¿Era realmente Sócrates tan sabio?

Emilio F. Gómez-Caminero Parejo
(Grupo de Lógica, Lenguaje e información, Universidad de Sevilla)
egomezcaminero@us.es

1. Generalidades

Como es bien sabido, la lógica epistémica trata sobre aquellos razonamientos que los agentes de un cierto grupo hacen o pueden hacer sobre lo que otros agentes, o ellos mismos, saben o ignoran[1]. La forma habitual de introducir el concepto de conocimiento es tratar esta lógica como un tipo de lógica multimodal, de forma que para cada agente a_i del conjunto $\mathcal{A}$ agentes añadiremos el operador K_{a_i}; con el significado "a_i sabe que …". Es frecuente también introducir como signo derivado el dual del operador de conocimiento, que escribiremos $\widehat{K}_{a_i}$ y que definimos:

$$\widehat{K}_{a_i}\varphi =_{def} \neg K_{a_i}\neg\varphi.$$

El lenguaje $\mathcal{L}^{LEP}$ de la lógica epistémica proposicional queda pues definido de la siguiente forma:

$$\varphi ::= p|\neg\varphi|\varphi \wedge \psi|K_{a_i}$$

Por supuesto, resulta fácil extender la sintaxis para obtener una lógica epistémica de primer orden, pero eso es algo de lo que hablaremos un poco más adelante.

Es también sabido que, igual que ocurre con otras lógicas modales, podemos disponer de un cierto número de sistemas de lógica epistémica, que podemos caracterizar tanto axiomáticamente como desde un punto de vista semántico. Los más conocidos de estos sistemas axiomáticos para m agentes se denominan T_m, $S4_m$ y $S5_m$. T_m se caracteriza por los siguientes axiomas y reglas:

A1: Todas las tautologías de la lógica proposicional

[1] La obra fundacional sobre lógica epistémica es Hintikka, 1962. Para el lector que quiera profundizar en este tema, dos obras clásicas de referencia son Fagin, 1995, y Meyer. J.J. y van der Hoek, 1995.

A2: $K_{a_i}\varphi \wedge K_{a_i}(\varphi \rightarrow \psi) \rightarrow K_{a_i}\psi$

A3: $K_{a_i}\varphi \rightarrow \varphi$

R1: $\dfrac{\varphi \rightarrow \psi \quad \varphi}{\psi}$

R2: $\dfrac{\varphi}{K_{a_i}\varphi}$

Si en lugar de T_m queremos trabajar con $S4_m$, deberemos añadir el llamado *axioma de introspección positiva*:

A4: $K_{a_i}\varphi \rightarrow K_{a_i}K_{a_i}\varphi$

Por último, si deseamos trabajar con el sistema $S5_m$, deberemos añadir el que se conoce como *axioma de introspección negativa*:

A5: $\neg K_{a_i}\varphi \rightarrow K_{a_i}\neg K_{a_i}\varphi$.

El axioma A3 es el que caracteriza al conocimiento (por oposición a la creencia, por ejemplo), y no resulta demasiado problemático. El axioma A2 ha dado lugar a una interesante y fructífera discusión sobre el llamado *problema de la omnisciencia lógica*[2]; y otro tanto sucede con la regla R2. Ésta es sin duda una cuestión apasionante, pero no es la que queremos discutir aquí. Nuestra discusión se centrará en el algo más modesto A5. Un axioma tenido por contraintuitivo pero que ha terminado por imponerse en la literatura. Nuestra pretensión es discutir por qué resulta contraintuitivo; y la conclusión a la que llegaremos es que su aspecto poco natural se debe a cuestiones relacionadas con la lógica epistémica de primer orden, y no con el nivel proposicional. Estas cuestiones están, además, relacionadas con dos viejas conocidas de la lógica modal: la Fórmula Barcan y su conversa.

Pero antes de abordar abiertamente estas cuestiones es conveniente recordar algunas cuestiones elementales sobre la semántica de la lógica epistémica. Como es bien sabido, el tratamiento habitual es la semántica kripkeana de mundos posibles, en la que, dados un conjunto no vacío P de variables proposicionales y un conjunto $\mathcal{A}$ de m agentes, definimos un modelo M como una estructura $M = \langle W, R_{a_1}, \cdots, R_{a_m}, v \rangle$, tal que $W \neq \emptyset$ es un conjunto de índices o mundos posibles $\{s, t, u, \cdots\}$; $v: P \times W \longmapsto \{0,1\}$ es una función de evaluación que asigna a cada variable proposicional un valor de verdad en cada mundo posible s (lo que escribiremos: $v(s,p) = 1$ ó $v(s,p) = 0$); y por último, $R_{a_i} \subseteq W^2$ (para $a_i \in \mathcal{A}$) es una relación binaria entre mundos posibles que representa la noción intuitiva de accesibilidad desde un mundo posible a otro para un sujeto dado a_i.

[2] Tal vez el artículo más conocido en torno a estas cuestiones sea Hocutt, 1972.

Una vez establecido el concepto de modelo kripkeano, es fácil definir la noción de verdad de una fórmula φ en un mundo posible s de un modelo dado M (en símbolos, $M, s \vDash \varphi$) de forma que respetemos nuestras intuiciones básicas sobre los operadores epistémicos. Los operadores proposicionales, por supuesto, se definen de la forma habitual. Para el caso del operador K, el concepto de verdad de una fórmula en un mundo posible se define:

$$M, s \vDash K_{a_i}\varphi \text{ si y sólo si (en adelante, syss) } M, t \vDash \varphi \text{ para todo } t \text{ tal que } sR_{a_i}t.$$

En cuanto a su dual, resulta ya bastante simple de definir:

$$M, s \vDash \widehat{K}_{a_i}\varphi \text{ syss } M, t \vDash \varphi \text{ para algún } t \text{ tal que } sR_{a_i}t.$$

Es fácil demostrar que las propiedades de la relación de accesibilidad son las que determinan qué fórmulas son válidas en un sistema dado; de modo que los axiomas A1 y A2, así como las reglas R1 y R2, son válidos en todos los modelos kripkeanos; el axioma A3 es válido en todos los modelos kripkeanos en los que la relación de accesibilidad es reflexiva (abreviadamente, en todos los modelos reflexivos); A4 es válido en todos los modelos reflexivos y transitivos y A5 es válido en todos aquellos modelos en los que la relación de accesibilidad es reflexiva, simétrica y transitiva.

2. Introspección negativa

Tendremos que volver más adelante a la semántica para extenderla a la lógica de primer orden; pero es mejor no adelantarse a los acontecimientos, así que dejemos la semántica de la lógica epistémica de primer orden para cuando la necesitemos y volvamos al tema principal: el axioma de introspección negativa.

La cuestión es por qué nos resulta contraintuitivo el axioma de introspección negativa. Con la introspección positiva no parece haber tantos problemas: podemos admitirla constatando, con Hintikka, que a su vez apela a toda la tradición filosófica, que saber que uno sabe algo equivale a saber algo *simpliciter*.

Con la introspección negativa, en cambio, nuestras intuiciones no resultan tan significativas. Hintikka, en este caso, contesta recurriendo a la más técnica noción de *implicación virtual*:

Nótese, sin embargo, que uno puede dejar de saber a menos que resulte ser tan sagaz como Sócrates aquello que ignora. Porque "$\neg K_a p$" no implica (virtualmente) "$K_a \neg K_a p$" como puede comprobarse fácilmente.[3]

Pero, ¿qué significa la noción de *implicación virtual*? Hablando algo toscamente, no consiste más que en la imposibilidad de construir un modelo que haga verdadera a una de ellas y falsa a la otra. Así definida, es cierto que $K_a\varphi$ implica virtualmente $K_aK_a\varphi$, mientras que $\neg K_a\varphi$ no implica virtualmente $K_a\neg K_a\varphi$, pero sólo porque Hintikka ha decido aceptar que la relación de accesibilidad es transitiva, pero no simétrica.

¿Y cuál es la razón por la que considera que la relación de accesibilidad no es simétrica? La respuesta, en este caso, resulta algo más oscura:

Puede verse que la relación no es simétrica. Para ello, recordemos que un conjunto modelo μ_2 es una alternativa a μ_1 si, y sólo si, intuitivamente hablando, no hay nada en el estado de hechos descrito por el primero que sea incompatible con lo que alguien sabe en el estado de hechos descrito por el segundo. Ahora bien, no está obviamente excluido por lo que yo sé ahora que debiera saber más que lo que ahora sé. Sin embargo tal conocimiento adicional puede muy bien ser incompatible con lo que, en la medida de mis conocimientos, ahora es todavía posible.[4]

Vemos que la argumentación dista mucho de ser clara. Además, términos como "ahora" o "todavía" tienden a introducir una interpretación temporal que no hace sino añadir complejidad a la cuestión. Tal vez el análisis de un par de ejemplos contribuya a aclarar nuestras intuiciones.

Empecemos con un ejemplo a favor del axioma de introspección negativa. Supongamos cuatro jugadores, digamos Ana, Berta, Carlos y David, jugando una partida de cartas. Supondremos, como es habitual, que los agentes son *razonadores perfectos*; pero también, y esto es importante, que se trata de una baraja estándar sin trucar —esto es, a la que no sobra ni falta ninguna carta— y que todos los agentes conocen la composición de la baraja. Por ejemplo, todos saben que hay exactamente cuatro reinas, etc. Para evitar los problemas relacionados con el tiempo y la memoria, podemos concentrarnos en el primer momento del juego.

Imaginemos ahora que el juego es tal que a cada jugador se le reparten dos cartas, una boca abajo y otra boca arriba, de forma que todos los jugadores pueden verla. Imaginemos también que en este primer reparto la carta visible de cada jugador

[3] Hintikka, 1962, Sec. 5.2. Tanto esta cita como la siguiente están tomadas de la traducción española: *Saber y Creer. Una Introducción a la Lógica de las Dos Nociones*. Tecnos. Madrid 1979. Traducción y prólogo de J. Acero.
[4] Ibid. Sec.3.3.

es una reina. Aparte de que las cartas no están bien barajadas, ¿qué conclusiones puede sacar cada jugador?

Si Ana, por ejemplo, se pregunta por las cartas de Berta, concluirá que la carta oculta no es , por ejemplo, la Reina de Corazones, puesto que esta carta ya está sobre la mesa. Por lo demás, Ana no puede concluir nada más, y sabe que no puede concluir nada más. Esto es, Ana no sabe si la carta oculta de Berta es, por ejemplo, el Rey de Picas, y sabe que no lo sabe. El axioma de introspección positiva, por lo tanto, se cumple.

Hasta aquí un ejemplo a favor del axioma de introspección negativa. Desgraciadamente, las cosas no son tan fáciles como esto, porque encontrar un ejemplo en contra resulta extremadamente fácil. Veamos uno.

Supongamos un determinado agente, digamos Antonio (en adelante, *a*) que jamás ha oído hablar del filósofo alemán Immanuel Kant y que, por supuesto, tampoco conoce el título de ninguna de sus obras. Parece evidente que la proposición "*a* no sabe que Kant escribió la *Crítica de la Razón Pura*" es verdadera. ¿Lo es también la proposición "*a* sabe que no sabe que (si) Kant escribió la *Crítica de la Razón Pura*"?

La intuición nos dice que no, puesto que para eso tendría que conocer al menos la existencia del autor y de su obra. ¿Nos dirá lo mismo nuestra semántica de modelos Kripkeanos? Veamos.

Sea $s \in W$ del modelo M un mundo posible que describe el mundo real, en el que es verdad que Kant escribió la *Crítica de la Razón Pura*. Puesto que el agente *a* no lo sabe, debe haber un mundo posible $t \in W$ del modelo M tal que t es una alternativa epistémica al mundo s para el agente *a* (en símbolos, $sR_a t$) en el cual la oración "Kant escribió la *Crítica de la Razón Pura*" es falsa. Si aceptáramos además, hipotéticamente, que *a* sabe que no sabe que Kant escribió la *Crítica de la Razón Pura*, tendríamos que exigir también que en todo mundo v accesible desde s (para el agente *a*) sea verdad que a no sabe que Kant escribió la *Critica de la Razón Pura*. Es decir, que haya un mundo posible accesible desde v en el que Kant no escribió la *Crítica de la Razón Pura*; y esta condición se cumple automáticamente si la relación de accesibilidad es, además de transitiva, simétrica. La razón es que al ser simétrica, se cumple que vRs, puesto que sRv; y dado que también es verdad que sRt y que la relación es transitiva, se cumple que para todo mundo posible v accesible desde s, vRt; y recordemos que t es precisamente, por hipótesis, un mundo posible en el que Kant jamás escribió su primera crítica.

Ahora bien, hemos dicho que *a* no sabe siquiera que Kant existe, y por supuesto tampoco conoce sus obras. Si hemos de tomarnos esto como parte de la descripción del modelo, tendremos que aceptar que hay mundos posibles accesibles desde *s* en los que o bien Kant o bien su primera crítica, o ambos, ni siquiera existen. La pregunta ahora es ¿puede ser simétrica la relación de accesibilidad en un modelo como éste? Aunque el asunto es técnicamente complejo y sería necesario precisar algunas cuestiones, podemos anticipar que la respuesta es no: si un modelo tiene dominios estrictamente decrecientes (o antimonótonos), como es el caso de éste, la relación de accesibilidad no es simétrica; y lo mismo ocurre cuando trabajamos con el tipo de dominios que llamamos. monótonos o crecientes.

3. Saltando al primer orden

Casi sin darnos cuenta, mientras discutíamos sobre el axioma de introspección negativa, hemos pasado a hablar de diferentes tipos de dominios. Pero no se habla de dominios en la semántica kirpkeana de mundos posibles. ¿Qué ha ocurrido?

Lo que ha ocurrido es que al analizar ejemplos como el de la oración como "*a* no sabe que Kant escribió la *Crítica de la Razón Pura*" estábamos dando por supuesto que la estructura interna de la proposición a la que afecta el operador de conocimiento, "Kant escribió la *Crítica de la Razón Pura*", era irrelevante a la hora de tratar problemas epistémicos de importancia como el de la introspección negativa; pero esto, formalizar estos ejemplos en el nivel proposicional, hurtaba un elemento central de la discusión. Necesitamos movernos en el nivel de la lógica de predicados para darnos cuenta de un hecho trascendental: es posible que el individuo designado por el nombre "Kant" ni siquiera exista en todos los mundos posibles de nuestro modelo.

Parece, pues, que nos vemos obligados a dar el salto a la lógica epistémica de primer orden si es que queremos analizar adecuadamente estos problemas[5]. La sintaxis que necesitamos es la extensión natural de la lógica de primer orden. La semántica, en cambio, no deja de ofrecer algunos problemas de interés más que considerable. Desgraciadamente, algunas de las cuestiones que se plantean son demasiado técnicas para tratarlas en tan reducidas dimensiones. En la medida de lo posible, remitiremos a la bibliografía cuando nos encontremos con demostraciones demasiado extensas que, por ser sobradamente conocidas, no aportan demasiado a la discusión.

Para empezar, debemos adoptar alguna decisión sobre el uso de las constantes individuales: hemos de decidir si designan a los mismos individuos en todos los

---

[5] Sobre lógica modal de primer orden en general, Vid. Fitting, 1988.

mundos posibles o, por el contrario, su referencia puede variar de un mundo posible a otro. Por razones más técnicas que estrictamente filosóficas, hemos decidido adoptar la primera de estas posibilidades; es decir, consideraremos las constantes individuales como lo que Kripke denominó "designadores rígidos"[6]. Esto, además de evitarnos la distinción entre modalidades *de dicto* y *de re*, sortea la cuestión de la identidad transmundana.

Además, y con esto llegamos al punto central, necesitamos disponer de un dominio de objetos; pero esto nos obliga a adoptar ciertos compromisos sobre los objetos que existen en cada mundo posible, lo que dará lugar a semánticas alternativas.

Una posibilidad es admitir un dominio único común a todos los mundos posibles de un modelo dado; con lo cual pueden variar las propiedades que los individuos tienen en un mundo respecto a las que tienen en otro, así como las relaciones que mantienen con otros individuos del dominio, pero no los individuos propiamente dichos, que son los mismos en todos los mundos. Llamaremos a éstas "semánticas de dominio constante".

Por contraposición, hablaremos de "semánticas de dominios variables" para referirnos a aquellas en las que cada mundo posible tiene un dominio propio, que no coincide necesariamente con el de otros dominios del modelo. Podemos imponer, además, la restricción de que el dominio de cada alternativa epistémica contenga al menos los mismos individuos que el del mundo desde el que es accesible. Esto es, si denominamos $\mathfrak{D}(s)$ al dominio del mundo s y $\mathfrak{D}(t)$ al del mundo t; si $sR_{a_i}t$, entonces $\mathfrak{D}(s) \subseteq \mathfrak{D}(t)$. Hablaremos en este caso de dominios crecientes o monótonos. En el caso contrario, si $sR_{a_i}t$, entonces $\mathfrak{D}(t) \subseteq \mathfrak{D}(s)$, hablamos de dominios decrecientes o antimonótonos. Ya veremos que según que optemos por una u otra de estas opciones obtendremos sistemas con diferentes fórmulas válidas; la fórmula Barcan, por ejemplo, es válida en todos los marcos[7] variables con dominios antimonótonos, y su conversa en los marcos con dominios monótonos. Por supuesto, ambas son válidas en los marcos de modelo constante.

3.1.Semántica de dominio constante

Para este tipo de semánticas no necesitamos más que añadir a nuestros modelos el dominio $\mathfrak{D}$ común a todos los mundos. Así, un modelo kripkeano

[6] Kripke, 1986.
[7] Toscamente hablando, un marco es un modelo sin función de interpretación.

extendido será una estructura $M = \langle \mathfrak{D}, W, R_{a_i}, \cdots, R_{a_n}, \upsilon \rangle$, y la función de evaluación υ ya no es una función que asigna a cada variable proposicional en cada mundo posible un valor de verdad, sino una función que asigna:

i.a cada constante individual un elemento del dominio, y

ii.a cada constante predicativa n-ádica una n-pla ordenada de elementos del dominio.

Así pues:

a)$\upsilon(s, a) \in \mathfrak{D}$

b)$\upsilon(s, P^n) \in \mathfrak{D}^n$

Puesto que además hemos decidido considerar las constantes de individuo como designadores rígidos; tendremos que imponer la siguiente condición: para todo $s, t \in W$ y para toda constante individual a, $\upsilon(s, a) = \upsilon(t, a)$ (para abreviar, escribiremos $\upsilon(a)$en lugar de $\upsilon(s, a)$, etc.)

La noción de verdad de una fórmula φ en un mundo s de un modelo M, que como siempre escribiremos $M, s \vDash \varphi$, se define de la forma habitual, añadiendo las cláusulas correspondientes a la lógica de primer orden. Por simplicidad, no tendremos en consideración formulas abiertas; esto es, asignaremos valores de verdad tan sólo a las sentencias. La cláusula para fórmula atómicas queda definida como sigue:

$$M, s \vDash P(a_1, \cdots, a_n) \text{ syss } \langle a_1, \cdots, a_n \rangle \in \upsilon(s, P)$$

Las condiciones de verdad para las conectivas proposicionales son las habituales y las de los operadores epistémicos son las que acabamos de ver en el caso proposicional. Las condiciones para los cuantificadores son las siguientes:

$\mathrm{M}, \mathrm{s} \vDash \forall \mathrm{x} \varphi$ syss $M', s' \vDash \varphi(a/x)$ para toda M', s' tales que $\mathrm{s}' = \mathrm{s}$ y $M' =_{\upsilon(a)} M$

$M, s \vDash \exists x \varphi$ syss $M', s' \vDash \varphi(a/x)$ para algúna M', s' tales que $s' = s$ y $M' =_{\upsilon(a)} M$

La expresión $\varphi(a/x)$ que aparece en estas dos cláusulas denota el resultado de sustituir uniformemente x por a en la fórmula φ. $M' =_{\upsilon(a)} M$ significa que $M' = M$ salvo, a lo sumo, respecto al valor que la función de evaluación υ asigna a la constante a.

Respecto a estos modelos con dominio constante, es interesante considerar las que se suelen llamar, por analogía con la lógica modal alética, *Fórmula Barcan* y *Conversa de la Fórmula Barcan*[8] (respectivamente: FB y CFB):

FB: $\forall x K_{a_i} \varphi(x) \rightarrow K_{a_i} \forall x \varphi(x)$

[8] Sobre el papel de la fórmula Barcan y su conversa en lógica modal alética, puede consultarse el clásico Hughes, 1968.

CFB: $K_{a_i}\forall x\varphi(x) \rightarrow \forall x K_{a_i}\varphi(x)$

Es fácil demostrar que todo modelo con dominio constante satisface ambas fórmulas[9]. La razón es, toscamente hablando, que la Fórmula Barcan expresa la propiedad de antimonotonía y su conversa la monotonía; y naturalmente, todo modelo constante es a la vez monótono y antimonótono.

Volvamos ahora al ejemplo de nuestro juego de naipes. Insistíamos al describirlo en una característica del juego que, por el hecho de ser la condición habitual, no se suele destacar; a saber, que se trataba de una baraja estándar sin trucar cuya composición conocen todos los agentes. En términos de nuestros modelos, esto significa que en todos las distribuciones de cartas que los agentes consideran posibles —en todos los mundos posibles— existen las mismas cartas, aunque probablemente distribuidas de diferente modo. Esto es, que estamos trabajando con modelos monótonos; y que por tanto deben cumplirse tanto **CFB** como, y esto nos interesa especialmente, **FB**.

¿Qué nos dice nuestra intuición? Exactamente lo mismo, veamos un ejemplo

Supongamos que a (la Ana del juego) sabe de cada una de las cartas que realmente existen en la baraja que es o bien roja o bien negra (pero no ambas cosas, claro). Entonces también es verdad que a sabe que todas las cartas son o rojas o negras; puesto que no considera posible que haya alguna carta de otro color de la que ella no tenga noticia. Puesto que el mismo argumento se puede usar para cualquier otro ejemplo que queramos poner, podemos considerar esto como una demostración intuitiva de que nuestro ejemplo cumple la fórmula Barcan.

Volvamos ahora al axioma de introspección negativa (A5). Llamamos $CS4_m$ a la extensión natural de la lógica de primer orden con el sistema $S4$, para m agentes, de lógica epistémica (la C es por los dominios constantes), y $CS5_m$ a la misma extensión con el sistema $S5$. ¿Qué decir de la fórmula Barcan? Pues bien, En el sistema $CS4_m$ debemos añadirla como axioma, ya que no se sigue del resto de los axiomas de la teoría y sin embargo resulta válida en todos los modelos constantes. En el sistema $CS5_m$ por el contrario, no es necesario contar con ella como axioma adicional; puesto que, al añadir A5, FB se vuelve una fórmula derivable en el sistema[10].

¿Quiere esto decir que para trabajar con este ejemplo necesitamos el axioma de introspección negativa? No necesariamente. Ya hemos visto que se puede añadir **FB** al sistema $S4$ de lógica epistémica. Pero sí es cierto que resulta bastante natural, ya

[9] La demostración puede verse, por ejemplo, en Fagin, 1995 y en Gómez-Caminero Parejo, 2011.
[10] **CFB** se demuestra también en todos los sistemas más débiles.

que sin más que extender LPO con los axiomas de la lógica epistémica, obtenemos una caracterización adecuada de todos estos ejemplos que podríamos llamar *situaciones de omnisciencia existencial*: aquellos en que todos los agentes tienen una información completa de todo lo que existe.

Esta condición, sin embargo, no tiene por qué cumplirse en todos los contextos epistémicos. Lo normal, de hecho, es que no se cumpla.

Si hablamos de conocimiento, por ejemplo, y nos ceñimos a un determinado campo —filósofos clásicos y sus obras, en el ejemplo anterior— lo natural es que algunos agentes ignoren la existencia de ciertos autores u obras (no tendremos en cuenta la posibilidad de que tengan creencias falsas). Estos casos, que podemos denominar *situaciones de ignorancia existencial positiva*[11] quedan adecuadamente descritos por modelos con dominios antimonótonos o decrecientes, que como hemos dicho son modelos en los que dados $s, t \in W$ del modelo en cuestión M, y dado un agente cualquiera a_i perteneciente al conjunto $\mathcal{A}$ de agentes, si $sR_{a_i}t$, entonces $\mathfrak{D}(t) \subseteq \mathfrak{D}(s)$.

Podemos estar tentados de tratar también *situaciones de ignorancia existencial negativa*, en la que ciertos agentes podrían no saber que ciertos individuos no existen, pero parece más natural utilizar para ello el concepto de creencia, y expresar esto diciendo que "es compatible con las creencias de a que el individuo c exista" —en signos, $\hat{B}_a\mathcal{E}(c)$[12]—. Este tratamiento permite expresar también la creencia en la existencia de seres que realmente no existen; lo cual, por supuesto, no es posible con el concepto de conocimiento. El tratamiento de estas *situaciones de falsa creencia existencial*, que es como debemos considerarlas, exige el uso de modelos con dominios monótonos o crecientes[13].

Resulta atractiva la idea de combinar ambos enfoques en una lógica epistémico-doxástica con modelos antimonótonos respecto a la relación de accesibilidad epistémica pero monótonos respecto a la doxástica. Ésta es una propuesta interesante que aún no hemos emprendido.

---

[11] Entendida como ignorancia de la existencia; por oposición a la *ignorancia existencial negativa*, o ignorancia de la inexistencia, de la que hablaremos a continuación.

[12] $\mathcal{E}$ es el predicado de existencia, del que hablaremos más adelante.

[13] Puesto que hemos visto que estas situaciones deben estudiarse en el contexto de una lógica doxástica, en adelante nos limitaremos a la que hemos denominado *ignorancia existencial positiva*. Cuando hablemos de ignorancia existencial, sin más, nos estaremos refiriendo a ella.

3.2. Semántica de dominios variables

Estamos llegando, con esto, al núcleo central de nuestra argumentación: sostenemos que el axioma de introspección negativa resulta contraintuitivo porque la mayoría de las situaciones que podríamos considerar naturales son situaciones de *ignorancia existencial*, y tales situaciones requieren ser formalizadas en una lógica de dominios variables antimonótonos. La razón por la que esto es un obstáculo para aceptar A5 es que en el sistema $S5$ la relación de accesibilidad es reflexiva, y por tanto todo modelo con dominios antimonótonos es también monótono, y por tanto constante[14]. Pero antes de argumentar esta posición, debemos hablar, siquiera someramente, de los modelos de dominios variables.

Para trabajar con este tipo de modelos es interesante contar con el predicado de existencia $\mathcal{E}$, que hemos mencionado hace un momento. Tal predicado puede ser introducido como primitivo, Pero también se puede definir como $\mathcal{E}(c) =_{def} \exists x(x = c)$

Tal fórmula, por supuesto, es válida en una lógica clásica de primer orden, y también en nuestra lógica de dominios constantes, pero no lo va a ser en una lógica de dominios variables. De manera que $\mathcal{E}(c)$ va a resultar verdadera precisamente en aquellos mundos en los que el individuo designado por la constante c existe.

La semántica de este tipo de lógicas requiere un dominio específico $\mathfrak{D}_s$ para cada mundo posible s; además del dominio $\mathfrak{D}$ del modelo, que debe cumplir que $\mathfrak{D} = \{\mathfrak{D}_s \cup \mathfrak{D}_t \cup \cdots\}$ (para todo $s, t, \cdots \in W$, de la estructura M). Hay varias maneras posibles de presentar esto, una de ellas es ampliar la definición de la función de evaluación v de forma que asigne a cada mundo posible un subconjunto de $\mathfrak{D}$. Un modelo kripkeano extendido variable es como antes una estructura $M = \langle \mathfrak{D}, W, R_{a_i}, \cdots, R_{a_n}, v \rangle$, donde W y $R_{a_1}, \cdots, R_{a_m}$ se interpretan como hasta ahora, $\mathfrak{D}$ es un conjunto no vacío de individuos y v es una función de evaluación tal que:

a) $v(s, \mathfrak{D}) \in \mathcal{P}(\mathfrak{D})$ (escribiremos $\mathfrak{D}_s$ en lugar de $v(s, \mathfrak{D})$).

b) $v(s, a) \in \mathfrak{D}$ (como anteriormente, estipulamos que para todo $s, t \in W$, $v(s, a) = v(s, a)$; por lo que abreviaremos escribiendo simplemente $v(a)$).

c) $v(s, P^n) \in \mathfrak{D}^n$.

[14] Técnicamente no se trataría de un modelo de dominio constante, sino de un modelo de dominio localmente constante; pero estos dos tipos de modelos son equivalentes en el sentido de que el conjunto de fórmulas válidas es el mismo en ambos.

La verdad de una fórmula en un mundo posible $s \in W$ de la estructura M se define de forma semejante al caso de los dominios constantes. El cambio más relevante es el que se refiere a los cuantificadores:

$M, s \vDash \forall x\varphi$ syss $M, s \vDash \varphi(a/x)$ para todo M', s' tales que $s = s'$ y $M' = M[\upsilon(a)/s]$

$M, s \vDash \exists x\varphi$ syss $M, s \vDash \varphi(a/x)$ para algún M', s' tales que $s = s'$ y $M' = M[\upsilon(a)/s]$

La expresión $M' = M[\upsilon(a)/s]$, (que leeremos "M' es una variante de M para a en s") significa que M' difiere de M a lo sumo en el valor que la función de evaluación asigna a la constante a y que además este valor pertenece a $\mathcal{D}_s$.

La intuición que subyace a estas cláusulas es que el ámbito de variabilidad de los términos es el dominio del modelo, mientras que el de los cuantificadores es el dominio de cada mundo posible, de manera que puede ocurrir que una fórmula de la forma $\forall x\varphi(x)$ sea verdadera en un mundo s de un modelo M, y que sin embargo $\varphi(t)$ sea falsa en ese mismo mundo, siempre que $\upsilon(t)$ no pertenezca a $\mathcal{D}_s$. Por ejemplo, en el mundo real es verdad que todos los hombres son mortales, y también que Aquiles era inmortal (salvo por un pequeño problema con el talón); el que ambas cosas no supongan inconsistencia se debe, por supuesto, al hecho de que Aquiles no es un personaje real.

Esta semántica tiene la peculiaridad de que, si bien las constantes individuales designan a un mismo individuo del dominio en todos los mundos posibles, ese individuo no existe necesariamente en cada uno de ellos. Esta es una característica propia de una familia de lógicas conocidas como Lógicas Libres[15]. Obsérvese que la función de evaluación υ asigna a cada predicado n-ádico, en cada mundo posible s, un conjunto de n-plas ordenadas de $\mathcal{D}_n$, y no de a $\mathcal{D}_s$; de manera que una fórmula de la forma $P(a)$ puede ser verdadera en un mundo posible s incluso en el caso de que $a \notin \mathcal{D}_s$; y por tanto, $M, s \vDash \neg\mathcal{E}(a)$. Cuando esto ocurre, se habla de una lógica libre positiva. Hay otras opciones interesantes, tanto desde un punto de vista técnico como filosófico, pero no las tendremos en consideración en este lugar.

Hemos mencionado anteriormente la posibilidad de imponer a nuestros dominios variables las condiciones adicionales que llamamos *monotonía* (si $sR_{a_i}t$, entonces $\mathfrak{D}(s) \subseteq \mathfrak{D}(t)$) y *antimonotonía* (si $sR_{a_i}t$, entonces $\mathfrak{D}(t) \subseteq \mathfrak{D}(s)$). Por supuesto, los modelos de dominio constante son a la vez monótonos y antimonótonos; pero los modelos de dominios variables pueden tener o no estas dos

[15] Una buena presentación de las lógicas libres se encuentra en Priest, 2008. Sobre Lógica Epistémica Libre, en particular, aunque con una presentación algo diferente, Lenzen, 2001.

propiedades. Por supuesto, pueden tener ambas propiedades a la vez; pero entonces se trata de dominios localmente constantes; que, como hemos dicho, satisfacen las mismas fórmulas que modelos de dominio constante.

Es el momento de volver a otro punto que anticipábamos al principio de este trabajo: la propiedad de antimonotonía es expresada por la Fórmula Barcan y la monotonía por su conversa. Efectivamente, respecto de la semántica que acabamos de esbozar es posible demostrar estos dos teoremas[16]:

Teorema 1: Sea $\mathcal{F}^+$ el conjunto de todos los marcos[17] kripkeanos con dominios monótonos. Se cumple que $\mathcal{F} \vDash K_{a_i}\forall x P(x) \rightarrow \forall x K_{a_i} P(x)$ syss $\mathcal{F} \in \mathcal{F}^+$.

Teorema 2: Sea $\mathcal{F}^-$ el conjunto de todos los marcos kripkeanos con dominios antimonótonos. Se cumple que $\mathcal{F} \vDash \forall x K_{a_i} P(x) \rightarrow K_{a_i}\forall x P(x)$ syss $\mathcal{F} \in \mathcal{F}^-$.

¿Qué tiene que ver esto con las situaciones de ignorancia existencial y omnisciencia existencial de las que estamos hablando? Es fácil de ver si sustituimos en las dos fórmulas anteriores la letra predicativa P por el predicado de existencia $\mathcal{E}$. Combinando CFB y FB tenemos:

$$K_{a_i}\forall x \mathcal{E}(x) \leftrightarrow \forall x K_{a_i}\mathcal{E}(x)$$

La parte de la izquierda de esta equivalencia es trivial: nos dice tan sólo que a_i sabe que todo existe, lo cual es siempre verdad de las cosas que existen en cada mundo posible (que bien podrían ser distintas). La parte de la derecha, en cambio, dista mucho de ser trivial: nos dice que de todo lo que existe (en el mundo concreto en el que evaluemos) a_i sabe que existe. Y puesto que la parte izquierda es válida en todos los mundos posibles y la relación de accesibilidad es simétrica (nos movemos en S5), tampoco puede suceder que el agente considere posible la existencia de más seres de los que existen en el mundo real, ya que en ese mundo posible el agente debe saber que esos seres existen; y por tanto, existirán también en todos los mundos accesibles desde él, incluido el real. Resumiendo, esta fórmula caracteriza adecuadamente lo que hemos dado en llamar *situaciones de omnisciencia existencial*.

Atemos por fin los últimos cabos de nuestro argumento. Hemos mencionado, y es sobradamente conocido, que el Axioma de Introspección Negativa caracteriza sintácticamente el sistema S5 de lógica epistémica, que a su vez se caracteriza

[16] No presentamos la demostración de estos dos teoremas. El lector interesado puede consultarlas, por ejemplo, en Fagin, 1995 o en Gómez-Caminero Parejo, 2011.

[17] Como ya hemos dicho, un marco es un modelo sin función de interpretación. Por razones técnicas, hablamos de marcos y no de modelos; igual que presentamos la Fórmula Barcan y su conversa como fórmulas concretas, y no como esquemas de fórmulas. Pero esto no afecta al núcleo de la cuestión; a saber, que la Fórmula Barcan caracteriza la propiedad de antimonotonía y su conversa la monotonía.

semánticamente porque la relación de accesibilidad es reflexiva, simétrica y transitiva. Ahora bien, en una relación tal todo modelo monótono es a la vez antimonótono; y por tanto, localmente constante. Ya hemos visto que en los modelos de este tipo valen tanto la Fórmula Barcan como su conversa; y que estas dos fórmulas, instanciadas con el predicado de existencia, caracterizan las situaciones que hemos denominado *de omnisciencia existencial*. Tales situaciones, por supuesto, pueden darse; pero sólo en contextos restringidos y artificiales. Esta es la razón de que el axioma A5 nos resulte tan antinatural.

4. Conclusiones

Nuestro propósito en este trabajo era intentar determinar por qué el denominado Axioma de Introspección Negativa, por lo demás, ampliamente aceptado en la literatura, resultaba tan poco natural para nuestro sentido común. El análisis de un par de ejemplos nos llevó a la conclusión de que debíamos fijarnos en la lógica de primer orden. Una primera versión de la lógica epistémica de primer orden se obtiene utilizando una semántica de dominio constante, que da lugar a la extensión natural de la lógica clásica de primer orden. Esta lógica, en la que valen tanto la Fórmula Barcan como su conversa, tiene la limitación de que sólo podemos tratar el tipo de situaciones en las que los agentes saben exactamente cuáles son los objetos del dominio, ni ignoran la existencia de ninguno de ellos ni consideran posible la existencia de seres que no existan realmente.

Es tipo de situaciones, por supuesto, se da, pero son infrecuentes y artificiales, lo cual nos lleva a trabajar con semánticas donde los elementos del dominio pueden cambiar de un mundo posible a otro y en las que es posible considerar situaciones de *ignorancia existencial positiva* (dejamos para otro momento la falsa creencia existencial). Esta ignorancia existencial queda adecuadamente caracterizada por una semántica de dominios antimonótonos o decrecientes. Ahora bien, es bien sabido que el Axioma de Introspección Negativa caracteriza el sistema S5 en que la relación de accesibilidad es reflexiva, simétrica y transitiva y en la que todos los modelos antimonótonos son también monótonos, y por tanto constantes; por lo que de nuevo, valen tanto la Fórmula Barcan como su conversa. Estos modelos siguen pues caracterizando la situación de *omnisciencia existencial*, que es precisamente la situación que queríamos evitar.

Si, como sostenemos, la situación más natural es la que denominamos *ignorancia existencial positiva* y queda adecuadamente descrita por una semántica de dominios antimonótonos, necesitamos una lógica epistémica libre con los axiomas A1-A4 más la Fórmula Barcan, pero no su conversa[18].

Ha llegado la hora de responder a la pregunta planteada en el título de este artículo: ¿era realmente Sócrates tan sabio?

¡Naturalmente que sí! Incluso sin considerar nuestro axioma, es obvio que reconocer la propia ignorancia exige sabiduría. Si lo que Sócrates hubiera querido decir es que de cada proposición que ignoraba, sabía que la ignoraba, que no es el caso, eso implicaría saber exactamente todo lo que existe en el mundo; y esto, sin duda, es una sabiduría inalcanzable.

[18] Estudiamos estas posibles combinaciones en Gómez-Caminero Parejo, 2011.

Referencias

Fagin, R., Halpern, J.Y., Moses, Y. y Vardy, M.Y. *Reasoning About Knowledge.* Cambridge: The MIT Press, 1995.
Fitting, M. y Mendelsohn, R. L. *First-Order Modal Logic.* Dordrecht: Kluwer Academic Publishers, 1988.
Gómez-Caminero Parejo, E. F. *Tablas Semánticas para Lógica Epistémica.* Sevilla: Fénix Editora, 2011.
Hintikka, J. *Knowledge and Belief.* Cornell: Cornell University Press, 1962.
Hocutt, M. O. "Is Epistemic Logic Posible." *Notre Dame Journal of Formal Logic* (Duke University Press) XXII, no. 4 (October 1972): pp.: 433-453.
Hughes, G. E. y Cresswell, M. J. *An Introduction to Modal Logic*. London: Methuen and Co Lt., 1968.
Kripke, S.A. *Naming and Necessity*. Oxford: Basil Blackwell, 1986.
Lenzen, W. "Free Epistemic Logic." In *New Essays in Free Logic*, by E. y Hieke, A. (eds) Morscher, 117-124. Dordrecht: Kluwer Academic Publishers, 2001.
Meyer. J.J. y van der Hoek, W. *Epistemic Logic for AI and Computer Sciences.* Cambridge: Cambridge University Press, 1995.
Priest, G. *An Introduction to Non-Classical Logic: from if to is.* Cambridge: Cambridge University Press, 2008.