente, por el modo en que construimos P', sucederá que los conceptos definidos coinciden:

$$R \vDash^{VI} \varphi \Leftrightarrow P' \vDash^{VI} \varphi$$
$$R,\sigma \vDash^{VE} \varphi \Leftrightarrow M_H \vDash^{SV} \varphi$$
$$R \vDash^{VC} \varphi \Leftrightarrow M_B \vDash^{SV} \varphi$$

Y lo mismo para los conceptos duales de falsedad. Esto muestra que la afirmación de Asmus de que el revisionismo no es supervaluacionismo disfrazado es, en este sentido, falsa. Hay otro sentido en que ambas propuestas se diferencian, y puede ya notarse si vemos que diversas

calcular el valor de oraciones moleculares, pero consideramos este caso para establecer la equivalencia con el modelo de TRV.

secuencias pueden corresponder con el mismo modelo H, dado que, si bien podrían estabilizar del mismo modo a las mismas oraciones, esto podría suceder en distintas etapas en cada una. Volveremos sobre esto en la última sección.

Por último, podemos dar por establecido el segundo teorema

TEOREMA. $\Gamma \vDash_{SVG} \Phi \Rightarrow \Gamma \vDash_{TRV} \Phi$

Dado un contramodelo revisionista para $\Gamma \vDash \Phi$, construimos un espacio *P*' como indicamos antes. Dadas las equivalencias mencionadas más arriba, *P'* también será un contramodelo para $\Gamma \vDash \Phi$. □

Se podría objetar que en los modelos de TRV, los conceptos de verdad estable, irrevisable y categórica surgen naturalmente, mientras que sus contrapartes supervaluacionistas requieren que restrinjamos *ad hoc* los espacios de precisificaciones. Sin embargo, la situación en TRV es exactamente la misma, puesto que debemos restringir metateóricamente las secuencias en S si queremos que verdad estable e irrevisable no coincidan.

Por último, hay otra distinción que -podría argumentarse- no puede ser representada por SPV, que es la que existe en TRV entre los dos tipos de oraciones problemáticas: las meramente patológicas y las paradójicas. Pero dado que en la estructura supervaluacionista hemos representado a los hablantes por los modelos H, podemos fácilmente caracterizar ambas categorías del modo esperable:

DEFINICIÓN *(Patológico en SPV)*

Una oración es patológica en SPV si y sólo si es no es superverdadera ni superfalsa en algún modelo H.

DEFINICIÓN *(Paradójico en SPV)*

Una oración es paradójica en SPV si y sólo si es no es superverdadera ni superfalsa en ningún modelo H.

5. Parsimonia

Mostramos hasta ahora cómo los conceptos de consecuencia de ambas teorías coinciden y no son clásicos, con lo cual, en tanto teorías lógicas de la vaguedad, la más simple debe ser preferible por sobre la otra.

Hay un sentido intuitivo en el que es claro que TRV tiene más modelos que SPV, aunque darle forma precisa a esta idea no es sencillo. Por un lado, siempre tenemos tantos modelos como conjuntos, sólo por el hecho de que cada conjunto puede constituir el dominio de un modelo. De este modo, los modelos de ambas teorías conforman clases propias y como los números cardinales son conjuntos, esas clases no tendrán cardinalidad.

A veces se usa -en discusiones acerca de lógicas no clásicas- la relación de inclusión como parámetro de tamaño. Por ejemplo, K3 tiene "más modelos" que CL, puesto que todo modelo clásico es un modelo trivaluado, pero no a la inversa. Pero $\subseteq$ es defectuosa como medida de tamaño, dado que es una relación parcial, y en nuestro caso particular de hecho sucede que ninguna de las clases de modelos incluye a la otra. Sin embargo, podemos considerar en algún sentido los modelos de SPV como la subclase de modelos de TRV con secuencias de revisión constantes y equivalencias vacías, y por ende, una subclase propia, dado que en TRV hay modelos con secuencias no constantes y equivalencias no vacías.[14]

Uno podría argumentar que los modelos que le faltan a SPV son relevantes, puesto que son aquellos que representan correctamente la semántica del lenguaje natural –en particular, la inestabilidad de ciertas oraciones. La situación sería análoga a una teoría de modelos trivalente en la cual ½ se comportara exactamente igual que 0. Podría ser que las oraciones falsas sean clasificables en dos grupos en función de alguna propiedad, pero

[14] Estrictamente, estamos hablando de una clase de modelos SPV* que es igual a SPV, pero que en lugar de modelos clásicos tiene ω-secuencias constantes de modelos clásicos. Como SPV* es una subclase propia de TRV, y es equivalente a ella, debería ser preferible. Por último, también por consideraciones de parsimonia, podemos argumentar que SPV es preferible a SPV*: si bien la "cantidad" de modelos es la misma, los espacios de SPV* son estructuralmente más complejos que los de SPV, de modo innecesario.

desde el punto de vista lógico, esa propiedad es irrelevante, puesto que esos modelos "extra" no están cumpliendo ningún rol.

Del mismo modo, los procesos de revisión podrían ciertamente estar involucrados en nuestras prácticas lingüísticas, no hay motivo para pensar que deban ser representados en los modelos como parte del aparato lógico. Hay muchos aspectos semánticamente interesantes de nuestro lenguaje que no son representados en la teoría de modelos. Por supuesto, el punto contra TRV no es que las secuencias de revisión son irrelevantes: ello sería una petición de principio. Por el contrario, el punto es que son irrelevantes para la determinación de la relación misma de consecuencia que Asmus parece considerar correcta.

TRV es –gracias a esa cantidad de modelos- sin dudas una teoría expresivamente mucho más poderosa, y si bien mostramos que SPV puede caracterizar sin problemas las distinciones semánticas que consideramos, esto no quiere decir que toda clasificación de oraciones que realice TRV pueda ser capturada por SPV. Si pensamos en la versión original de la teoría de la revisión, nos encontramos con que dentro de las oraciones estables, las hay estables a secas y cercanamente estables, que son aquellas que se estabilizan antes de un ordinal límite. El motivo por el cual Gupta y Belnap distinguen ambos tipos de estabilidad es que ello redunda en dos relaciones de consecuencia distintas. Sin embargo, la mera multiplicación de categorías semánticas no es en sí misma un valor. Al menos, no desde un punto de vista lógico.

Referencias

ASMUS, C. M. "Vagueness and revision sequences". Synthese, Vol. 190, No. 6, pp 953-974, 2013.

BEALL, JC. *Liars and heaps*. Oxford: Clarendon Press, 2003.

COBREROS, P. "Supervaluationism and Classical Logic". En: R. Nouwen, R., R. van Rooij et al (eds.) (2011) pp 51-63.

FINE, K. "Vagueness, Truth and Logic". *Synthese*, Vol. 30, No. 3/4, On the Logic Semantics of Vagueness, pp 265-300, 1975.

GUPTA, A. & BELNAP, N., *The revision theory of truth*. Cambridge: The MIT Press, 1993.

KEEFE, R. *Theories of vagueness*. Cambridge: Cambridge University Press, 2000.

NOUWEN, R., VAN ROOIJ, R., et al. *Vagueness in Communication*. Berlin: Springer, 2011.

SMITH, N. *Vagueness and degrees of truth*. Nueva York: Oxford University press, 2008.

PRIEST, G. "A site for sorites". En JC Beall, (Ed.) (2003), pp 9-23.

VARZI, A. "Supervaluationism and its logics". Mind, Vol. 116, No. 463, pp 663-676, 2007.