Casa abierta al tiempo
UNIVERSIDAD AUTÓNOMA METROPOLITANA
Unidad Iztapalapa

División de Ciencias Sociales y Humanidades
Posgrado en Humanidades
Línea en Fílosofía de las Ciencias y del Lenguaje

Idónea Comunicación de Resultados que para obtener el grado de Maestro en

Humanidades, Línea de Filosofía de las Ciencias y del Lenguaje, presenta:

Adán Contreras Quintero

Consideraciones en torno a la Lógica Epistémica

Asesores:

Dra. Yolanda Magda Torres Falcón y Dr. José Marcos Nicolás Javier de Teresa Ochoa

Jurado:

Presidente: Dr. José Marcos Nicolás Javier de Teresa Ochoa

Secretaria: Dra. Yolanda Magda Torres Falcón

Vocal: Dr. Victor Cantero Flores

Vocal: Dr. José Jorge Max Fernández de Castro Tapia

Lectores:

Dr. José Jorge Max Fernández de Castro Tapia y Dr. Victor Cantero Flores

Iztapalapa, Ciudad de México, 2 de marzo de 2021

Esta tesis se realizó con el apoyo de CONACYT

Índice

Introducción

Este trabajo es una introducción a la Lógica Epistémica de corte modal. De ello hay que tomar en consideración dos cosas:

I) No se trata aquí de la Lógica Epistémica en general, sino de un tipo particular de ella, a saber, una que pertenece al conjunto de la Lógica Modal[1]. La lógica que aquí se analiza emplea las herramientas formales desarrolladas para el estudio de las nociones de "necesidad" y "posibilidad", y las aplica al análisis de algunos conceptos epistemológicos tales como "conocimiento" y "creencia". Por consiguiente, el objeto de nuestro estudio es la Lógica Epistémica *Modal*, la cual es eminentemente formal y hace de las nociones epistémicas formas de modalidad, a las que se les da el nombre de *modalidades epistémicas*. De manera elemental, una modalidad es una manera de afirmar una proposición. Por ejemplo, dada una proposición *p*, yo puedo afirmarla de formas distintas: "*p* es verdadera", "*p* es necesaria", "*p* es posible"...; cada una de éstas es una modalidad de *p*. Así, entender los conceptos epistémicos como modalidades significa que declaraciones del tipo " sé que *p*", "se cree que *p*" o "*p* está justificada", constituyen formas de afirmar la proposición *p*.

En términos generales, la importancia de toda lógica modal, incluida la epistémica, reside, en parte, en su valor como herramienta para el análisis de conceptos filosóficos —en este sentido, constituye una alternativa lógico-formal al puro análisis conceptual a través del uso ordinario de las expresiones—; razón por la cual las lógicas modales forman parte de la *lógica filosófica*. En cuanto a la Lógica Epistémica Modal, se espera de ella el esclarecimiento de nociones de la Epistemología o Teoría del Conocimiento, *verbi gratia*, “creencia”, “verdad”, “justificación” y “conocimiento”. Así,

[1] Algunos desarrollos paralelos que merecen el nombre de lógicas epistémicas son: la lógica de John Dewey, tal como la presenta en su *Logic: Theory of Inquiry* (1938); y la lógica del descubrimiento (ver Aliseda, Atocha, *Abductive Reasoning_ Logical Investigations into Discovery and Explanation*, Synthese Library 330, Springer, 2006).

la pregunta por su éxito o fracaso en semejante tarea conforma el problema filosófico general de la Lógica Epistémica.

II) Su carácter *introductorio*. Mi intención es proveer de las herramientas necesarias a cualquiera que desee iniciarse en el estudio de esta lógica. Por esta razón, la mitad de estas páginas está dedicada a la explicación y demostración de nociones y afirmaciones básicas, las cuales, aunque prescindibles para aquellos bien informados en Lógica Modal, resultan indispensables para el profano. Como consecuencia, el lector encontrará múltiples ejemplos de la demostración de teoremas (derivaciones de fórmulas al interior de una teoría formal) y metateoremas (pruebas de afirmaciones generales acerca de una teoría formal); también hallará diagramas diversos, cuya finalidad es hacerle más asequible alguna definición o prueba. La principal motivación que tuve para ello fue el carácter escueto de muchas introducciones a la Lógica Epistémica, las cuales, aunque muy útiles, son semejantes al oráculo de Delfos, pues *no revelan ni ocultan la verdad, sólo la insinúan*[2].

Además de esta labor de hormiga en la construcción de explicaciones y pruebas, este trabajo es introductorio en otro sentido, ya que las páginas restantes, la otra mitad, están consagradas a temas epistemológicos o, con mayor precisión, al vínculo entre la Teoría del Conocimiento y la Lógica Epistémica. Es éste un punto olvidado en la literatura contemporánea[3], aunque no por ello deja de ser un punto de partida obligatorio: tal conexión se hace patente desde el nombre mismo 'Lógica Epistémica'. Cabe aclarar aquí que no tomo a mi cargo los problemas concernientes a los orígenes de esta lógica, o las razones y motivos que llevaron al abandono de su relación con la Teoría del Conocimiento. En su lugar, me concentro en el estatus epistemológico de los principios fundamentales de esta lógica, *videlicet*: la cláusula modal de conocimiento

---

[2] *Confer*, por ejemplo, Freund, Max, Lógica Epistémica, en: *Enciclopedia iberoamericana de filosofía (Volumen 7)*, Trotta (Ed.), 1995, p. 207; Gochet P. y Gribomont P., "Epistemic logic", en: *Handbook of the history of logic* (vol 7), M. Gabbay y Wood J. (Editores), 2006; Hans van Ditmarsch, Joseph Y. Halpern, Wiebe van der Hoek, Barteld Kooi (Eds.), *Handbook of Epistemic Logic*, College Publications (2015).

[3] Para hacer patente este olvido, también es útil ver los dos handbooks sugeridos en la nota anterior.

(según la cual un agente conoce una proposición si y sólo si ésta es el caso en todas las posibilidades a las que él tiene acceso en un estado determinado); el supuesto aristotélico (o la tesis de que todo lo demostrado constituye conocimiento); el supuesto de transmisión de conocimiento (*id est*, que las propiedades del saber se heredan a través de la deducción); el principio de conocimiento (o bien, la idea de que el conocimiento proposicional precisa de la verdad para ser conocimiento auténtico); el principio de introspección positiva (*i. e.*, la tesis de que se conoce una proposición sólo si se sabe que se la conoce), y el principio de introspección negativa (o el supuesto de que siempre que se ignora una proposición, se sabe que se la ignora). Desde luego, dichos principios no surgieron con la Lógica Epistémica, sino que pertenecen al ámbito de la Teoría del Conocimiento, y aquélla no hace sino abrigarlos como punto de arranque para examinar sus implicaciones inferenciales y consecuencias. Mi trabajo consiste, en gran medida, en la determinación de tales consecuencias para la Epistemología[4].

Los puntos I) y II) delimitan los alcances de nuestra investigación.

En cuanto a su estructura, este escrito se halla seccionado en seis capítulos. En el primero de ellos, como su nombre lo indica, se presentan nociones básicas o fundamentales, en el sentido de que resultan necesarias para el estudio de la Lógica Epistémica Modal, así como para la comprensión de los capítulos subsecuentes en este trabajo. Entre estas nociones básicas, destaca la cláusula modal del conocimiento. Los capítulos del II al IV contienen el análisis de cuatro teorías axiomáticas "estándar" —es decir, teorías definidas en un lenguaje formal con un único operador modal, el de conocimiento— para esta lógica, así como de los supuestos epistémicos fundamentales a ellas relacionados: la teoría $\mathbb{K}$ y los supuestos aristotélico y de preservación del conocimiento (cap. II); la teoría $\mathbb{T}$ y el principio de conocimiento (cap. III); la teoría $\mathbb{S}4$

[4] Algunos trabajos muy interesantes acerca del vínculo entre Epistemología y Lógica Epistémica son: Hilpinen, R.: 1970, “Knowing that One Knows”, Synthese 21, 109-131; Hintikka, J.: 1970, “Knowing that One Knows’ Reviewed”, Synthese 21, 141-162; y, por su puesto, Lewis, D.: 1996, “Elusive Knowledge”, The Australasian Journal of Philosophy 74, 549-567.

y el principio de introspección positiva, así como la teoría S5 y el principio de introspección negativa (cap. IV). En el capítulo V, se presenta una manera de traducir fórmulas de un lenguaje modal a fórmulas de un lenguaje de primer orden, lo cual abre una reflexión acerca de la importancia de los lenguajes modales. En el capítulo VI y final, se introduce al estudio de la Lógica Multimodal: aquella que trabaja con lenguajes con dos o más operadores modales [*v. g.*, uno de conocimiento y otro de creencia]. En las conclusiones, el lector encontrará un breve comentario sobre uno de los problemas clásicos en Lógica Epistémica, a saber, el problema de la omnisciencia lógica; además de un balance de las discusiones epistemológicas de los capítulos precedentes. Por último, el cierre de este escrito está dado por tres apéndices: el primero incluye la demostración de algunos teoremas que se obtienen en los sistemas K, T, S4 y S5; el segundo, el metateorema de completez para cada uno de estos sistemas; y el tercero, el metateorema de corrección para cuatro sistemas de Lógica Epistémica Multimodal. La inclusión de estos tres apéndices obedece al cumplimiento de la promesa de abundar en las demostraciones, pero también al intento de descargar los capítulos de consideraciones formales y, de este modo, tornar mucho más sencilla la lectura de las partes centrales de esta investigación.

Como se desprende de lo anterior, este trabajo es híbrido, pues si bien cada uno de sus capítulos involucra un desarrollo puramente formal de la Lógica Epistémica, también contempla una discusión epistemológica de sus supuestos fundamentales. Tales secciones epistemológicas aparecen bajo el subtítulo de "Consideraciones...", y son las que le dan nombre a esta tesis, a la vez que constituyen, de alguna manera, su parte más original.

Finalmente, dejo una disculpa para el lector acostumbrado a las abundantes referencias bibliográficas al final de un trabajo de semejante naturaleza. Las mías son mínimas, y ello sólo tiene una explicación: <<*porque en esta ocasión al menos, yo también me sentí "poltrón y perezoso de andarme buscando autores que digan lo que yo me sé decir sin ellos"*>>.

1. Nociones fundamentales

1.1. Antecedentes y noción de la Lógica Epistémica[5]

Considérese el siguiente argumento:

i) Edipo *sabe* que la peste en Tebas es causada por el asesino de Layo.
ii) Si la peste en Tebas es causada por el asesino de Layo, ésta no cesará hasta que aquél expíe su crimen.
∴
c) La peste no cesará hasta que el asesino de Layo expíe su crimen.

Desde el punto de vista de la lógica clásica, este argumento es incorrecto pues sus premisas podrían ser todas verdaderas y su conclusión ser falsa, y ello principalmente porque la proposición expresada en i) no es la misma que aquella que hace de antecedente en ii); no tienen, de hecho, las mismas condiciones de verdad. Sin embargo, alguien podría oponer que el razonamiento, después de todo, es correcto, ya que la verdad de la proposición 'la peste en Tebas es causada por el asesino de Layo' se encuentra implicada en la proposición 'Edipo sabe que la peste en Tebas es causada por el asesino de Layo'. Pero con ello se está reconociendo que su corrección no descansa únicamente en la definición de verdad de los conectivos y operadores de los que se ocupa la lógica clásica y, por tanto, en los principios de razonamiento estudiados por ésta, sino sobre todo en las propiedad y relaciones lógicas de la expresión 'sabe', que aparece en la primera premisa del razonamiento. A este tipo de razonamientos, *i. e.*, a los razonamientos en cuya validez o invalidez juegan un papel crucial expresiones epistémicas tales como 'conocer', ´saber', 'creer', etcétera, así como sus derivados, los referiremos como *razonamientos en torno al conocimiento*.

[5] Para la noción de lógica epistémica y sus antecedentes, me he apoyado principalmente en el primer apartado, "Setting the Stage", de *Epistemic logic: a survey of the logic of knowledge* (2005), de Nicholas Rescher.

Caracterizaremos, entonces, a la Lógica Epistémica (LE) como aquella rama de la lógica filosófica que se encarga de la investigación de los principios generales de los razonamientos en torno al conocimiento, con el objeto de elucidar sus implicaciones inferenciales y sus consecuencias[6]. Dicha investigación se ha centrado principalmente en el conocimiento proposicional: aquel que hace referencia a un hecho susceptible de ser expresado mediante una proposición; podemos ejemplificar esta forma de conocimiento con el enunciado <<Aquileo sabe que *morirá junto a las puertas Esceas*>>, o bien, con este otro <<Sherlock Holmes no sabe que *la Tierra gira alrededor del Sol*>>. No obstante, también se ha prestado atención, aunque de manera secundaria, al conocimiento interrogativo: aquel que se expresa mediante adverbios interrogativos (*qué, cómo, cuándo,* etcétera); como una ilustración de éste téngase el enunciado <<Aquileo sabe *dónde* morirá>> o <<Sherlock Holmes no sabe *quién* es Thomas Carlyle>>. Tal restricción al conocimiento proposicional y al conocimiento interrogativo limita el alcance de la investigación en LE; no obstante, mediante el primero es posible expresar otras formas de conocimiento, como lo son el *conocimiento de objetos* (que involucra tanto objetos individuales como colecciones de éstos), así como un tipo de saber-cómo [*know-how*], llamado *conocimiento procedimental*[7]. Por ejemplo, la declaración:

---

[6] La LE también se puede entender como el análisis lógico-formal de las expresiones u operadores de la forma <<S sabe que α>> (donde S y α son, respectivamente, un agente de conocimiento y una proposición cualesquiera) así como otras expresiones afines, *v. g.*, <<S cree que α>>, <<α es cognoscible>>, <<α es cognoscible de forma constructiva>>, <<α puede ser probado>>, <<α es de conocimiento común>>, etcétera; a los operadores de este tipo se les conoce generalmente como *operadores o modalidades epistémicos*. Por consiguiente, podemos caracterizar a la lógica epistémica como el estudio de las propiedades y relaciones lógicas de los operadores epistémicos. *Cf.,* Freund, Max, *Lógica epistémica*, en: Enciclopedia iberoamericana de filosofía (Volumen 7), Trotta (Ed.), 1995, p. 207.

[7] En general, existen dos tipos de saber-cómo. El saber-cómo procedimental: *x* sabe cómo la actividad A es realizada, en el sentido de que *x* puede dar instrucciones para realizar A; y el saber-cómo performativo: *x* sabe cómo hacer la actividad A, en el sentido de que *x* puede hacer A. De estos dos tipos de saber-cómo, únicamente el primero es susceptible de una reducción o reformulación mediante el conocimiento proposicional.

‘Tiresias conoce la identidad del asesino del rey Layo’ [conocimiento de objetos] puede ser reformulada diciendo que:

> ‘Hay una proposición *p* tal que ésta expresa quién es el asesino del rey Layo (*v. g.,* <<Edipo es el asesino de Layo>>) y Tiresias sabe que *p* (*i. e*., Tiresias sabe que *Edipo es el asesino de Layo*)’ [conocimiento proposicional].

Por otro lado, la afirmación:

> ‘Bonvicino de la Riva sabe cómo hay que comportarse en la mesa’ [saber-cómo procedimental],

se puede reformular, empleando la maquinaria del conocimiento proposicional, como sigue:

> ‘Bonvicino de la Riva sabe que *no se debe beber de la sopera ni abalanzarse sobre la fuente, ni mucho menos hacer ruidos con la boca como una foca’* [conocimiento proposicional][8].

Si bien la LE se desarrolló en la Edad Media, particularmente durante los siglos XII al XV, la investigación contemporánea sobre LE se retrotrae a la segunda mitad del siglo XX. Algunos preliminares al respecto se hallan en *Meaning and necessity* (1947), de Rudolf Carnap, en sus deliberaciones en torno a los enunciados de creencia. Un año después, en 1948, encontramos en el artículo *Many valued logics and the formailization of intensional fuction*, del lógico polaco Jerzy Łoś, un desarrollo de lo que él llama una lógica de la “creencia” o “aceptación”; en este trabajo, Łoś presenta una axiomatización para el operador ‘Lxp’, que emplea para simbolizar expresiones de la forma <<el individuo x cree (o está comprometido con) la proposición p>>. Durante los años 50, las publicaciones de algunos lógicos como Arthur Prior, G. H. von Wright, entre otros, ampliaron el rango de investigación en este campo, mas no fue sino hasta los 60 cuando vio la luz el primer libro sobre la materia, *Knowledge and belief: an introduction to the*

---

[8] Extraído de un texto que refleja el comportamiento de la clase alta del S. XIII. Citado en: Elías Norbert, *El proceso de la civilización*, Fondo de Cultura Económica (Ed.), México, cap. 2, 4, II.

logic of the two notions (1962), escrito por Jaakko Hintikka. Desde entonces, la investigación en Lógica Epistémica ha sido continua y con el presente trabajo, por modesto que sea, se intenta contribuir al desarrollo de la misma.

1.2. Lenguajes formales epistémicos

En el curso de la investigación, la LE emplea lenguajes formales (*i. e.*, sin significado) para el análisis de los razonamientos sobre el conocimiento, esto con la finalidad de no caer en *el embrujo del lenguaje* –para emplear una expresión de Frege–, es decir, para evitar ambigüedades y demás oscuridades que emergen con facilidad cuando se intenta hacer lógica en las diversas lenguas existentes. Los lenguajes formales que emplea esta lógica constituyen extensiones del lenguaje formal proposicional, así como de los lenguajes de primer orden y órdenes superiores. A continuación, se presenta una definición general de los lenguajes proposicionales y de primer orden para la LE.

Para definir un lenguaje formal epistémico tenemos que dar los símbolos que lo componen, así como sus reglas de formación para fórmulas bien formadas (*fbf's*).

1.2.1. *ℒ-epistémico proposicional*

Un lenguaje epistémico proposicional consta de los siguientes símbolos:

I) Un conjunto infinito numerable de letras proposicionales p_1, p_2, p_3, …

II) ¬, →

III) (,)

IV) Un conjunto de n operadores[9]

[9] Según se esté trabajando con modalidades epistémicas absolutas o relativas, se emplean operadores de la forma ‘O’ (sin subíndices) o de la forma ‘Om’ (con subíndices). Las modalidades absolutas son aquéllas en donde no aparece explícitamente el agente de conocimiento, como ocurre en las expresiones <<α es cognoscible>> o<<α puede ser probado>>, donde solamente se afirma algo sobre la proposición ‘α’, mas no se menciona nada sobre el sujeto de conocimiento. Por el contrario, en las modalidades relativas, figura explícitamente el agente de conocimiento, como en las expresiones <<S sabe que α>> o <<S cree que α>>, en donde ‘S’ es dicho agente. En este trabajo, emplearemos lenguajes epistémicos cuyos

O_1, ..., O_n

Reglas de formación de *fbf* ´s de $\mathcal{L}$:

I) Toda letra proposicional es una *fbf*.

II) Si α y β son *fbf* ´s, también lo son:

$(\neg\alpha)$, $(\alpha\rightarrow\beta)$, $(O_j\alpha)$ con con $j=1, 2, ..., n$.

III) No hay más *fbf* ´s.

Definiciones:

$\alpha\wedge\beta \stackrel{\text{def}}{=} \neg(\alpha\rightarrow\neg\beta)$

$\alpha\vee\beta \stackrel{\text{def}}{=} (\neg\alpha\rightarrow\beta)$

$\alpha\leftrightarrow\beta \stackrel{\text{def}}{=} (\alpha\rightarrow\beta)\wedge(\beta\rightarrow\alpha)$

1.2.2. $\mathcal{L}$- epistémicos de primer orden

Un lenguaje epistémico de primer orden $\mathcal{L}$ consta de los siguientes símbolos:

Símbolos lógicos:

I) Un conjunto numerable de variables individuales $\{x_i: i\in\mathbb{N}\}$

$x_1, x_2, \ldots x_n, \ldots$

II) Símbolo de identidad

$\approx$

III) Conectivos lógicos

$\neg, \rightarrow$

IV) Cuantificadores

---

conjuntos de operadores tienen subíndices, pues éstos, a diferencia de las modalidades absolutas, permiten expresar relaciones entre los agentes de conocimiento, razón por la cual se les llama lenguajes epistémicos de multi-agentes. Cf. Gochet P. y Gribomont P., "Epistemic logic", en: Handbook of the history of logic (vol 7), M. Gabbay y Wood J. (Editores), 2006.

$\forall$, $\exists$

V) Paréntesis

(,)

Símbolos no lógicos:

VI) Un conjunto (posiblemente vacío) de constantes individuales $\{c_k: k \in K\}$, donde K es un conjunto cualquiera, que funciona como el conjunto de índices.

VII) Un conjunto (posiblemente vacío) de símbolos relacionales o de predicados $\{P_i: i \in I\}$, donde I es un conjunto cualquiera, que funciona como el conjunto de índices.

VIII) Un conjunto (posiblemente vacío) de símbolos funcionales $\{f_j: j \in J\}$, donde J es un conjunto cualquiera, que funciona como el conjunto de índices.

IX) Un conjunto (posiblemente vacío) de operadores $\{O_m: m \in M\}$, donde M es un conjunto cualquiera, que funciona como el conjunto de índices[10].

Notación: $\mathscr{L} = \{\{c_k: k \in K\}, \{P_i: i \in I\}, \{f_j: j \in J\}, \{O_m: m \in M\}\}$

Observación: a cada símbolo relacional y funcional se le asocia un número natural ≥ 1, su aridad[11].

Reglas de formación de *fbf* 's de $\mathscr{L}$:

Términos de $\mathscr{L}$:

[10] Aunque K, I, J y M son conjuntos arbitrarios, muchas veces se les considera como subconjuntos de los números naturales.

[11] Esto último se puede hacer por medio de funciones:

$\mu: I \longrightarrow \mathbb{N} \setminus \{0\}$, donde $\mu(i)$ es la aridad de P_i;

$\nu: J \longrightarrow \mathbb{N} \setminus \{0\}$, donde $\nu(j)$ es la aridad de f_j.

I) Variables individuales

II) Constantes individuales

III) f_j ($t_1 \ldots t_n$) si la aridad de f_j es n y $t_1, \ldots, t_n$ son términos de $\mathcal{L}$.

fbf 's:

I) $t_1 \approx t_2$, donde t_1 y t_2 son términos de $\mathcal{L}$.

II) $P_i\, t_1 \ldots t_n$ si $t_1, \ldots, t_n$ son términos de $\mathcal{L}$ y P_i es un predicado n-ario

III) Si α y β son *fbf 's*, también lo son:

$(\neg\alpha), (\alpha\rightarrow\beta), (O_j\alpha)$ con $j=1, \ldots, m$

IV) Si α es una *fbf* y v es una variable individual, también es *fbf*:

$\forall v\alpha, \exists v\alpha$

V) No hay más *fbf 's*[12].

Definiciones:

$\alpha\wedge\beta \overset{\text{def}}{=} \neg(\alpha\rightarrow\neg\beta)$

$\alpha\vee\beta \overset{\text{def}}{=} (\neg\alpha\rightarrow\beta)$

$\alpha\leftrightarrow\beta \overset{\text{def}}{=} (\alpha\rightarrow\beta)\wedge(\beta\rightarrow\alpha)$

$\exists v\alpha \overset{\text{def}}{=} \neg\forall v\neg\alpha$

---

[12] Aunque aquí sólo se ha dado una definición general de los lenguajes formales epistémicos proposicionales y de primer orden, habrá que decir que también es posible definir un lenguaje formal epistémico de orden superior. En cualquier caso, la extensión a los lenguajes epistémicos consiste, como mínimo, en agregar a los lenguajes en cuestión un conjunto de operadores, lo que da lugar a un conjunto adicional de fbf's. Digo 'como mínimo' porque existen lenguajes formales epistémicos a los que se agregan otros símbolos además de los operadores. *Cf.*, por ejemplo, el lenguaje epistémico que desarrolla N. Rescher en *Epistemic Logic…* (ver la primera nota al pie del presente trabajo). De cualquier forma, en los lenguajes epistémicos se puede notar que la lógica epistémica es una extensión de la lógica proposicional, de primer orden o de órdenes superiores.

Una vez definido un lenguaje formal epistémico, la LE puede desarrollarse en dos direcciones de análisis lógico de los razonamientos en torno al conocimiento: una, que llamaremos *sintáctica*, y la otra, conocida como *semántica-formal.* Veamos brevemente en qué consiste cada una de ellas.

1.3. Desarrollo sintáctico de la Lógica Epistémica

Dado un lenguaje formal epistémico $\mathcal{L}$, se puede estudiar la LE desde un punto de vista sintáctico al definir un sistema o teoría en $\mathcal{L}$, el cual puede ser al menos de dos tipos, un sistema de deducción natural, o bien, un sistema axiomático. En ambos casos, un sistema consta de lo siguiente:

I) Una noción de lo que cuenta como una demostración en el sistema.

II) Una noción de teorema (de lo que significa *ser demostrable en el sistema*).

III) Un conjunto de reglas de inferencia (o reglas de producción de fórmulas a partir de fórmulas anteriores).

En el caso de un sistema axiomático, se debe agregar también:

IV) Un conjunto de axiomas (que es un subconjunto del conjunto de *fbf*'s de $\mathcal{L}$).

El concepto central en el estudio sintáctico de la LE es el de *demostrabilidad*, y dicho estudio se puede desarrollar con diferentes objetivos. Por ejemplo, en caso de que ya se conozcan *todas las verdades lógicas* de la LE, un análisis sintáctico podría tener como objeto la formulación de un sistema *correcto* y *completo* para la LE, *i. e.*, un sistema en el que se demuestren todas las verdades lógicas epistémicas y sólo estas verdades, de modo que todo lo verdadero epistémicamente sea demostrable en el sistema y viceversa. Por otro lado, también se podría emprender un análisis sintáctico únicamente con la finalidad de explorar las implicaciones inferenciales, así como las consecuencias de asumir ciertas opiniones en torno al conocimiento, en cuyo caso con los axiomas o reglas de inferencia del sistema en cuestión se buscaría expresar o capturar tales opiniones. A manera de ejemplo, definamos un lenguaje proposicional epistémico y un sistema axiomático en dicho lenguaje.

Definimos el lenguaje proposicional epistémico de n-agentes ($\mathcal{L}$-{K_1, ..., K_n}), *i. e.*, el lenguaje proposicional epistémico cuyo conjunto de operadores es K_1, ..., K_n, para algún número natural n).

Símbolos de $\mathcal{L}$-{K_1, ..., K_n}:

I) Un conjunto infinito numerable de letras proposicionales
p_1, p_2, p_3, ...

II) ¬, →

III) (,)

IV) Un conjunto de n operadores epistémicos
K_1, ..., K_n

Fbf´s de $\mathcal{L}$-{K_1, ..., K_n}:

I) Toda letra proposicional es *fbf* de $\mathcal{L}$.

II) Si α y β son *fbf ´s*, también lo son:
$(\neg\alpha)$, $(\alpha\rightarrow\beta)$, $(K_j\alpha)$ con j=1, 2, ..., n.

III) No hay más *fbf ´s*.

En este lenguaje, definimos el sistema axiomático H[13]:

[13] El sistema H figura en la breve introducción a la Lógica Epistémica que hace Max Freund en el opúsculo al que se ha hecho referencia anteriormente (ver nota al número 2). Allí, el autor atribuye el sistema H a Jaakko Hintikka, específicamente en su obra clásica de 1962. Sin embargo, esto es un error, ya que la lógica epistémica que presenta Hintikka en *Knowledge and belief...* concuerda mucho mejor con el sistema S4, un sistema de Lógica Epistémica que analizaremos en el capítulo IV y que no incluye, a diferencia de H, el axioma A5.

Definición de demostración:

Una demostración en H es una lista finita de *fbf*'s de L-{K_1, ..., K_n}, tal que en cada una de sus líneas aparece un axioma de H o una *fbf* obtenida de fórmulas que aparecen en líneas anteriores por aplicación de alguna regla de inferencia de H.

Definición de teorema:

Se considera un teorema de H a la *fbf* que aparece en la última línea de una demostración.

Axiomas de H[14]

Sean α y β *fbf*'s cualesquiera de $\mathcal{L}$-{K_1, ..., K_n} y sea j=1, ..., n.

P: Todas las instancias de tautologías proposicionales [*Propiedad de tautología*].
K: $K_j(\alpha\rightarrow\beta)\rightarrow(K_j\alpha\rightarrow K_j\beta)$ [Propiedad de distribución].
T: $K_j\alpha\rightarrow\alpha$ [Propiedad de conocimiento].
A4: $K_j\alpha\rightarrow K_jK_j\alpha$ [Propiedad de introspección positiva].
A5: $\neg K_j\alpha\rightarrow K_j\neg K_j\alpha$ [Propiedad de introspección negativa].

Reglas de inferencia de H

Sean α y β *fbf*'s cualesquiera de $\mathcal{L}$-{K_1, ..., K_n}.

MP: Si $\vdash\alpha$ y $\vdash\alpha\rightarrow\beta$, entonces $\vdash\beta$ [*Modus ponens*] y
KG: Si $\vdash\alpha$, entonces $\vdash K_i\alpha$ [Generalización del conocimiento].

El lenguaje $\mathcal{L}$-{K_1, ..., K_n} constituye un lenguaje puramente formal y, por tanto, los axiomas de H carecen de significado. No obstante, si se interpretan las fórmulas del tipo << $K_i\sigma$>> como <<el agente i sabe que σ>>, se caerá en cuenta que mientras P y MP son elementos de la lógica clásica, los axiomas K-A5, así como la regla KG, expresan

[14] A continuación, se enlistan los cinco axiomas de H y sus dos reglas de inferencia, cada uno con su abreviación y nombre usual. En los capítulos subsecuentes, cada uno de los axiomas, desde K hasta A5, nos servirá para definir distintos sistemas axiomáticos para la LE, cuyas características lógicas esclareceremos.

opiniones o puntos de vista en torno al conocimiento, más precisamente, en torno al conocimiento proposicional. Por ejemplo, con K se afirma que si un agente conoce una proposición y sabe que tal proposición implica lógicamente a otra, entonces dicho agente conoce también esta última proposición; T pone de manifiesto la idea de que un componente esencial del saber es la verdad, mientras que A4 y A5 son axiomas que afirman que el conocimiento y la ignorancia son reflexivos, *i. e.*, que si un agente conoce una proposición, entonces él sabe que la conoce, y en caso de que la ignore, él sabe que la ignora. Por otro lado, la regla KG nos dice que los agentes conocen todas las proposiciones que tienen demostración o prueba.

Se ha discutido ampliamente si el sistema H es adecuado para capturar formalmente el conocimiento proposicional en seres humanos ya que involucra una concepción del agente como un conocedor ideal, esto es, sin limitaciones de razonamiento, así como de carácter espacio-temporal. Por ejemplo, KG, como se acaba de mencionar, obliga a pensar al agente del conocimiento como capaz de conocer todas las proposiciones demostrables; K y KG, tomados en conjunto, llevan a la conclusión de que el agente conoce todas las consecuencias lógicas de las proposiciones por él sabidas; A4 y A5, tomados en conjunto, conducen a la conclusión de que los agentes conocen un número infinito de proposiciones. Todo esto, por supuesto, es incompatible con una interpretación de los agentes de conocimiento como seres humanos[15].

1.4. Desarrollo semántico-formal de la Lógica Epistémica

Para desarrollar la LE desde un punto de vista semántico, se debe partir, como en el punto de vista sintáctico, de un lenguaje formal epistémico $\mathcal{L}$, y ofrecer una definición de interpretación y una noción de verdad para dicho lenguaje. Lo que es una interpretación para $\mathcal{L}$ se puede definir en términos de una *estructura de Kripke*, para

[15] Por el momento, no se tomará posición alguna respecto a la adecuación de la Lógica Epistémica Modal para modelar el conocimiento humano. Sin embargo, el lector podrá hallar algunas consideraciones al respecto en las conclusiones finales de este trabajo.

después, apoyándose en esta estructura, dar una noción de verdad para cada *fbf* de $\mathcal{L}$.

Como se puede notar, los conceptos centrales en el desarrollo semántico de la LE son los de *interpretación* y *verdad*, y tiene por objetivo el conocimiento de las verdades lógicas epistémicas. Como ejemplo, consideremos una vez más el lenguaje formal epistémico L-{K_1, ..., K_n}, para algún número natural n, y definamos una estructura de Kripke (una interpretación) así como una noción de verdad para tal lenguaje.

Una estructura de Kripke, *i. e.*, una interpretación, para $\mathcal{L}$-{K_1, ..., K_n} es un tuplo:

$\mathcal{E} = \langle S, \pi, R_1, \ldots, R_n \rangle$[16].

Donde:

1) $S \neq \emptyset$ S es un conjunto no vacío (llamado el conjunto de estados s de $\mathcal{E}$).

2) $\pi: L \longrightarrow P(S)$ π es una función que va de L al conjunto potencia de S, donde L es el conjunto de letras proposicionales de $\mathcal{L}$-{K_1, ..., K_n} ($\pi(p_i) \subseteq S$).

3) Para cada j=1, ..., n, $R_j \subseteq S^2$ (R_j se llama la relación de accesibilidad o posibilidad de j).

Así definida una estructura o interpretación, al lenguaje epistémico $\mathcal{L}$-{K_1, ..., K_n} le corresponde una familia de estructuras, pero, como iremos viendo a lo largo de este trabajo, tal familia puede cambiar si imponemos ciertas restricciones a las relaciones binarias que son parte de las estructuras (R_1, ..., R_n).

Demos ahora una noción de verdad para nuestro lenguaje, *i. e.*, definamos lo que quiere decir que una *fbf* de $\mathcal{L}$-{K_1, ..., K_n} sea verdadera en una interpretación o estructura.

---

[16] El número de las R's en la estructura corresponde exactamente al número de operadores en el lenguaje (en este caso, las K's). Es decir, el conjunto de índices para las relaciones de accesibilidad y para los operadores de L-{K_1, ..., K_n} será el mismo.

Definimos primero recursivamente, para cada $s \in S$ y cada *fbf* σ de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$, lo que significa:

$\mathcal{E}$, $s \vDash \sigma$ (se lee: la fórmula σ es verdadera (realizada o satisfecha) en el estado s bajo la interpretación (estructura) $\mathcal{E}$).

Cláusula 1 [*Cláusula básica*]: σ es una letra proposicional.
Por lo tanto, $\sigma \in L$ (σ es una letra proposicional p_i).
$\mathcal{E}$, $s \vDash p_i$ sii $s \in \pi(p_i)$.
Cláusula 2 [*Cláusula de la negación*]: $\mathcal{E}$, $s \vDash \neg\sigma$ sii $\mathcal{E}$, $s \nvDash \sigma$.
Cláusula 3 [*Cláusula de la implicación*]: $\mathcal{E}$, $s \vDash \alpha \rightarrow \beta$ sii $\mathcal{E}$, $s \nvDash \alpha$ o $\mathcal{E}$, $s \vDash \beta$.
Cláusula 4 [*Cláusula modal*]: Para $j=1, \ldots, n$, $\mathcal{E}$, $s \vDash K_j\sigma$ sii $\mathcal{E}$, $s' \vDash \sigma$ para todo s' tal que $\langle s,s' \rangle \in R_j$.

A partir de lo anterior, podemos dar las siguientes dos definiciones, ambas de vital importancia desde el punto de vista semántico

Sea σ cualquier *fbf* de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$.

Definición de verdad: σ es verdadera en una estructura $\mathcal{E}$ sii para todo estado s de $\mathcal{E}$ se tiene que $\mathcal{E}$, $s \vDash \sigma$.

Notación: $\mathcal{E} \vDash \sigma$

Definición de validez: σ es válida (verdad lógica epistémica) sii para toda estructura $\mathcal{E}$ se tiene que $\mathcal{E} \vDash \sigma$

Notación: $\vDash \sigma$

Si bien sintáctica y semántica corresponden a dos enfoques distintos en el desarrollo de la LE, en este campo se tiene como un *desiderátum* la elaboración de un sistema formal en el que se demuestre única y exclusivamente el conjunto de fórmulas válidas para el lenguaje formal epistémico en cuestión, de tal forma que lo demostrable en el sistema y lo semánticamente válido lleguen a corresponder con exactitud. Sin embargo, a diferencia de lo que ocurre con la lógica clásica, para los lenguajes de LE no se cuenta con una única semántica y, por tanto, con único conjunto de fórmulas válidas que habría que demostrar en un sistema; ello se explica, en parte, por la ausencia de una comprensión diáfana del significado de conceptos epistémicos fundamentales como *creencia, saber, justificación*, etcétera, que se pudiera rescatar en una semántica de aceptación general. *Ipso facto*, en esta lógica, la perspectiva sintáctica ha presentado un mayor desarrollo en comparación con la semántica, mientras que esta última se ha ido construyendo, en gran medida, a partir de los diversos intentos por rescatar los teoremas de los distintos sistemas formales, de ahí la pluralidad de las semánticas[17].

En los dos capítulos subsecuentes, se analizarán algunos de los sistemas axiomáticos más importantes para la LE, así como la semántica correspondiente a cada uno de ellos.

1.5. Consideraciones en torno a la semántica anterior

En conjunto, la noción de *interpretación* y la noción de *verdad* anteriores conforman lo que se conoce como *Semántica de Kripke* o *Semántica de Mundos Posibles*, la cual ha resultado ser adecuada para el estudio formal de algunas modalidades, *v. g.,* las modalidades aléticas (lo necesario, lo posible y lo contingente) y deónticas (lo obligatorio, lo permitido y lo prohibido)[18]. Lo anterior significa, entre otras cosas, que esta semántica rescata, bajo la forma de verdades lógicas, algunas de nuestras intuiciones fundamentales acerca de las modalidades en cuestión. Además, en el caso de la Lógica Alética, cuando los operadores modales O_j se interpretan como operadores de necesidad, la cláusula modal de nuestra definición de verdad [Para $j=1, \ldots, n$, $\mathcal{E}, s \vDash O_j\sigma$ sii $\mathcal{E}, s' \vDash \sigma$ para todo s' tal que $\langle s,s' \rangle \in R_j$] adquiere una lectura francamente intuitiva: en este caso, los mundos que son accesibles o visibles para el agente *j* desde un mundo *s* constituyen el conjunto de mundos que *j* considera posibles cuando mira desde *s*; en consecuencia, naturalmente, una proposición α será necesaria para *j* en el mundo *s*

[17] *Cf.* Freund, Max, *Lógica epistémica*..., p. 213.

[18] De hecho, antes de Saul Kripke, esta semántica parece haber sido desarrollada por Stig Kanger y Jaakko Hintikka.

siempre y cuando dicha proposición sea verdadera en todos los mundos a los que éste tiene acceso desde *s*. Es justo esta adecuación, con diversas modalidades, lo que motiva el empleo de la semántica de Kripke para el análisis lógico de las modalidades epistémicas. No obstante, cabe la pregunta directa acerca de estas últimas: ¿es la semántica de Kripke en efecto adecuada para el estudio de las modalidades epistémicas?, ¿disponemos de una lectura intuitiva de la cláusula modal de nuestra definición de verdad en el caso de la LE cuando los operadores modales se interpretan como operadores de conocimiento? Intentaré responder ahora a esto último[19]. En realidad, la segunda interrogante equivale a preguntar acerca de la intuición sobre el conocimiento que se halla detrás de la cláusula modal. Una respuesta la podemos ver esbozada en las líneas siguientes:

> Si tal significado intuitivo aplica al operador de conocimiento es una cuestión de opinión. Asumamos que un estado *t* es accesible desde el estado *s* en una estructura *M*. Esto podría significar que, sobre la base de la información disponible para el "agente epistémico" en el estado *s*, él o ella no puede descartar al estado *t* como un candidato a ser el estado "real". De modo que, si el agente sabe *p* desde el estado *s*, entonces *p* debe ser verdadera en *t* en cualquier estado accesible desde *s* (incluido *s* probablemente). Por el contrario, el agente no sabe *p* desde *s* si *p* resulta ser falsa en algún estado accesible desde *s*, por ejemplo, *t* o *s*[20].

[19] En cuanto a la pregunta más general sobre la adecuación de la semántica de Kripke para el estudio de la LE, únicamente dejaré anotadas en este momento dos observaciones. Primero, que haya una lectura intuitiva de la cláusula modal para el caso de esta lógica [intentaré mostrar que la hay] brinda un apoyo racional, aunque mínimo, a tal adecuación. Segundo, una contestación más cabal de este problema se puede obtener mediante los resultados de aplicar la semántica de Kripke a la LE. Es decir, esta semántica dividirá las *fbf's* de nuestro lenguaje epistémico en dos conjuntos, *viz.,* las fórmulas que son lógicamente válidas y aquellas que no lo son. Si la división es intuitivamente aceptable, entonces diremos que la semántica también lo es; en caso contrario, juzgaremos que la semántica no es apropiada para nuestro estudio (cf. Gochet P. y Gribomont P., "Epistemic logic"…, p. 101). La evaluación de la semántica de Kripke a partir de sus resultados se desarrollará explícitamente en las conclusiones del presente trabajo.

[20] Gochet P. y Gribomont P., "Epistemic logic"…, p. 101. Repárese en que esto sólo explica uno de los dos condicionales involucrados en nuestra cláusula.

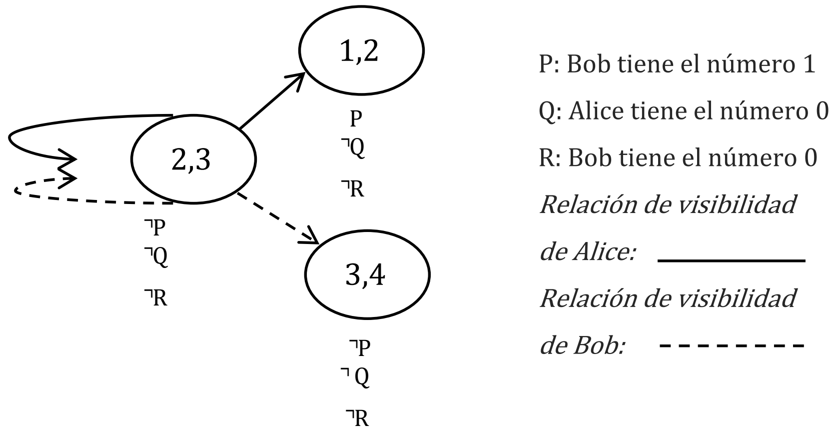
Este párrafo nos deja ver que los mundos posibles en las estructuras de Kripke no son únicamente asignaciones de valores de verdad a las letras proposicionales, sino también estados informacionales, *i. e.*, en cada uno de ellos los agentes poseen cierta información, y con base en ella tienen acceso o les son visibles ciertos mundos, justamente aquellos mundos posibles que no pueden descartar como candidatos al mundo real a partir de la información de que disponen en el mundo en cuestión. Por consiguiente, resulta razonable afirmar que los agentes conocen una proposición sólo si ésta es verdadera en todos los mundos a los que ellos tienen acceso. Con todo, es ésta una explicación parcial de la cláusula modal para el caso de la LE. Para una explicación más robusta, considérese la siguiente situación:

> *Decimos a Alice y a Bob que serán informados, cada uno en secreto, de dos números naturales consecutivos (incluido el 0). Alice recibirá el número par y Bob el impar. Todo esto es conocimiento común. Entonces, Alice recibe en secreto el número 2 y Bob el 3*[21]*.*

En esta situación, Alice, que tiene el número 2, concibe dos posibilidades, de entre las cuales no puede distinguir la correspondiente a la realidad: una posibilidad en la que Bob tiene el número 1 y una posibilidad en la que Bob tiene el número 3. Por su parte, Bob, con el número 3, concibe a su vez dos mundos posibles, y tampoco él puede discernir cuál de éstos es el mundo real: un mundo posible en el que Alice tiene el número 2 y otro en el que Alice tiene el número 4. Estas posibilidades que nuestros dos agentes conciben en función del número que poseen constituyen el conjunto de mundos

[21] Lo anterior es parte de un acertijo cuya continuación, irrelevante para la explicación en curso, sería:
Ambos mantienen el siguiente diálogo.
A: No sé tu número.
B: Tampoco sé tu número.
A: Ahora sé cuál es tu número.
B: Yo también sé cuál es tu número ahora.
¿Cómo es esto posible?

posibles que les son accesibles desde un mundo a partir de la información de la que disponen en ese mismo mundo. Veamos esto en un diagrama:



En el diagrama figuran tres mundos posibles, entre los que se encuentra el mundo "real" (el mundo que aparece a la izquierda). Note que ambos, Bob y Alice, lo conciben como una de dos posibilidades, pero, dada la información de que disponen en él, no pueden identificarlo como tal, ya que no pueden distinguir cuál de estas dos posibilidades constituye el mundo real. Por otro lado, he asociado a cada mundo posible tres proposiciones que resultan verdaderas en éste, por lo que será relativamente fácil darse cuenta de lo siguiente: Alice *no sabe* que Bob tiene el número 1 pues, según la información que posee (ella tiene el número 2), existe la posibilidad de que Bob tenga el 3 y no el 1; en contraste, Bob *sabe* que él no tiene el número 1 pues, de acuerdo a la información con la que cuenta (él posee el número 3), no importa qué número tenga Alice, el 2 ó el 4, seguirá siendo verdad que él no tiene el 1 sino el 3. Por último, nótese que tanto Alice como Bob *saben* que el otro no tiene el número 0[22].

Este ejemplo, tal como fue señalado en una nota al pie, proviene de un acertijo –conocido como el acertijo de los números consecutivos–, razón por la que presenta un aire un tanto atípico y artificial; aun así, es suficiente para mostrar que existen situaciones, por infrecuentes que parezcan, en las que se emplea el vocablo 'saber' justo

[22] Podríamos agregar también, por ejemplo, que Alice no sabe que Bob sabe que ella no tiene el 0, pues ésta concibe una posibilidad en la que Bob no sabe que ella no posee el 0, a saber, en caso de que Bob tuviera el número 1.

en el sentido capturado por la cláusula modal de la Lógica Epistémica: un agente sabe una proposición siempre y cuando ésta sea verdadera en todas las posibilidades que él concibe con base en determinada información, *i. e.*, en todos los mundos posibles a los que puede ver o acceder. Por otra parte, el ejemplo es útil también para explorar una interpretación específica de esta cláusula, que la tornará un poco más intuitiva o, por lo menos, más cercana a la discusión filosófica en torno al conocimiento.

En nuestro ejemplo, los agentes saben aquello de lo que pueden estar completamente seguros o ciertos, *v. g.*, Bob tiene completa certeza de que Alice no posee el 0: él recibió el 3 y desde un principio se le dijo que los números que recibirían serían consecutivos. Con fundamento en esta certeza, él sabe que Alice no tiene el 0. No sabe, en cambio, si Alice posee o no el número 2, pues no tiene certeza al respecto: como él cuenta con el 3, ella bien podría tener el 2 ó el 4, puesto que ambos son pares y satisfarían el requisito de que se trate de números consecutivos. En otros términos, este ejemplo sugiere la siguiente interpretación de nuestra cláusula modal:

Sean S y α un sujeto y una proposición cualesquiera, respectivamente.

C: S sabe que α sii S tiene completa certeza sobre α.

Para esclarecer semejante interpretación, habrá que distinguir entre certeza psicológica y certeza epistémica[23]. La primera tiene que ver con la fuerza de nuestras convicciones; así, tener completa certeza psicológica acerca de una proposición significa estar absolutamente convencido de su verdad (tanto como nos sea posible); es estar persuadido de su verdad a un grado tal que su falsedad nos resulta inconcebible. A la luz de esta forma de certeza, la interpretación **C** carece en gran medida de interés epistémico, ya que con ella se diría simplemente que los agentes que saben una proposición sii se *sienten* completamente ciertos de su verdad (*i. e.*, sii se hallan tan convencidos de que es verdadera que no pueden aceptar la posibilidad de que sea falsa). Pero esta forma de certeza poco o nada tiene que ver con el conocimiento,

[23] La idea básica de estas dos formas de certeza la he tomado de; Feldman, Richard, "Skepticism (I)", *Epistemology*, New Jersey, Prentice Hall, 2003, p. 117.

y de ningún modo es condición necesaria para éste –a no ser, desde luego, que por 'estar completamente cierto' se haga referencia al mero acto de creer, en cuyo caso sería una condición necesaria para saber, mas no suficiente–. Es verdad, en el caso de Alice y Bob podemos atribuirles una sensación de plena seguridad respecto a las proposiciones por ellos conocidas; sin embargo, esta sensación no está implicada en su saber ni lo garantiza. La cuestión no es, por tanto, si los agentes experimentan o no una certeza psicológica respecto a ciertas proposiciones, sino si esa certeza se encuentra o no plenamente justificada, *i. e.*, fundada en razones que la legitiman.

La segunda forma de certeza, la certeza epistémica, está determinada por la fuerza o solidez de nuestros fundamentos para creer. Así, tener completa certeza epistémica de una proposición significa que nuestros fundamentos o razones para creerla son máximamente sólidos, donde esto último quiere decir *razones absolutamente concluyentes* o *evidencia infalible*[24]. Considerando esta forma de certeza, la interpretación **C** de la cláusula modal se leería como sigue:

> **C**$_e$: S sabe que α sii S tiene completa certeza epistémica respecto de α [*i. e.*, sii las razones de S para creer α son absolutamente concluyentes][25].

Es esta lectura epistémica, y no la psicológica, la que interesa en el presente análisis. Por tanto, habrá que preguntarse, primero, si en nuestro ejemplo Alice y Bob cuentan con una certeza semejante y, segundo, si la cláusula modal admite ser leída desde la certeza epistémica. Con el objeto de responder a estos cuestionamientos, debe tenerse en cuenta que hemos caracterizado a la certeza epistémica completa a partir de la noción de *razones o evidencia infalible* (=razones absolutamente concluyentes), y que por ésta entendemos aquella evidencia que es *incompatible* con la falsedad de la

---

[24] La idea es que la certeza de tipo epistémico requiere que nuestras razones para creer sean tan sólidas como sea posible; pero 'posible', en este contexto, tiene por lo menos dos sentidos: hace referencia al grado máximo de justificación que los seres humanos, de hecho, somos capaces de alcanzar, o bien, refiere al grado máximo de justificación imaginable, con independencia de si podemos o no satisfacerlo. Nosotros estamos considerando aquí este último sentido.

[25] Como veremos al final del capítulo, esta lectura de la cláusula modal nos permitirá conectarla con una forma de escepticismo.

proposición en cuestión. En consecuencia, si un agente cree alguna proposición sobre la base de evidencia infalible, entonces tiene derecho, dado que ésta es incompatible con la falsedad de la proposición, a descartar cualquier posibilidad de que la proposición falle[26]. Consideremos ahora nuestro ejemplo a la luz de esta noción.

En él, dijimos, Bob sabe que su número es el 3 y que el número de Alice no es 0. No sabe, por el contrario, si el número de Alice es el 2 ó el 4. Estas atribuciones de conocimiento e ignorancia no son particularmente problemáticas, ya que, en última instancia, resultan necesarias para solucionar el acertijo de donde proviene nuestro ejemplo[27]. Bien, ¿diremos en este caso que Bob posee completa certeza epistémica de las primeras dos proposiciones, mientras que carece de ella en cuanto a la tercera se refiere? Aparentemente, no es así, pues su evidencia a favor de que su número es 3 y el de Alice no es 0 se limita a la información que se le proporcionó desde un inicio (se le dijo que su número era el 3 y que su número y el de Alice serían números naturales consecutivos), así como a las conclusiones que él va extrayendo a partir de dicha información. No obstante, ¿cómo esta evidencia –o cualquier otra– podría descartar por completo la posibilidad de error? Por ejemplo, él pudo haber oído mal el número que se le proporcionó, o simplemente no recordarlo correctamente; además, podría cometer errores de razonamiento o no comprender bien el concepto de *números consecutivos*. Su evidencia, entonces, parece dejar un hueco para el error y, en consecuencia, no sería evidencia infalible.

Sin embargo, todo esto es sólo aparente; para notarlo basta con recordar, una vez más, que nuestro ejemplo proviene de un acertijo lógico. En él, como en muchos otros acertijos, se presuponen muchas cosas para su solución. Primero, se parte del supuesto de que los agentes involucrados son seres ideales cognitivamente, *i. e.*, razonan correctamente y no tienen fallos lingüísticos, de memoria o perceptuales. En otras palabras, el ejemplo, por su construcción misma, proscribe estas posibilidades de

---

[26] Nótese que la evidencia infalible, como su nombre sugiere, no sólo destierra la posibilidad de error de la proposición creída, sino que ella misma es inmune al error.

[27] Ver nota número 21 al pie, en este capítulo.

error. En segundo lugar, el acertijo también presupone que quien proporciona la información inicial a nuestros agentes es máximamente confiable, *i. e.*, nos les da información falsa y su simple acto de afirmar garantiza la verdad de lo dicho. Esto último no resulta tan controversial como podría parecer en un principio, ya que es esta persona (o grupo de personas) quien determina las condiciones iniciales en la situación que describe el ejemplo. Así, cuando se le dice a Bob que su número es el 3, este acto constituye, *stricto sensu*, un acto de habla, pues con él no sólo se afirma que su número será el 3, sino que se le *asigna* dicho número[28]. Por consiguiente, la afirmación resulta verdadera en el acto mismo. Por lo tanto, en nuestro ejemplo, dada su construcción o diseño, nuestros agentes poseen completa certeza epistémica (evidencia infalible) acerca de algunas proposiciones: Bob está completamente cierto de tener el 3, pues se le dijo que ése era su número y esto es suficiente, bajo las condiciones recién expuestas, para descartar racionalmente cualquier posibilidad de error al respecto; del mismo modo se explica su certeza en cuanto a que su número y el de Alice son consecutivos. De estas dos proposiciones Bob deduce que el número de Alicia no puede ser el 0, por lo que tiene derecho a dejar cualquier duda de lado y, en este sentido, tiene completa certeza también de su conclusión –después de todo, razonó correctamente a partir de proposiciones ciertas–.

En contraste, Bob no posee completa certeza epistémica sobre el número que se le proporcionó a Alice, el 2 ó el 4: no puede determinar qué número tiene Alice debido a que su evidencia es compatible con ambos números (por ello no le permite descartar por completo la posibilidad de error para ninguno de ellos.

Pero las proposiciones de las que nuestro agente está epistémicamente cierto coinciden exactamente con aquellas proposiciones que, según dijimos desde el comienzo, él conoce[29]. En suma, en el ejemplo propuesto parece valer la interpretación

[28] Para la noción misma de *acto de habla*, ver Austin, J. L., *How to Do Things With Words*, Cambridge (Mass.), 1962.

[29] Más todavía, diremos que nuestras atribuciones de conocimiento para Alice y Bob se basan, en cierto modo, en el hecho de que éstos poseen certeza epistémica en torno a determinadas proposiciones.

Ce, según la cual los agentes epistémicos saben una proposición, siempre y cuando tengan una completa certeza epistémica respecto a ella.

Abordemos ahora la cuestión de si la cláusula modal admite o no la interpretación **Ce**. Textualmente, la cláusula declara: para j=1, …, n, $\mathcal{E}$, s⊨K_jα sii $\mathcal{E}$, s'⊨α para todo s' tal que <s,s'>∈R_j [se lee " *j* sabe que α en el mundo s sii α es verdadera en todos los mundo posibles epistémicamente accesibles para *j* desde s"]. Si por "mundo epistémicamente accesible para *j*" entendemos aquellas posibilidades compatibles con la evidencia de *j*, *i. e.*, que ésta no descarta o elimina (como candidatos al mundo real), entonces leemos "*j* sabe que α en s sii α es el caso en todas las posibilidades que la evidencia de *j* (en s) no logra eliminar". Pero esto es equivalente a "… sii la evidencia de *j* (en s) elimina cualquier posibilidad en la que α es falsa. Pero ésta es justo la interpretación **Ce**. Por lo tanto, según parece, tal interpretación resulta compatible –más aún, es natural– con la cláusula modal. Además, entender así la cláusula modal de conocimiento permite conectarla con la discusión filosófica en Epistemología, particularmente, con el problema del escepticismo filosófico de *altos estándares*[30]. Considérese el siguiente argumento:

Sea *j* un sujeto arbitrario y sea α cualquier proposición creída por él.

P1) α podría ser falsa.

P2) Si α podría ser falsa, entonces *j* no sabe que α.

∴

C) *J* no sabe que α [de P1) Y P2), por MP][31].

---

[30] Se le llama *escepticismo filosófico de altos estándares* a aquél cuyos argumentos, implícita o explícitamente, exigen estándares de justificación muy elevados, prácticamente insatisfacibles. Además del argumento que aquí se presenta, en el capítulo siguiente emplearé otro razonamiento de esta forma de escepticismo con el objeto de discutir el principio K de la Lógica Epistémica.

[31] Una versión acotada de este argumento (para el escepticismo del mundo exterior), la cual emplearemos en el capítulo III, aparece en Feldman, Richard, "Skepticism (I)", *Epistemology*, New Jersey, Prentice Hall, 2003, pp. 114-115.

En este razonamiento, conocido como *El razonamiento de la posibilidad de error*, la conclusión hace referencia a cualquier proposición creída [por cualquiera], de modo que está diseñado para apoyar el escepticismo general, y no éste y otro escepticismo de carácter local. Tiene la forma de un *Modus Ponens*; se trata de un razonamiento correcto. Su primera premisa señala la posibilidad de error para cualquiera de nuestras creencias, y tiene como trasfondo las hipótesis del escepticismo cartesiano, pues éstas presentan escenarios que nos hacen desconfiar hasta de nuestras creencias más evidentes[32].

La segunda premisa constituye el corazón del argumento; con ella se declara que, si hay la posibilidad de que nuestras creencias yerren, entonces no constituyen casos de conocimiento. Comparemos esta premisa (P2) con la próxima afirmación (P2'): si α es falsa, entonces *j* no sabe que α. En ella se declara que la falsedad *efectiva* de una creencia es suficiente para que ésta no sea conocimiento. Detrás de ella se encuentra simplemente la idea de que el conocimiento requiere verdad, *i. e.*, no se puede conocer lo que es falso; algo aceptable para cualquier análisis que tiene a la verdad como un componente insoslayable del saber. Por el contrario, con P2 se afirma algo más fuerte, sujeto a controversia, pues nos dice que la simple posibilidad de falsedad, se actualice o no, es suficiente para que no se tenga conocimiento. ¿Qué intuición se encuentra detrás de P2? La idea básica parece ser la siguiente: <<si puedes estar equivocado acerca de algo, entonces no estás absolutamente cierto de ello. Si no estás absolutamente cierto de algo, entonces no lo sabes>>[33]. De manera más expresa: si α

[32] Se suele creer que las ideas evidentes son inmunes a las dudas planteadas por los escenarios escépticos, pero yo no comparto esta tesis; después de todo, como ya señala Descartes –quien es el proponente original de tal forma de escepticismo– una divinidad (o un genio maligno) terriblemente poderosa podría hacer que nos engañáramos hasta en aquello que nos parece más evidentes: <<[…] así como a veces juzgo que los demás se engañan en las cosas que mejor creen saber con mayor certeza, ¿qué sé yo si Dios no ha querido que yo también me engañe cuando adiciono dos y tres, o enumero los lados de un cuadrado, o cuando juzgo de cosas más fáciles que ésas, si es que puede imaginarse algo que sea más fácil?>> (*AT.* VII, 21).

[33] Feldman, Richard, "Skepticism (I)", *Epistemology*, New Jersey, Prentice Hall, 2003, p. 124.

podría ser falsa a pesar de nuestra evidencia (*i. e.*, si ésta no es absolutamente concluyente), entonces tal evidencia no nos justifica para creer α (justo porque no elimina la posibilidad de error al respecto) y, en consecuencia, no nos permite conocer α. Pero si esto es así, P2 no es sino una consecuencia de nuestra cláusula modal –entendida desde la interpretación **Ce**–, del mismo modo que P2' es una consecuencia de la propiedad de conocimiento [el saber implica verdad]. Esto sugiere fuertemente que quizá tanto la Lógica Epistémica Modal como el escepticismo de altos estándares parten, en el fondo, del mismo concepto de saber: uno que hace del saber algo infalible. En suma, la interpretación **Ce** es una de las más fructíferas e interesantes de la cláusula modal en su versión epistémica.

2. Sistemas de Lógica Epistémica (El Sistema $\mathbb{K}$)

En este capítulo se estudiará el sistema axiomático más elemental para la lógica proposicional epistémica, el sistema $\mathbb{K}$ –mientras que en el capítulo III estudiaremos el sistema $\mathbb{T}$, y en el capítulo IV los sistemas $\mathbb{S}4$ y $\mathbb{S}5$, para esta lógica–[34]. Nuestra manera de proceder será la siguiente: primero, a partir de la definición de un lenguaje proposicional epistémico $\mathcal{L}$, definiremos el sistema en cuestión en dicho lenguaje. Esto significa dar sus axiomas y sus reglas de inferencia, así como una definición de *demostración* y de *teorema* para el mismo. Una vez hecho esto, analizaremos sus propiedades sintácticas, para luego definir la semántica correspondiente, *i. e.*, dar una noción de *interpretación* y de *verdad* para nuestro lenguaje epistémico. Por último, demostraremos el metateorema de corrección para $\mathbb{K}$ y comentaremos sus presupuestos epistemológicos.

2.1. El lenguaje proposicional epistémico $\mathcal{L}$-$\{K_1, \ldots, K_n\}$ de n-agentes

Para el estudio de los sistemas axiomáticos de que trataremos en este capítulo y en el próximo emplearemos el lenguaje proposicional epistémico $\mathcal{L}$-$\{K_1, \ldots, K_n\}$ de n-agentes que definimos en el capítulo I. Recordemos su definición:

Definición de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$

Símbolos de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$:

I) Un conjunto infinito numerable de letras proposicionales
$p_1, p_2, p_3, \ldots$

II) $\neg, \rightarrow$

---

[34] El sistema $\mathbb{K}$ fue nombrado así en honor al filósofo y lógico Saul Kripke. En realidad, tanto este sistema como los tres sistemas posteriores que se analizarán en los capítulos siguientes ($\mathbb{T}$, $\mathbb{S}4$ y $\mathbb{S}5$) surgieron en la lógica modal alética (aquella que tiene a su cargo la investigación de los conceptos de *necesidad* y *posibilidad*), y lo que aquí se expone es una versión de los mismos para la LE.

III) (,)

IV) Un conjunto de n operadores epistémicos: $K_1, \ldots, K_n$

Fbf´s de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$:

I) Toda letra proposicional es *fbf* de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$.

II) Si α y β son *fbf ´s*, también lo son:

$(\neg\alpha)$, $(\alpha\rightarrow\beta)$, $(K_j\alpha)$ con j=1, 2, …, n.

III) No hay más *fbf ´s*.

Definiciones:

$\alpha\wedge\beta \stackrel{\text{def}}{=} \neg(\alpha\rightarrow\neg\beta)$

$\alpha\vee\beta \stackrel{\text{def}}{=} (\neg\alpha\rightarrow\beta)$

$\alpha\leftrightarrow\beta \stackrel{\text{def}}{=} (\alpha\rightarrow\beta)\wedge(\beta\rightarrow\alpha)$

Una vez dado nuestro lenguaje $\mathcal{L}$-$\{K_1, \ldots, K_n\}$, podemos pasar a definir nuestro primer sistema axiomático para la LE, el sistema $\mathbb{K}$.

2.2. El sistema $\mathbb{K}$

El sistema $\mathbb{K}$ es el sistema axiomático más simple para la LE, cuyo axioma epistémico, el axioma K, goza prácticamente de una aceptación general entre los epistemólogos; no obstante, de entre los cuatro sistemas axiomáticos que mencionamos al inicio, es el de menor poder deductivo. Definamos, pues, el sistema $\mathbb{K}$:

Definimos en $\mathcal{L}$-$\{K_1, \ldots, K_n\}$ el sistema axiomático $\mathbb{K}$.

Axiomas de $\mathbb{K}$

Sean α, β y γ *fbf ´s* cualesquiera de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$ y sea j=1,… , n.

A1: $\alpha\rightarrow(\beta\rightarrow\alpha)$,

A2: $(\alpha\rightarrow(\beta\rightarrow\gamma))\rightarrow((\alpha\rightarrow\beta)\rightarrow(\alpha\rightarrow\gamma))$,

A3: $(\neg\alpha\rightarrow\neg\beta)\rightarrow((\neg\alpha\rightarrow\beta)\rightarrow\alpha)$, K: $K_j(\alpha\rightarrow\beta) \rightarrow (K_j\alpha\rightarrow K_j\beta)$ [Propiedad de distribución].

Reglas de inferencia de K

Sean α y β *fbf ´s* cualesquiera de $\mathcal{L}$-{K_1, …, K_n}.

MP: Si $\vdash^K \alpha$ Y $\vdash^K \alpha\rightarrow\beta$, entonces $\vdash^K\beta$ [*Modus ponens*] y

KG: Si $\vdash^K \alpha$, entonces $\vdash^K K_i\alpha$ [Generalización del conocimiento][35].

La definición de lo que significa demostrar en K y de lo que significa ser teorema de K son básicamente las mismas que para el sistema H. Veamos.

Definición de demostración:

Una demostración en K es una lista finita de *fbf ´s* de $\mathcal{L}$-{K_1, …, K_n}, tal que en cada una de sus líneas aparece un axioma de K o una *fbf* obtenida de fórmulas que aparecen en líneas anteriores por aplicación de alguna regla de inferencia de K.

Definición de teorema:

Se considera un teorema de K a la *fbf* que aparece en la última línea de una demostración.

Los axiomas A1-A3 rescatan la esencia de la lógica proposicional clásica. El axioma A1 nos deja saber que la implicación con la que trabajamos es una implicación material, esto es, que para que β implique a α no es necesaria ninguna conexión real entre estos enunciados: cuando α es verdadera, la implicación se cumple, sin importar quién es β o su valor de verdad. El axioma A2 es una versión del silogismo hipotético o principio de transitividad, y nos dice cómo se comporta la implicación. Por último, el axioma A3 es el principio de reducción al absurdo —si suponer la falsedad de una proposición nos lleva a una contradicción, entonces la proposición en cuestión debe ser verdadera— y, dado como está formulado, nos hace saber que no estamos en una lógica intuicionista. Estos tres axiomas, junto con la regla MP, son suficientes para tener los metateoremas de corrección y completez para la lógica proposicional clásica. Por otra

[35] Empleo la notación ‘$\vdash^K$’ para significar ‘ser teorema de K’. Esto para diferenciar los teoremas de K de los teoremas correspondientes a los otros sistemas.

parte, K y KG son, respectivamente, un axioma y una regla de inferencia puramente epistémicos. El axioma K, como el nombre 'propiedad de distribución' indica, nos dice que los operadores epistémicos K_j (con j=1, ..., n) se distribuyen con la implicación, mientras que la regla KG afirma que si se llega a un teorema en $\mathbb{K}$, entonces se le puede agregar un operador epistémico K_j a la izquierda y el resultado será un nuevo teorema de $\mathbb{K}$. Veamos una demostración en $\mathbb{K}$. Demostremos la fórmula '$(K_j(\alpha\rightarrow\beta)\wedge K_j(\beta\rightarrow\alpha))\rightarrow(K_j\alpha\leftrightarrow K_j\beta)$'[36], para ello tendremos que dar una lista de *fbf's* de $\mathcal{L}$-$\{K_1, ..., K_n\}$ que finalice en dicha fórmula y tal que en cada una de sus líneas aparece un axioma de $\mathbb{K}$, o bien, una fórmula que se sigue de anteriores por aplicación de la regla MP o KG:

T1$\vdash^{K}$ $(K_j(\alpha\rightarrow\beta)\wedge K_j(\beta\rightarrow\alpha))\rightarrow(K_j\alpha\leftrightarrow K_j\beta)$

1. $K_j(\alpha\rightarrow\beta) \rightarrow(K_j\alpha\rightarrow K_j\beta)$	*K*
2. $K_j(\beta\rightarrow\alpha) \rightarrow(K_j\beta\rightarrow K_j\alpha)$	*K (α/β, β/α)*
3. $(K_j(\alpha\rightarrow\beta)\wedge K_j(\beta\rightarrow\alpha))\rightarrow((K_j\alpha\rightarrow K_j\beta)\wedge (K_j\beta\rightarrow K_j\alpha))$	*1, 2, Dilema constructivo y MP(x2)*
4. $(K_j(\alpha\rightarrow\beta)\wedge K_j(\beta\rightarrow\alpha))\rightarrow(K_j\alpha\leftrightarrow K_j\beta)$	*3, Def.* $\leftrightarrow$

Nótese que he enumerado las líneas de la demostración y que, al final de cada renglón he introducido, en *cursivas*, la justificación de la fórmula introducida en dicho renglón. Por otra parte, en el tercer renglón de la demostración se lee lo siguiente:

3. $(K_j(\alpha\rightarrow\beta)\wedge K_j(\beta\rightarrow\alpha))\rightarrow((K_j\alpha\rightarrow K_j\beta)\wedge (K_j\beta\rightarrow K_j\alpha))$ *1, 2, Dilema constructivo y MP(x2)*.

He hecho esto para simplificar la demostración, pues de lo contrario, la demostración hubiera continuado, después del renglón 2, de la siguiente manera:

[36] Dado que los axiomas A1-A3 de $\mathbb{K}$ resultan suficientes para demostrar cualquier instancia de tautología en el sistema (por metateorema de completez para la lógica clásica), nos permitiremos introducir, en caso de que sea necesario, cualquier instancia de tautología en alguna línea de nuestras demostraciones, pues de antemano sabemos que constituyen teoremas de $\mathbb{K}$. Asimismo, cada que obtengamos un nuevo teorema, nos tomaremos la libertad de emplearlo en la demostración de teoremas subsecuentes, sin reproducir, digamos, su demostración completa, aunque podríamos reproducirla si fuese menester hacerlo.

3.$(K_j(\alpha\to\beta)\to(K_j\alpha\to K_j\beta))\to((K_j(\beta\to\alpha)\to(K_j\beta\to K_j\alpha))\to((K_j(\alpha\to\beta)\wedge K_j(\beta\to\alpha))\to$ $((K_j\alpha\to K_j\beta)\wedge(K_j\beta\to K_j\alpha)))$ *Dilema constructivo* [Se trata de la tautología '$(\alpha\to\varphi)\to((\beta\to\psi)\to((\alpha\wedge\beta)\to(\varphi\wedge\psi)))$']

4. $(K_j(\beta\to\alpha)\to(K_j\beta\to K_j\alpha))\to((K_j(\alpha\to\beta)\wedge K_j(\beta\to\alpha))\to((K_j\alpha\to K_j\beta)\wedge(K_j\beta\to K_j\alpha)))$ *1, 3 y MP*

5. $((K_j(\alpha\to\beta)\wedge K_j(\beta\to\alpha))\to((K_j\alpha\to K_j\beta)\wedge(K_j\beta\to K_j\alpha))$ *2, 4 y MP* [note que se trata del renglón 3 de nuestra demostración original]

6. $(K_j(\alpha\to\beta)\wedge K_j(\beta\to\alpha))\to(K_j\alpha\leftrightarrow K_j\beta)^1$ *3, Def. ↔.*

En lugar de ello, decidí simplificar la demostración y escribir al final del tercer renglón como justificación: *1, 2, Dilema constructivo y MP(x2)*. Esto significa que, para introducir el renglón 3, se ha utilizado el dilema constructivo y aplicado MP dos veces, empleando para ello los renglones 1 y 2 de la demostración.

Otra manera de simplificar las demostraciones consiste en el empleo de reglas derivadas de inferencia. Por ejemplo, a partir del axioma K, obtendremos una regla derivada de inferencia para $\mathbb{K}$, la regla RD1[37]:

Regla derivada 1 (RD1)

Si $\vdash^{K} \alpha\to\beta$, entonces $\vdash^{K} K_j\alpha\to K_j\beta$, para cualquier j= 1, …, n.

Con tal regla se afirma que, si se llega a un teorema con forma condicional, entonces se puede agregar un operador epistémico K_j inmediatamente a la izquierda tanto de su antecedente como de su consecuente y el resultado será un nuevo teorema de $\mathbb{K}$. Probemos esto:

Sean α y β *fbf's* cualesquiera de $\mathcal{L}$-{K_1, …, K_n} y sea j=1, …, n.

Supongamos que $\vdash^{K} \alpha\to\beta$. Entonces existe una demostración en $\mathbb{K}$ que termina en la fórmula '$\alpha\to\beta$'. Podemos entonces continuar la demostración como sigue y llegar a la fórmula '$K_j\alpha\to K_j\beta$':

----

m. $\alpha\to\beta$

[37] En realidad, tanto esta regla como cualquier regla derivada para $\mathbb{K}$ tiene aplicación en todos los sistemas axiomáticos que estudiaremos.

m+1. $K_j(\alpha\rightarrow\beta)$	*m, KG*
m+2. $K_j(\alpha\rightarrow\beta) \rightarrow (K_j\alpha\rightarrow K_j\beta)$	*K*
m+3. $K_j\alpha\rightarrow K_j\beta$	*m+1, m+2 MP*

Veamos una demostración en $\mathbb{K}$ empleando la regla RDI, la de la fórmula '$K_j(\alpha\wedge\beta)\leftrightarrow(K_j\alpha\wedge K_j\beta)$':

T2$\vdash^{K}$ $K_j(\alpha\wedge\beta)\leftrightarrow(K_j\alpha\wedge K_j\beta)$

1. $(\alpha\wedge\beta)\rightarrow\alpha$	*Simplificación*
2. $K_j(\alpha\wedge\beta)\rightarrow K_j\alpha$	*1, RD1*
3. $(\alpha\wedge\beta)\rightarrow\beta$	*Simplificación*
4. $K_j(\alpha\wedge\beta)\rightarrow K_j\beta$	*3, RD1*
5. $K_j(\alpha\wedge\beta)\rightarrow(K_j\alpha\wedge K_j\beta)$	*2, 4, Implicación y MP(x2)*
6. $\alpha\rightarrow(\beta\rightarrow(\alpha\wedge\beta))$	*Adjunción*
7. $K_j\alpha\rightarrow K_j(\beta\rightarrow(\alpha\wedge\beta))$	*6, RD1*
8. $K_j(\beta\rightarrow(\alpha\wedge\beta))\rightarrow(K_j\beta\rightarrow K_j(\alpha\wedge\beta))$	*K (α/β, β/(α∧β))*
9. $K_j\alpha\rightarrow(K_j\beta\rightarrow K_j(\alpha\wedge\beta))$	*7, 8, Transitividad y MP(x2)*
10. $(K_j\alpha\wedge K_j\beta)\rightarrow K_j(\alpha\wedge\beta)$	*9, Importación/Exportación*
11. $K_j(\alpha\wedge\beta)\leftrightarrow(K_j\alpha\wedge K_j\beta)$	*5, 10, Equivalencia y MP(x2)*

Una vez explicado todo lo anterior, veamos algunos de los teoremas y reglas derivadas de inferencia que se obtienen en el sistema $\mathbb{K}$[38]:

Teoremas de $\mathbb{K}$

T1$\vdash^{K}$ $(K_j(\alpha\rightarrow\beta)\wedge K_j(\beta\rightarrow\alpha))\rightarrow(K_j\alpha\leftrightarrow K_j\beta)$	T9$\vdash^{K}$ $K_j(\neg\alpha\rightarrow\alpha)\leftrightarrow K_j\alpha$
T2$\vdash^{K}$ $K_j(\alpha\wedge\beta)\leftrightarrow(K_j\alpha\wedge K_j\beta)$	T10$\vdash^{K}$ $K_j(\alpha\rightarrow\neg\alpha)\leftrightarrow K_{j\neg}\alpha$
T3$\vdash^{K}$ $(K_j(\alpha\rightarrow\beta)\wedge K_j(\beta\rightarrow\alpha))\leftrightarrow K_j(\alpha\leftrightarrow\beta)$	T11$\vdash^{K}$ $(K_j(\beta\rightarrow\alpha)\wedge K_j(\neg\beta\rightarrow\alpha))\leftrightarrow K_j\alpha$
T4$\vdash^{K}$ $K_j\neg(\alpha\vee\beta)\leftrightarrow (K_j\neg\alpha\wedge K_j\neg\beta)$*	T12$\vdash^{K}$ $(K_j(\alpha\rightarrow\beta)\wedge K_j(\alpha\rightarrow\neg\beta))\leftrightarrow K_j\neg\alpha$
T5$\vdash^{K}$ $\neg K_j\neg(\alpha\vee\beta)\leftrightarrow(\neg K_j\neg\alpha\vee\neg K_j\neg\beta)$*	T13$\vdash^{K}$ $K_j\alpha\rightarrow K_j(\beta\rightarrow\alpha)$
T6$\vdash^{K}$ $K_j(\alpha\rightarrow\beta)\rightarrow(\neg K_j\neg\alpha\rightarrow\neg K_j\neg\beta)$*	T14$\vdash^{K}$ $Kj\neg\alpha\rightarrow K_j(\alpha\rightarrow\beta)$

[38] Para la demostración de estos teoremas y reglas de inferencia ver el Apéndice I.

T7 $\vdash^{K} (K_j\alpha \vee K_j\beta) \rightarrow K_j(\alpha \vee \beta)$	T15 $\vdash^{K} K_j\alpha \rightarrow (\neg K_j\neg\beta \rightarrow \neg K_j\neg(\alpha \wedge \beta))$
T8 $\vdash^{K} \neg K_j\neg(\alpha \wedge \beta) \leftrightarrow (\neg K_j\neg\alpha \wedge \neg K_j\neg\beta)$*	

Reglas derivadas de inferencia para $\mathbb{K}$:

Regla derivada 1 (RD1)	Regla derivada 3 (RD3)
Si $\vdash^{K} \alpha \rightarrow \beta$, entonces $\vdash^{K} K_j\alpha \rightarrow K_j\beta$.	Si $\vdash^{K} \alpha \rightarrow \beta$, entonces $\vdash^{K} \neg K_j\neg\alpha \rightarrow \neg K_j\neg\beta$.
Regla derivada 2 (RD2)	Regla derivada 4 (RD4)
Si $\vdash^{K} \alpha \leftrightarrow \beta$, entonces $\vdash^{K} K_j\alpha \leftrightarrow K_j\beta$.	Si $\vdash^{K} \alpha$, entonces $\vdash^{K} \neg K_j\neg\beta \rightarrow \neg K_j\neg(\alpha \wedge \beta)$.

Respecto al significado de los teoremas de $\mathbb{K}$, nos encontramos en la misma situación que con el sistema $\mathbb{H}$: al estar definido en el lenguaje $\mathcal{L}$-$\{K_1, \ldots, K_n\}$, un lenguaje puramente formal, los teoremas del sistema $\mathbb{K}$, *stricto sensu,* no poseen significado alguno. Sin embargo, cuando interpretamos las fórmulas de la forma << $K_i\sigma$>> como <<el agente i sabe que σ>>, se hace patente que los teoremas anteriores constituyen, en general, afirmaciones ampliamente aceptadas sobre el conocimiento proposicional. T1 afirma que si un agente sabe que dos proposiciones se implican mutuamente y conoce, además, cualquiera de estas dos proposiciones, entonces también conoce la otra proposición en cuestión. T2 nos dice que los operadores K_j se distribuyen mediante la conjunción, pero también declara que un agente conoce una conjunción si y sólo si conoce ambos conyuntos. T3 es una versión epistémica del principio de *equivalencia material*, pues afirma que sabemos que dos proposiciones se implican entre sí siempre y cuando conozcamos el bicondicional de ambas proposiciones. T4 y T5 declaran, básicamente, que sabemos que una disyunción es falsa si y sólo si sabemos que ambos disyuntos son falsos. T6 constituye una afirmación de gran interés ya que se trata de una versión del *Modus Tollens* (*MT*), al tiempo que es un principio de amplio uso en casi cualquier investigación racional. Se trata del siguiente principio: si sabemos que una proposición implica otra e ignoramos que la primera es falsa, entonces también ignoramos que la segunda lo es[39]; dicho de otra forma, si conocemos que una proposición implica otra y sabemos que esta última es falsa, entonces también sabemos que la primera proposición lo es. T7 nos dice que conocemos una disyunción siempre que conocemos alguno de sus disyuntos, mientras que T8 afirma que ignoramos que una conjunción es falsa si y sólo si ignoramos la falsedad de cada uno de sus conyuntos. En cuanto a los teoremas T9 a T14, note que se trata de versiones epistémicas de las paradojas de la implicación material. De entre estos teoremas, vale la pena resaltar que T12 es un equivalente epistémico del principio de *reducción al absurdo*. Por último, T15 declara que, para conocer la falsedad de una conjunción, respecto de la cual sabemos que uno de sus conyuntos es verdadero, es necesario conocer que el conyunto restante es falso[40].

Al margen de la interpretación anterior de los teoremas, cabe plantear algunas interrogantes respecto al poder deductivo o demostrativo del sistema $\mathbb{K}$, esto es,

---

[39] En este contexto, cuando decimos 'S ignora o no conoce que la proposición α es falsa' no se quiere decir con ello que α sea falsa y S ignore dicha falsedad, sino simplemente que S no sabe que α sea falsa, sin que ello involucre ninguna afirmación respecto al valor veritativo de α.

[40] En cuanto a las reglas de inferencia derivadas, su interpretación sería la siguiente. Con RD1 se afirma que si es posible demostrar que una proposición implica otra y los agentes epistémicos conocen la primera de estas proposiciones, entonces también conocen la segunda, la proposición implicada. Dado que en el sistema $\mathbb{K}$ son demostrables todas las instancias de tautologías proposicionales, esta regla derivada declara que los agentes epistémicos conocen todas las consecuencias lógicas de las proposiciones por ellos conocidas. RD2 afirma que si es demostrable la equivalencia de dos proposiciones, entonces los agentes epistémicos conocen cualquiera de estas dos proposiciones si y sólo si conocen la otra. RD3 nos dice que si una implicación es demostrable, entonces basta con que los agentes epistémicos ignoren que el antecedente de dicha implicación es falso para que ignoren también la falsedad del consecuente. RD4 afirma, por último, que si una proposición tiene demostración, entonces los agentes epistémicos ignoran que la conjunción de esta proposición con cualquier otra es falsa a menos que sepan que esta última proposición es falsa. Téngase en cuenta que mientras la interpretación de los axiomas da como resultado afirmaciones sobre el conocimiento proposicional más o menos plausibles, la interpretación de las reglas de inferencia derivadas, por el contrario, resulta en afirmaciones contraintuitivas si se piensa en el conocimiento proposicional de los seres humanos. Quizá esto es así porque en la prueba de cada una de estas cuatro reglas interviene la regla de inferencia KG que, como vimos en el capítulo I, es en sí misma contraintuitiva (ver Apéndice I).

respecto a qué *fbf*'s de nuestro lenguaje epistémico son demostrables o no a partir de los axiomas y reglas de inferencia de $\mathbb{K}$. Podemos inquirir, por ejemplo, si en $\mathbb{K}$ se demuestra lo que hemos llamado la *propiedad de conocimiento* ($K_j\alpha\rightarrow\alpha$), que equivale a afirmar que el conocimiento proposicional implica verdad, o bien, si cualquier instancia del esquema '$K_j(\alpha\vee\beta)\rightarrow(K_j\alpha\vee K_j\beta)$', que afirma que siempre que se conoce una disyunción se conoce alguno de sus disyuntos, es teorema de $\mathbb{K}$. En el primer caso, el del esquema '$K_j\alpha\rightarrow\alpha$', habrá que volver a decir que la idea de que el conocimiento proposicional implica verdad, además de intuitiva, goza de una amplia aceptación en la Epistemología contemporánea, por lo que cabría esperar que cualquier sistema epistémico "adecuado" para modelar la forma de conocimiento proposicional tenga, entre sus teoremas, a cualquier instancia del esquema '$K_j\alpha\rightarrow\alpha$', o a cualquier instancia de un algún esquema equivalente en su lenguaje. En el caso del esquema '$K_j(\alpha\vee\beta)\rightarrow(K_j\alpha\vee K_j\beta)$', la situación es contraria a la anterior. Intuitivamente, la idea de que si conocemos una disyunción, entonces conocemos uno de sus disyuntos, parece errónea; de hecho, contamos con números contraejemplos al respecto. Por ejemplo, sabemos la siguiente disyunción: que las obras fundacionales de la literatura occidental, la *Ilíada* y la *Odisea*, fueron compuestas por un poeta, al que se le dio el nombre de Homero, en dos momentos distintos de su vida, o bien, que fueron compuestas por una ráfaga de aedos durante varias generaciones; no obstante, no sabemos cuál de estas posibilidades corresponde con la realidad[41]. Por ello, podemos esperar que en un sistema epistémico adecuado no se pueda demostrar cualquier instancia del esquema en cuestión o cualquier instancia de algún otro esquema equivalente a éste. Por consiguiente, la determinación de la capacidad o poder demostrativo de un sistema de lógica epistémica resulta relevante para juzgar sobre su adecuación a la hora de modelar una forma de conocimiento. Por esta razón, indagaremos ahora en torno a la fuerza demostrativa del sistema $\mathbb{K}$, y para ello definiremos primero una semántica para

[41] Este problema forma parte de las *cuestiones homéricas*, un conjunto de preguntas en torno a la vida y obra de Homero. Ver el prólogo a la traducción de la *Ilíada* de Óscar Martínez García, Alianza (Ed.), Madrid, 2010, pp. 7-56.

el lenguaje de este sistema, el lenguaje $\mathcal{L}$-{K_1, …, K_n}; después, apoyándonos en tal semántica, estudiaremos las propiedades metalógicas del sistema, lo que nos permitirá, a larga, dar cuenta de su poder demostrativo.

2.3. Semántica formal para el sistema $\mathbb{K}$

Definir una semántica para un lenguaje formal significa dar una noción de interpretación, así como una noción de verdad para éste (una noción de lo que significa que una *fbf* del lenguaje sea verdadera bajo una interpretación). En el capítulo I, definimos una semántica para el lenguaje $\mathcal{L}$-{K_1, …, K_n}, empleando para ello *estructuras de Kripke*. Recordemos esta semántica ya que en la sección subsiguiente nos serviremos de ella con el fin de estudiar las propiedades metalógicas del sistema $\mathbb{K}$.

Definición de interpretación para $\mathcal{L}$-{K_1, …, K_n}

Una estructura de Kripke, *i. e.*, una interpretación, para $\mathcal{L}$-{K_1, …, K_n} es un tuplo: 42

$\mathcal{E}$= <S, π, R_1, …, R_n>.

Donde:

1) S≠Ø S es un conjunto no vacío (llamado el conjunto de estados s de $\mathcal{E}$).

2) π: L⟶P(S) π es una función que va de L al conjunto potencia de S, donde L es el conjunto de letras proposicionales de $\mathcal{L}$-{K_1, …, K_n} ($\pi(p_i)\subseteq S$).

3) Para cada j=1, …, n, $R_j\subseteq S^2$ (R_j se llama la relación de accesibilidad o posibilidad de j).

Definición de verdad para $\mathcal{L}$-{K_1,…,K_n}

Definimos primero, recursivamente, para cada s∈S y cada *fbf* σ de $\mathcal{L}$-{K_1, …, K_n}, lo que significa:

$\mathcal{E}$, s⊨σ (se lee: la fórmula σ es verdadera (realizada o satisfecha) en el estado s bajo la interpretación (estructura) $\mathcal{E}$).

Definición de satisfacibilidad:

Cláusula 1: σ es una letra proposicional

Por lo tanto, $\sigma \in L$ (σ es una letra proposicional p_i)

$\mathcal{E}, s \vDash p_i$ sii $s \in \pi(p_i)$.

Cláusula 2: $\mathcal{E}, s \vDash \neg\sigma$ sii $\mathcal{E}, s \nvDash \sigma$.

Cláusula 3: $\mathcal{E}, s \vDash \alpha \rightarrow \beta$ sii $\mathcal{E}, s \nvDash \alpha$ o $\mathcal{E}, s \vDash \beta$.

Cláusula 4: Para $j=1, \ldots, n$, $\mathcal{E}, s \vDash K_j\sigma$ sii $\mathcal{E}, s' \vDash \sigma$ para todo $s' \in S$ tal que $\langle s, s' \rangle \in R_j$.

A partir de lo anterior, podemos dar la siguiente definición de verdad para nuestro lenguaje.

Sea σ cualquier *fbf* de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$.

Definición de verdad:

σ es verdadera en una estructura $\mathcal{E}$ sii para todo estado s de $\mathcal{E}$ se tiene que $\mathcal{E}, s \vDash \sigma$.

Notación: $\mathcal{E} \vDash \sigma$

Por último, definamos lo que significa que σ sea una fórmula válida de nuestro lenguaje (una verdad lógica epistémica de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$).

Definición de validez lógca:

σ es válida sii para toda estructura $\mathcal{E}$ se tiene que $\mathcal{E} \vDash \sigma$

Notación: $\vDash \sigma$

Ilustremos estas nociones con algunos ejemplos. Consideremos la siguiente $\mathcal{E}$ interpretación o estructura para $\mathcal{L}$-$\{K_1, \ldots, K_n\}$.

$\mathcal{E} = \langle S, \pi, R_1, \ldots, R_n \rangle$

$S=\{r, s, t\}$

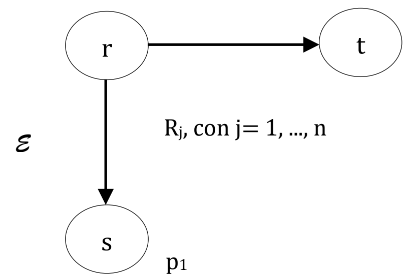
$\pi: L \longrightarrow P(S)$

$\pi(\psi)=\emptyset$, para toda letra proposicional ψ, excepto la letra proposicional p_1.

$\pi(p_1)=\{s\}$

$R_j= \{\langle r, s \rangle, \langle r, t \rangle\}$, con $j= 1, \ldots, n$

Con frecuencia, las interpretaciones se representan mediante diagramas en los que se emplean nodos y flechas, como podemos ver a continuación:



Este es un diagrama de la interpretación $\mathcal{E}$, en el que los nodos representan los estados o elementos del conjunto S, mientras que la flecha representa la relación de accesibilidad R_j (con j= 1, ..., n) entre tales estados. Tomando en cuenta esta interpretación, y mediante el uso de las definiciones anteriores, podemos probar lo siguiente:

a) $\mathcal{E} \vDash p_1 \rightarrow K_1 p_1$.

Prueba:

Como $r \notin \pi(p_1)$ y $t \notin \pi(p_1)$, tenemos, respectivamente, por cláusula 1 de nuestra definición de satisfacibilidad:

1) $\mathcal{E}$, $r \nvDash p_1$ y

2) $\mathcal{E}$, $t \nvDash p_1$.

De 1) y 2) se sigue, respectivamente, por cláusula 3 de nuestra definición de satisfacibilidad:

3) $\mathcal{E}$, $r \vDash p_1 \rightarrow K_1 p_1$ y

4) $\mathcal{E}$, $t \vDash p_1 \rightarrow K_1 p_1$.

Ahora bien, como $<s, s^*> \notin R_1$, para cualquier $s^* \in S$, tenemos que, por cláusula 4 de nuestra definición de satisfacibilidad,

5) $\mathcal{E}$, $s \vDash K_1 p_1$, pues $\mathcal{E}$, $s' \vDash p_1$, para todo $s' \in S$ tal que $<s, s'> \in R_1$.

De 5) y cláusula 3 de nuestra definición de satisfacibilidad, se sigue que:

6) $\mathcal{E}$, $s \vDash p_1 \rightarrow K_1 p_1$.

Por último, de 3), 4), 6) y nuestra definición de verdad, se sigue que: $\mathcal{E} \vDash p_1 \rightarrow K_1 p_1$.

Podemos probar, a su vez, afirmaciones generales como:

b) $\models\alpha\to(\beta\to\alpha)$[42].

Prueba:

Demostración por reducción al absurdo.

Supongamos que existe una estructura $\mathcal{E}$ y existe s de $\mathcal{E}$ tal que

1) $\mathcal{E}$, s$\not\models\alpha\to(\beta\to\alpha)$. De esta suposición y cláusula 3 de definición de satisfacibilidad, tenemos que:

2) $\mathcal{E}$, s$\models\alpha$ y

3) $\mathcal{E}$, s$\not\models(\beta\to\alpha)$.

De 3) y cláusula 3 de definición de satisfacibilidad, se sigue que:

4) $\mathcal{E}$, s$\models\beta$ y

5) $\mathcal{E}$, s$\not\models\alpha$**!**

Pero 5) se contradice con 2).

Por lo tanto,

$\models\alpha\to(\beta\to\alpha)$.

Esta semántica resulta adecuada para nuestro sistema epistémico $\mathbb{K}$, en el sentido de que rescata única y exclusivamente los teoremas del sistema, *i. e.*, con ella podemos probar los metateoremas de corrección y completez para $\mathbb{K}$. En la sección subsecuente probaremos el primero de estos metateoremas, el metateorema de corrección[43].

2.4. Metateorema de corrección para $\mathbb{K}$

El metateorema de corrección para el sistema epistémico $\mathbb{K}$ se puede formular como sigue:

> Sea φ un teorema cualquiera de $\mathbb{K}$. Entonces se tiene que:
> $\models\varphi$.

---

[42] En realidad, lo que se está afirmando aquí es que cualquier instancia del esquema '$\alpha\to(\beta\to\alpha)$' es una fórmula válida.

[43] La demostración del metateorema de completez para $\mathbb{K}$, así como para los sistemas epistémicos T, S4 y S5, se presentará en el Apéndice II de este trabajo.

Note que esta afirmación establece una relación entre nuestro sistema sintáctico y la semántica que recién se presentó, pues declara que todo teorema de $\mathbb{K}$ es una fórmula válida, *i. e.*, es verdadero bajo cualquier interpretación o estructura. Para probar tal afirmación, estableceremos primero que todos los axiomas de $\mathbb{K}$ son fórmulas válidas y que sus reglas de inferencia preservan validez.

Al final de la sección anterior, probamos que toda instancia del esquema axiomático '$\alpha\rightarrow(\beta\rightarrow\alpha)$' es una fórmula válida. Probemos ahora que cualquier instancia de los tres axiomas restantes de $\mathbb{K}$, A2, A3 y K, es una fórmula válida.

c) $\vDash(\alpha\rightarrow(\beta\rightarrow\gamma)\rightarrow(\alpha\rightarrow\beta)\rightarrow(\alpha\rightarrow\gamma)$.

Prueba por reducción al absurdo.

Supongamos que existe una estructura $\mathcal{E}$ y existe s de $\mathcal{E}$ tal que

1) $\mathcal{E}$, s$\nvDash(\alpha\rightarrow(\beta\rightarrow\gamma))\rightarrow((\alpha\rightarrow\beta)\rightarrow(\alpha\rightarrow\gamma))$. De esta suposición se sigue que:

2) $\mathcal{E}$, s$\vDash\alpha\rightarrow(\beta\rightarrow\gamma)$ y 3) $\mathcal{E}$, s$\nvDash(\alpha\rightarrow\beta)\rightarrow(\alpha\rightarrow\gamma)$.

De 3), tenemos que: 4) $\mathcal{E}$, s$\vDash\alpha\rightarrow\beta$ y 5) $\mathcal{E}$, s$\nvDash\alpha\rightarrow\gamma$.

De 5), se sigue que: 6) $\mathcal{E}$, s$\vDash\alpha$ y 7) $\mathcal{E}$, s$\nvDash\gamma$.

De 6) y 2), tenemos que: 8) $\mathcal{E}$, s$\vDash(\beta\rightarrow\gamma)$.

De 8) y 7), se sigue que: 9) $\mathcal{E}$, s$\nvDash\beta$.

De 9) y 6), tenemos que: 10) $\mathcal{E}$, s$\nvDash\alpha\rightarrow\beta$**!**

Pero 10) se contradice con 4). Por lo tanto,
$\vDash(\alpha\rightarrow(\beta\rightarrow\gamma)\rightarrow(\alpha\rightarrow\beta)\rightarrow(\alpha\rightarrow\gamma)$.

d) $\vDash(\neg\alpha\rightarrow\neg\beta)\rightarrow((\neg\alpha\rightarrow\beta)\rightarrow\alpha)$.

Prueba por reducción al absurdo.

Supongamos que existe una estructura $\mathcal{E}$ y existe s de $\mathcal{E}$ tal que

1) $\mathcal{E}$, s$\nvDash(\neg\alpha\rightarrow\neg\beta)\rightarrow((\neg\alpha\rightarrow\beta)\rightarrow\alpha)$. De esta suposición se sigue que:

2) $\mathcal{E}$, s$\vDash\neg\alpha\rightarrow\neg\beta$ y 3) $\mathcal{E}$, s$\nvDash(\neg\alpha\rightarrow\beta)\rightarrow\alpha$.

De 3), tenemos que: 4) $\mathcal{E}$, s$\vDash\neg\alpha\rightarrow\beta$ y 5) $\mathcal{E}$, s$\nvDash\alpha$.

De 5), se sigue: 6) $\mathcal{E}$, s$\vDash\neg\alpha$.

De 6) y 2), tenemos que: 7) $\mathcal{E}$, s$\models\neg\beta$.

De 7), se sigue que: 8) $\mathcal{E}$, s$\not\models\beta$.

De 6) y 8), tenemos que: 9) $\mathcal{E}$, s$\not\models\neg\alpha\rightarrow\beta$**!**

Pero 9) se contradice con 4). Por lo tanto,

$\models(\neg\alpha\rightarrow\neg\beta)\rightarrow((\neg\alpha\rightarrow\beta)\rightarrow\alpha)$.

e) $\models K_j(\alpha\rightarrow\beta)\rightarrow(\ (K_j\alpha\rightarrow K_j\beta)$.

Prueba por reducción al absurdo.

Supongamos que existe una estructura $\mathcal{E}$ y existe s de $\mathcal{E}$ tal que

1) $\mathcal{E}$, s$\not\models K_j(\alpha\rightarrow\beta)\rightarrow(K_j\alpha\rightarrow K_j\beta)$. De esta suposición se sigue que:

2) $\mathcal{E}$, s$\models K_j(\alpha\rightarrow\beta)$ y 3) $\mathcal{E}$, s$\not\models K_j\alpha\rightarrow K_j\beta$.

De 3), tenemos que: 4) $\mathcal{E}$, s$\models K_j\alpha$ y 5) $\mathcal{E}$, s$\not\models K_j\beta$.

De 5), se sigue que: existe s' de $\mathcal{E}$ tal que $<s, s'>\in R_j$. y 6) $\mathcal{E}$, s'$\not\models\beta$.

De 4), como $<s, s'>\in R_j$, se sigue que: 7) $\mathcal{E}$, s'$\models\alpha$. 47

De 6) y 7), tenemos que: 8) $\mathcal{E}$, s'$\not\models\alpha\rightarrow\beta$.

De 8), como $<s, s'>\in R_j$, se sigue que: 9) $\mathcal{E}$, s $\not\models K_j(\alpha\rightarrow\beta)$**!**

Pero 9) se contradice con 2). Por lo tanto,

$\models K_j(\alpha\rightarrow\beta)\rightarrow(\ (K_j\alpha\rightarrow K_j\beta)$.

Una vez que hemos establecido que los axiomas de $\mathbb{K}$ son fórmulas válidas, probemos que sus reglas de inferencias preservan validez. Las reglas de inferencia de $\mathbb{K}$ son MP (*Modus ponens*) y KG (*Generalización del conocimiento*). Sabemos que el *Modus ponens* es una forma de razonamiento correcto, *i. e.*, preserva verdad y, por tanto, preserva también validez. Respecto a si la regla KG preserva validez, considérese la siguiente prueba:

Sea α una *fbf* de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$ tal que $\models\alpha$ y sean $\mathcal{E}$ una estructura cualquiera y s un estado cualquiera de $\mathcal{E}$.

Si aplicamos a α la regla KG, obtendremos la fórmula $K_j\alpha$, con $j= 1, \ldots, n$. Como $\models\alpha$, tendremos que $\mathcal{E}, s\models K_j\alpha$, pues $\mathcal{E}, s'\models\alpha$ para cualquier s' de $\mathcal{E}$ t. q. $<s. s'>\in R_j$. Como s es un estado cualquiera de $\mathcal{E}$, tenemos, por definición de verdad, que $\mathcal{E}\models K_j\alpha$. Pero como $\mathcal{E}$ es una estructura cualquiera, tenemos, por definición de validez, que $\models K_j\alpha$. Por lo tanto, la regla KG también preserva validez.

Ahora que se ha establecido que todos los axiomas del sistema $\mathbb{K}$ son fórmulas válidas y que sus reglas de inferencia preservan validez, probemos el metateorema de corrección para este sistema.

Metateorema de corrección para $\mathbb{K}$

Sea φ una *fbf* cualquiera $\mathcal{L}$-$\{K_1, \ldots, K_n\}$.

Todo teorema de $\mathbb{K}$ es una fórmula válida. Es decir, si $\vdash^{K}\varphi$, entonces $\models\varphi$.

Demostración

Demostración por inducción matemática sobre el número m de pasos de la demostración de φ.

Base: $m=1$ (la demostración de σ tiene un renglón)

Entonces φ se demuestra en un paso. Por lo tanto, φ es un axioma de $\mathbb{K}$ (A1-K).

Pero sabemos que todo axioma de $\mathbb{K}$ es una fórmula válida. Por lo tanto, $\models\varphi$.

$m= n$, para $n>1$ (la demostración de φ consta de n renglones, con $n>1$)

Hipótesis de Inducción: El metateorema vale para todas las fórmulas que aparecen en los renglones anteriores a n.

Por Demostrar. El condicional vale para φ (que aparece en el renglón n).

Hay tres casos para φ: φ es axioma de $\mathbb{K}$, es consecuencia de anteriores por MP, o bien, es consecuencia de anteriores por KG.

Caso 1: φ es axioma de $\mathbb{K}$. Se procede igual que en la base.

Caso 2: φ es consecuencia de anteriores por MP. Entonces φ se obtuvo, por MP, de las fórmulas $\alpha \rightarrow \varphi$ y α, que aparecen en renglones anteriores de la demostración. Por consiguiente, a estas dos fórmulas les podemos aplicar HI, de modo que $\vDash \alpha \rightarrow \varphi$ y $\vDash \alpha$. Como la regla de inferencia MP preserva validez, tenemos que $\vDash \varphi$.

Caso 3: φ es consecuencia de anteriores por KG. Entonces φ es de la forma $K_j\alpha$ y se obtuvo, por KG, de la fórmula α, que aparece en un renglón anterior de la demostración. En consecuencia, a esta fórmula se le puede aplicar HI, de forma que tenemos que $\vDash \alpha$. Como la regla de inferencia KG preserva validez, tenemos que $\vDash \varphi$.

De la base y los casos 1-3, se sigue, por principio de inducción matemática completa (PIMC), que:

Si $\vdash^{K} \varphi$, entonces $\vDash \varphi$.

2.5. El poder demostrativo del sistema $\mathbb{K}$

Ahora que hemos probado el metateorema de corrección para el sistema $\mathbb{K}$, podemos emplearlo para contestar las interrogantes respecto a su fuerza demostrativa: ¿es cualquier instancia de los esquemas '$K_j\alpha \rightarrow \alpha$' y '$K_j(\alpha \vee \beta) \rightarrow (K_j\alpha \vee K_j\beta)$' un teorema de $\mathbb{K}$? Para responder, habrá que notar que el metateorema de corrección nos deja saber que todos los teoremas de $\mathbb{K}$ comparten una propiedad, *viz.*, son *fbf's* válidas de nuestro lenguaje formal; por consiguiente, basta con mostrar que una fórmula no es válida para concluir que no es teorema de $\mathbb{K}$[44]. Esto es relativamente fácil de establecer para algunas instancias de los dos esquemas anteriores. Consideremos las fórmulas '$K_1p_1 \rightarrow p_1$'

[44] Para establecer que una fórmula válida de nuestro lenguaje epistémico es teorema de $\mathbb{K}$, bastará con dar la demostración correspondiente, o bien, emplearemos el metateorema de completez para este sistema (toda fórmula válida de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$ es teorema de $\mathbb{K}$), cuya demostración la encontrará el lector en el apéndice II de este trabajo.

[instancia del esquema ‘$K_j\alpha\rightarrow\alpha$’] y ‘$K_1(p_1\vee\neg p_1)\rightarrow(K_1p_1\vee K_1\neg p_1)$’ [instancia del esquema y ‘$K_j(\alpha\vee\beta)\rightarrow(K_j\alpha\vee K_j\beta)$’] y veamos su valor de verdad en la siguiente estructura:

$\mathcal{E}= <S, \pi, R_1, \ldots, R_n>$

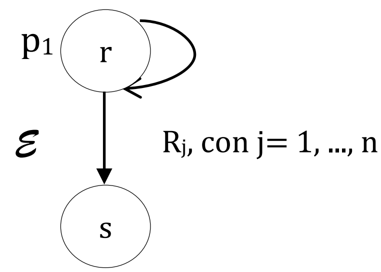
$S=\{r, s\}$

$\pi: L \longrightarrow P(S)$

$\pi(\psi)=\emptyset$, para toda letra proposicional ψ, excepto la letra proposicional p_1.

$\pi(p_1)=\{r\}$

$R_j= \{<r, r>, <r, s>\}$, con $j= 1, \ldots, n$



Bajo esta estructura, tenemos que:

1) $\mathcal{E}, r\models p_1$ y 2) $\mathcal{E}, s\not\models p_1$. De 1) y 2) se sigue respectivamente, por la cláusula de la negación, que 3) $\mathcal{E}, r\not\models\neg p_1$ y 4) $\mathcal{E}, s\models\neg p_1$. Ahora, dada la relación de accesibilidad R_1 y nuestra cláusula modal, se tiene que 5) $\mathcal{E}, s\models K_1p_1$. De 5), 2) y cláusula del condicional se tiene que 6) $\boldsymbol{\mathcal{E}, s\not\models K_1p_1\rightarrow p_1}$. Por otra parte, de 3) y 4) tenemos respectivamente, por la cláusula del condicional, que 7) $\mathcal{E}, r\models\neg p_1\rightarrow\neg p_1$ y 8) $\mathcal{E}, s\models\neg p_1\rightarrow\neg p_1$. De 7) y 8), por la relación de accesibilidad R_1 y la cláusula modal, se sigue que 9) $\mathcal{E}, r\models K_1(\neg p_1\rightarrow\neg p_1)$. No obstante, de la misma cláusula y la misma relación de accesibilidad se tiene, por 2) y 3) respectivamente, que 10) $\mathcal{E}, r\not\models K_1p_1$ y 11) $\mathcal{E}, r\not\models K_1\neg p_1$. De 10) y cláusula de la negación se sigue que 12) $\mathcal{E}, r\models\neg K_1p_1$. Pero de 12), 11) y cláusula de la implicación tenemos que 13) $\mathcal{E}, r\not\models\neg K_1p_1\rightarrow K_1\neg p_1$. Y de 9) y 13) se tiene respectivamente, por $\stackrel{def}{=}\vee$, que 14) $\mathcal{E}, r\models K_1(p_1\vee\neg p_1)$ y 15) $\mathcal{E}, r\not\models K_1p_1\vee K_1\neg p_1$. Por último, de 14), 15) y cláusula de la implicación se sigue que 16) $\boldsymbol{\mathcal{E}, r\not\models K_1(p_1\vee\neg p_1)\rightarrow(K_1p_1\vee K_1\neg p_1)}$.

Lo anterior constituye la prueba de que estas dos instancias de los esquemas en cuestión no son *fbf's* válidas de nuestro lenguaje epistémico y, en consecuencia, tampoco son teoremas del sistema $\mathbb{K}$. Por lo tanto, no toda instancia de los esquemas ‘$K_j\alpha\rightarrow\alpha$’ y ‘$K_j(\alpha\vee\beta)\rightarrow(K_j\alpha\vee K_j\beta)$’ es teorema de $\mathbb{K}$. Esto significa que nuestra teoría axiomática no reconoce como verdades o principios epistémicos: i) que el conocimiento implica la verdad de la proposición sabida y ii) que siempre que se conoce una disyunción se conoce alguno de los disyuntos. Por las razones antes expuestas, la omisión de ii) es un acierto para la teoría $\mathbb{K}$, pues ésta deja fuera, de las verdades respecto al conocimiento proposicional, algo que es francamente falso. Por el contrario, la omisión de i) es una anomalía para $\mathbb{K}$, pues ésta no contempla entre los principios epistémicos una idea que se considera una verdad fundamental de conocimiento proposicional, a saber, que éste implica la verdad de la proposición conocida. Para percatarse de la importancia de esta idea en la historia de la epistemología, basta con recordar las palabras de Immanuel Kant en el parágrafo VII de la "Introducción" a su *Lógica*: <<Una perfección fundamental del conocimiento, condición sin duda esencial e inseparable de toda perfección del mismo, es la *verdad* [...]>>[45]. Por otra parte, en un gran número de análisis tanto internistas (*v. g.,* fundacionalismo y coherentismo) como externistas (*v. g.,* causalismo y confiabilismo) del concepto de conocimiento proposicional, si bien se discute en torno a si la justificación es un ingrediente esencial del saber y cuál es la naturaleza de ésta, la verdad figura como un componente indiscutible de dicho concepto; además, la verdad suele invocarse como aquella característica que distingue al conocimiento de la mera creencia[46]. Es pues esta tradición en el análisis del concepto de conocimiento proposicional lo que hace que la omisión de i) se considere una anomalía del sistema $\mathbb{K}$.

2.6. Consideraciones en torno al sistema $\mathbb{K}$

El sistema $\mathbb{K}$ es la teoría axiomática más elemental de entre aquellas que analizaremos aquí, esto es, tiene el menor poder demostrativo, así como también hace el menor número de presupuestos epistémicos (a tal punto que, como acabamos de ver, ni

---

[45] Kant, Immanuel, *Lógica. Acompañada de Una Selección de Reflexiones del Legado de Kant,* Akal (Ed.), Madrid, 2001, p. 111.

[46] Un punto de quiebre con esta tradición lo constituye la reformulación que hace Luis Villoro del análisis tradicional del conocimiento en el capítulo 8 de *Creer, saber, conocer*, en donde se sustrae a la verdad como una nota del conocimiento y se le da la función de una idea regulativa. En el capítulo III se analizará con mayor detenimiento la relación entre saber y verdad.

siquiera reconoce entre sus principios que el conocimiento implica la verdad de lo conocido). Quisiera detenerme ahora, antes de pasar al estudio de otro sistema formal, precisamente en los supuestos epistémicos de los que sí parte la teoría. Me ha parecido que la presuposición epistémica fundamental de la teoría $\mathbb{K}$ es la siguiente: *la deducción es relevante para el conocimiento*. Esta declaración encierra, en realidad, dos ideas. A la primera de ellas me referiré con el nombre de *Supuesto aristotélico*, y consiste en la idea de que todo aquello que tiene demostración vale como conocimiento [47]; tal pensamiento se halla expresado por la regla de inferencia KG: Si $\vdash^{K} \alpha$, entonces $\vdash^{K} K_i\alpha$[48]. Es cierto que si por *demostración* entendemos *una demostración en la teoría K*, este supuesto parece una estipulación sin importancia; sin embargo, si por *demostración* entendemos *un razonamiento deductivo no circular que parte de premisas bien establecidas*, entonces nos hallaremos ante una idea que no sólo ha estado presente en la historia de la filosofía, notablemente en la obra de Aristóteles, sino que sigue vigente en las disciplinas formales, *i. e.*, en la lógica y la matemática –de lo contrario, ¿cómo se explica el empeño arduo por demostrar cada una de sus afirmaciones? La explicación más razonable es que en estas disciplinas la demostración eleva a las creencias al rango de conocimiento, pues con ella se satisface la cláusula de justificación para las proposiciones lógicas y matemáticas–[49]. Además de lo anterior, este supuesto revela

[47] Aristóteles llama a la demostración [*apódeixis*] o deducción demostrativa [apodeiktikós *syllogismós*] razonamiento científico, y afirma: <<[...] llamo científico a aquel <razonamiento> en virtud de cuya posesión sabemos>> (*Analíticos Segundos,* 71b 16-19).

[48] En realidad, la regla KG afirma algo más fuerte que el supuesto que he llamado aristotélico, a saber, que toda proposición o fórmula demostrada es sabida por *todos* los agentes epistémicos. El supuesto en cuestión se expresaría fielmente en un lenguaje epistémico para modalidades absolutas, es decir, con operadores epistémicos sin subíndices para los agentes (ver nota al pie número 5 del capítulo I). Si definiéramos la teoría $\mathbb{K}$ en un lenguaje así, la regla KG tendría la forma siguiente: Si $\vdash^{K} \alpha$, entonces $\vdash^{K} K\alpha$. Intuitivamente, esta regla nos dice que cualquier proposición demostrada es conocida (*i. e.*, es sabida por alguien o por la humanidad en su conjunto, aunque no necesariamente por todo agente epistémico); y en esto consiste justamente el supuesto aristotélico.

[49] Resulta interesante, sin embargo, explorar las relaciones entre ambos conceptos de demostración, el aristotélico y el que se definió para el sistema $\mathbb{K}$. Aristóteles caracteriza a la demostración <<[...] como el razonamiento [que] parte de cosas verdaderas y primordiales, o de cosas cuyo conocimiento se origina

que aun si la noción de justificación no se encuentra explícitamente presente en el sistema $\mathbb{K}$ (y en ningún otro de los sistemas formales aquí estudiados), sí está contenida, en cierto modo, en la regla de inferencia KG. La segunda idea en $\mathbb{K}$ en torno al saber, y a la que he nombrado el *Supuesto de preservación del conocimiento*, es la de que a través del *Modus Ponens* se preservan las propiedades del conocimiento, cualesquiera que éstas sean. En otras palabras, si un agente sabe que una proposición implica a otra y conoce, además, la primera de ellas, entonces (por *Modus Ponens*) también conoce la segunda proposición. Así, por ejemplo, si adoptáramos el análisis tradicional del conocimiento –según el cual el conocimiento es creencia verdadera

---

a través de cosas primordiales y verdaderas>> (*Tópicos*, 100a 25-30)), y de éstas últimas dice que <<son *verdaderas* y *primordiales* las cosas que tienen credibilidad, no por otras, sino por sí mismas (en efecto, en los principios cognoscitivos no hay que inquirir el porqué, sino que cada principio ha de ser digno de crédito en sí mismo)>> (*Tópicos*, 100b, 20-25). Esta caracterización muestra que una primera diferencia entre ambas nociones de demostración es que, según parece para el estagirita, los puntos de arranque más básicos de cualquier demostración consisten en verdades fundamentales que se acreditan por sí mismas, mientras que para nosotros estos puntos de arranque, *i. e.*, los axiomas de K, no son verdades indiscutibles, sino, a lo más, afirmaciones intuitivamente aceptables (cuando se les interpreta, claro está). Otro punto de desencuentro importante sería que, a diferencia de la *apódeixis* aristotélica, las demostraciones en $\mathbb{K}$ no sólo admiten razonamientos estrictamente deductivos, esto es, que preservan verdad, sino que además del *Modus Ponens*, aceptan el empleo de la regla de inferencia KG, que es la contraparte sintáctica de una forma de razonamiento que no preserva verdad, pero sí validez lógica. En cuanto a las similitudes entre ambas nociones, cabe señalar que, como en Aristóteles, las demostraciones en $\mathbb{K}$ parten de cosas "primordiales", ya que sus axiomas pueden aparecer en las líneas de una demostración sin que medie para ello alguna regla de inferencia –justo por esto sirven como puntos de partida para las demostraciones–, y sin ellos no hay demostración. Además, tanto en la noción aristotélica como en el sistema $\mathbb{K}$, las demostraciones son, *stricto sensu*, argumentos: Aristóteles caracteriza a la demostración como una forma de razonamiento deductivo, mientras que en $\mathbb{K}$ las demostraciones son listas de *fbf's* de nuestro lenguaje formal, *i. e.*, enunciados de este lenguaje, construidas a partir de los axiomas y las reglas de inferencia. Por último, la contemplación de la regla KG revela todavía una semejanza más con la epistemología aristotélica: dado que la regla KG permite que se pase de un axioma al conocimiento de este axioma sin la necesidad de alguna otra regla de inferencia, el sistema $\mathbb{K}$ encierra en cierto modo la idea aristotélica de que los axiomas o puntos de partida de la demostración son conocidos por sí mismos y no a partir de algo más.

justificada—, nuestro supuesto equivaldría a afirmar que los razonamientos de la forma *Modus Ponens* no sólo preservan verdad, sino también la creencia justificada. La idea de la preservación del conocimiento se pone de manifiesto en K mediante el axioma que da nombre a este sistema, *i. e.*, el axioma K: $K_j(\alpha\rightarrow\beta)\rightarrow(K_j\alpha\rightarrow K_j\beta)$[50]. Respecto a este axioma, hay que agregar que si bien en conjunto con la regla KG (y el metateorema de completez para la lógica clásica proposicional) nos conduce al problema de la omnisciencia lógica (o deductiva), *i. e.*, a la tesis de que los agentes conocen todas las implicaciones lógicas de las proposiciones por ellos sabidas, por sí solo no genera tal problema, ni, al parecer, ninguna otra consecuencia falsa o absurda. No obstante, el axioma K se formula de modo distinto en ocasiones, concretamente, en la forma siguiente: K*: $(\alpha\rightarrow\beta)\rightarrow(K_j\alpha\rightarrow K_j\beta)$[51]. Tal formulación torna prácticamente irrelevante a la regla de inferencia KG, mientras que conduce directamente a la primera de nuestras reglas derivadas (RD1)[52]. Los inconvenientes de esta formulación del axioma es que lleva ya no a la omnisciencia lógica, sino a la *omnisciencia (o ignorancia) absoluta*. Veamos cómo:

1. $(\alpha\rightarrow\beta)\rightarrow(K_j\alpha\rightarrow K_j\beta)$	*K**
2. $\beta\rightarrow(\alpha\rightarrow\beta)$	*Tautología*
3. $\beta\rightarrow(K_j\alpha\rightarrow K_j\beta)$	*1, 2, Transitividad*
4. $K_j\alpha\rightarrow(\beta\rightarrow K_j\beta)$	*3, Tautología*

La fórmula que aparece al final de la demostración declara que si el agente *j* conoce alguna proposición, entonces conoce también todas las proposiciones verdaderas; es decir, *j* conoce todas las verdades o ignora cualquier proposición. A esto

---

[50] La lectura más usual de este axioma es la siguiente: éste declara que el conocimiento es transmisible a través de las implicaciones lógicas *conocidas por el agente*. Por ello también se le llama el *principio de transmisibilidad del conocimiento*. Ver Feldman, Richard, "Skepticism (I)", *Epistemology*, New Jersey, Prentice Hall, 2003, pp. 108-129.

[51] Ver especialmente la nota 2 del trabajo de Collier, Kenneth en Radu J. Bogdan (ed.), *'Jaakko Hintikka'*, 181-198, 1987, D. Reidel Publishing Company.

[52] Recordemos esta regla. RD1: Si $\vdash^{K} \alpha\rightarrow\beta$, entonces $\vdash^{K} K_j\alpha\rightarrow K_j\beta$.

es a lo que me he referido antes como la omnisciencia o ignorancia absolutas. Este resultado es problemático no sólo por tratarse de algo completamente contraintutivo, sino porque conduce directamente a una conclusión escéptica: dado este resultado, de K* y del hecho claramente manifiesto de que no conocemos todas las verdades, se sigue lógicamente que no conocemos proposición alguna (= la ignorancia absoluta). No obstante, la versión más moderada del axioma, *i. e.*, el principio K, también ha servido para la causa escéptica a través de un razonamiento que lleva por nombre el *argumento de la transmisibilidad*[53].

2.6.1. *El argumento de la transmisibilidad*

Sea ψ cualquier proposición ordinaria acerca del mundo exterior; sea HE cualquier hipótesis escéptica lógicamente incompatible con ψ (*i. e.*, ψ implica lógicamente la falsedad de HE; y sea S un sujeto arbitrario que sabe que ψ implica lógicamente la falsedad de HE.

P1) S sabe que ψ implica lógicamente la falsedad de HE [presuposición del argumento].

P2) Si S sabe que ψ implica lógicamente la falsedad de HE, entonces, si S sabe que ψ es verdadera, S sabe que HE es falsa [principio K].

PI) Si S sabe que ψ es verdadera, entonces S sabe que HE es falsa [de P1) y P2) por MP].

P3) S no sabe que HE es falsa [posición escéptica].

∴

C) S no sabe que ψ es verdadera [de PI) y P3) por MT].

---

[53] El argumento que se presenta a continuación está diseñado para objetar nuestras pretensiones de conocimiento acerca del mundo exterior y, para ello, se apoya en hipótesis o escenarios posibles a la manera como lo hace el escepticismo cartesiano. Para dos versiones de este argumento ligeramente distintas de la aquí presentada, cf. Dretske, Fred, "Epistemic Operators,"Journal of Philosophy 67 (1970): 1007-23; Feldman, Richard, "Skepticism (I)", *Epistemology*, New Jersey, Prentice Hall, 2003, pp. 108-129. De hecho, en la exposición del argumento sigo muy de cerca a este último.

Es éste un argumento escéptico que ataca nuestro conocimiento acerca del mundo exterior, y para tornarlo más transparente, recurriré a un ejemplo muy sencillo. Considérese a un individuo relativamente ordinario, por ejemplo, al escritor español Miguel de Cervantes, autor del *Quijote*. Tomemos ahora alguna de sus creencias ordinarias acerca del mundo exterior y consideremos la proposición correspondiente, *exempli gratia*:

MA: Cervantes tiene una mano anquilosada.

Finalmente, consideremos la hipótesis escéptica del *cerebro en el balde*, de Hilary Putnam, referida a Cervantes, *i. e.*,

*CEB: Cervantes es **simplemente** un cerebro en un balde (un estanque escéptico) que se halla conectado a una máquina de vida artificial.*

Nótese que, según esta hipótesis, Cervantes no tiene un cuerpo más allá del cerebro y el balde; por tanto, no posee una mano la cual pudiera estar anquilosada. En otras palabras, CEB es lógicamente incompatible con MA, lo que significa que MA implica lógicamente la falsedad de CEB. Supongamos ahora que el autor del *Quijote* es lo suficientemente taimado para percatarse de esta implicación, *i. e.*, que él conoce -intuitivamente o con justificación tal implicación. Entonces, de acuerdo al principio K, si Cervantes sabe que MA implica la falsedad de CEB y sabe, además, que tiene una mano atrofiada (MA), entonces debería saber también que no es un cerebro en un balde (CEB). Sin embargo, ¿Cervantes sabe esto último?, ¿qué recurso emplearía para saber algo semejante? En este caso, algunas posibles fuentes de conocimiento son la percepción y la memoria, las cuales incluyen, entre otras cosas, el cúmulo de sensaciones, percepciones y recuerdos del escritor acerca de sí mismo –*v. g.*, Cervantes podría recordar su participación en la batalla de Lepanto, además de observar el estado actual de su cuerpo–. No obstante, todo esto (*i. e.,* el conjunto completo de sus estados internos) es compatible con la hipótesis del cerebro en el balde, pues cualquiera de sus contenidos mentales bien podría haber sido generado o implantado por una súper computadora que alimenta con impulsos eléctricos el cerebro que es Cervantes. En otras palabras, la hipótesis CEB y la proposición MA, en general, la creencia de

Cervantes de que es una persona normal (un cerebro en un cuerpo) en un mundo normal, están subdeterminadas tanto por la percepción como por la memoria de éste, dado que no le permiten distinguir cuál de ellas es falsa y cuál de ellas es verdadera (si alguna lo es). Como esta forma de evidencia no le permite discernir la verdad del error, no lo justifica para creer que CEB es falsa y, en consecuencia, tampoco le proporciona conocimiento al respecto. Cualquiera que se detenga a pensar un momento se dará cuenta de que no es nada sencillo saber algo así, y de que quizá Cervantes, más todavía, cualquiera de nosotros, no posee conocimiento al respecto. Si esto es así, si después de todo Cervantes no sabe que no es un cerebro en un balde, entonces tampoco sabe –si nos atenemos a K– que tiene una mano atrofiada, aun si él mismo se encuentra completamente persuadido de ello.

En este argumento, la premisa P1) resulta verdadera por construcción, pero vale la pena percatarse de que el escepticismo sobre el mundo exterior no tiene por qué socavar el conocimiento lógico que se atribuye al agente S; la razón es que las verdades lógicas no son verdades empíricas o acerca del mundo externo y, por tanto, el conocimiento lógico no es conocimiento del mundo exterior. La premisa P2) es justo el principio K de nuestro sistema axiomático, y ya sea que le dé una lectura restringida (el conocimiento se hereda a través del *Modus Ponens*) o una lectura más amplia (el conocimiento se hereda mediante las implicaciones lógicas *conocidas por el agente*), es un principio completamente plausible. El paso intermedio PI) se obtiene a partir de estas dos premisas, mientras que la conclusión se deriva de PI) y la tercera premisa P3). Es esta última premisa, *S no sabe que HE es falsa*, la clave del argumento. Sin embargo, no la discutiré aquí, pues mi objetivo no es realizar un examen del escepticismo, sino señalar que uno de los principios fundamentales de la lógica epistémica, *viz.*, la propiedad de distribución o axioma K, ha tenido un papel en los argumentos escépticos.

Los dos supuestos anteriores –el aristotélico y el de preservación de conocimiento–, por problemáticos que pudieran ser, constituyen las intuiciones epistémicas básicas que se intentan formalizar mediante la teoría K, y están presentes también en las teorías subsecuentes, *i. e.*, en T, $\mathrm{S4}$ y $\mathrm{S5}$, en la medida en que todas ellas tienen como base a K. A continuación, haremos el análisis de estas tres teorías axiomáticas restantes, pero ahora que hemos explicado toda la terminología necesaria nuestro análisis podrá ser relativamente más breve.

3. Sistemas de Lógica Epistémica (El sistema T)

3.1. El sistema T

La tradición filosófica de análisis del concepto de conocimiento se originó en Grecia, gracias a Platón, y los análisis contemporáneos responden, casi por completo, al filósofo ateniense: en su mayoría trabajan con la materia platónica, *i. e.*, son variaciones del análisis platónico del saber. La discusión epistemológica se ha centrado principalmente en la noción de justificación, un componente del conocimiento ya presente en el *Teetetes*, pero al que el autor, según parece, no dotó de una significación precisa. De esta discusión derivan las dicotomías internismo-externismo y fundacionismo-coherentismo, en Epistemología; incluso, algunas formulaciones externistas descartan a la justificación como condición necesaria del saber[54]. Respecto de la verdad, el escenario es distinto. Que el conocimiento requiere verdad constituye una idea aceptada por la mayoría; es, por decirlo así, un *endoxon* u opinión reputada de la Epistemología contemporánea. Así lo evidencian varias definiciones de saber:

[La teoría de la ausencia de fundamentos falsos]

> NFG. S sabe p = df. (i) S cree p; **(ii) p es verdadera**; (iii) S está justificado en creer p; (iv) Todos los fundamentos de S para creer p son verdaderos[55].

[La teoría de la ausencia de socavadores]

> ND. S sabe p = df. (i) S cree p; **(ii) p es verdadera**; (iii) S está justificado en creer p; (iv) No hay una proposición verdadera t tal que, si S estuviera justificado en

---

[54] Así lo hace la teoría causal de Alvin Goldman, en donde la justificación es sustituida por una relación causal apropiada entre el hecho de conocimiento y la creencia del agente en ese hecho. Alvin Goldman, "A Causal Theory of Knowing", *Journal of Philosophy* 64 (1967): 357-72.

[55] Michael Clark, "Knowledge and Grounds: A Comment on Mr. Gettier's Paper," *Analysis* XXIV (1963): 46-48. Las **negritas** son mías tanto en esta definición como en las dos subsecuentes.

creer t, entonces S no estaría justificado en creer p. (Ninguna verdad socava la justificación de S para p)[56].

[La teoría de los rastreadores de verdad]

TT. S sabe p iff (i) S cree p; **(ii) p es verdadera**; (iii) La actitud de S hacia p rastrea la verdad de p: cuando p no es verdadera, S no cree p; y cuando p es verdadera, S cree p[57].

Las dos primeras definiciones son extensiones del análisis tradicional del conocimiento, pues a las tres cláusulas de éste –creencia, justificación y verdad– agregan una cuarta condición: la ausencia de algún fundamento falso, en el primer caso, y la ausencia de algún socavador para la justificación, en el segundo. La tercera definición, por el contrario, constituye una alternativa al análisis tradicional, ya que sustituye la cláusula de justificación por otra, específicamente, por la del rastreo de la verdad. No obstante, en todas ellas la *verdad*, más aún, la *creencia verdadera*, es un ingrediente fundamental del saber. Tan *cara amica* es la verdad para los epistemólogos que llaman a ésta la propiedad de conocimiento, pues lo distingue de la mera creencia, en donde no se exige la verdad sino sólo la consistencia[58]. Es esta *endoxon* de que el saber implica la verdad de lo conocido la que queda recogida por el sistema axiomático $\mathbb{T}$ de Lógica Epistémica, y ello a través del axioma que lo caracteriza, *i. e.*, el axioma T: $K_j\alpha\rightarrow\alpha$ [propiedad de conocimiento]. Tal sistema es una extensión de la teoría $\mathbb{K}$, pues se obtiene de ésta al agregar el axioma T. Por esto mismo, el sistema $\mathbb{T}$ hereda la fuerza demostrativa de $\mathbb{K}$: todo lo que es demostrable en $\mathbb{K}$ también es demostrable en $\mathbb{T}$; además, hereda los supuestos epistemológicos de aquella, *i. e.*, el supuesto aristotélico y el supuesto de preservación del conocimiento, a los que agrega un tercer supuesto,

[56] Peter Klein, "Knowledge and Grounds: A Comment on Mr. Gettier's Paper," *Analysis* XXIV (1963): pp. 46-48.

[57] Robert Nozick, *Philosophical Explanation* (Cambridge, MA: Harvard University Press), (1981), Chapter p. 3.

[58] Claro está, desde luego, que esta distinción no se sostiene para el caso de las creencias *verdaderas*.

viz., el de que sólo conocemos lo verdadero. Los axiomas y reglas de inferencia de T son, entonces, los siguientes:

Axiomas de T

Sean α, β y γ *fbf 's* cualesquiera de $\mathcal{L}$-{K_1, ..., K_n} y sea j=1, ..., n.

A1: $\alpha\rightarrow(\beta\rightarrow\alpha)$,

A2: $(\alpha\rightarrow(\beta\rightarrow\gamma))\rightarrow((\alpha\rightarrow\beta)\rightarrow(\alpha\rightarrow\gamma))$,

A3: $(\neg\alpha\rightarrow\neg\beta)\rightarrow((\neg\alpha\rightarrow\beta)\rightarrow\alpha)$,

K: $K_j(\alpha\rightarrow\beta)\rightarrow(K_j\alpha\rightarrow K_j\beta)$ [Propiedad de distribución],

T: $K_j\alpha\rightarrow\alpha$ [Propiedad de conocimiento].

Reglas de inferencia de T

Sean α y β *fbf 's* cualesquiera de $\mathcal{L}$-{K_1, ..., K_n}.

MP: Si $\vdash^{T}\alpha$ Y $\vdash^{T}\alpha\rightarrow\beta$, entonces $\vdash^{T}\beta$ [*Modus ponens*] y

KG: Si $\vdash^{T}\alpha$, entonces $\vdash^{T}K_i\alpha$ [Generalización del conocimiento][59].

En cuanto al poder demostrativo de T, además de todo lo demostrable en K, se obtienen algunos otros teoremas de gran interés, *v.g.*:

Teoremas de T

T16 $\vdash^{T}\alpha\rightarrow\neg K_j\neg\alpha$

T17 $\vdash^{T}K_j\alpha\rightarrow\neg K_j\neg\alpha$

T18 $\vdash^{T}(K_j\alpha\wedge K_j\beta)\rightarrow\neg(\alpha\rightarrow\neg\beta)$[60]

El primero de estos teoremas, el teorema 16, nos dice que no podemos conocer lo que es contrario a la verdad, mientras que los dos teoremas restantes son teoremas de consistencia: el teorema 17 declara que si un agente sabe una proposición, entonces no sabe la negación de ésta, *i. e.*, el conocimiento de los agentes es consistente; el

---

[59] Una vez más, empleo la notación '$\vdash^{T}$' para significar 'ser teorema de T'. Este sistema axiomático fue desarrollado por Feys y Von Wright.

[60] Para la demostración de estos teoremas, consúltese el apéndice I.

teorema 18, por su parte, afirma que si un agente sabe dos proposiciones cualesquiera, entonces estas proposiciones por sí solas deben ser consistentes entre sí. Que estos dos últimos teoremas se obtengan a partir del axioma T nos deja saber que la verdad en el conocimiento implica su consistencia.

Ahora bien, al igual que hicimos con el sistema K, podemos preguntarnos ahora por el alcance demostrativo de T e interrogar, por ejemplo, si en él se demuestra cualquier instancia del esquema de fórmulas $K_j\alpha \rightarrow K_jK_j\alpha$, conocido como el *principio de introspección positiva* o la *KK-tesis*, el cual será fundamental para la teoría S4 que se analizará posteriormente, y que declara que si un agente sabe una proposición, entonces sabe que la sabe. En otras palabras, con ese esquema se afirma que el conocimiento de primer orden implica siempre un conocimiento de segundo orden. Por otra parte, sabemos que cualquier instancia del esquema converso o recíproco a éste, *i. e.*, $K_jK_j\alpha \rightarrow K_j\alpha$, es teorema de T, pues se obtiene del axioma T: $K_j\alpha \rightarrow \alpha$ al sustituir simplemente 'α' por '$K_j\alpha$'. Sin embargo, de ello no se sigue que toda instancia de $K_j\alpha \rightarrow K_jK_j\alpha$ sea también teorema del sistema. Para responder a nuestra interrogante, procederemos igual que con la teoría K: definiremos una nueva semántica para nuestro lenguaje formal, a partir de la cual probaremos el metateorema de corrección para el sistema T; por último, usaremos este metateorema para establecer que no cualquier instancia del esquema anterior es teorema de T.

3.2. Semántica formal para el sistema T

En el capítulo anterior, probamos que algunas instancias del axioma T no son verdades lógicas o fórmulas válidas de nuestro lenguaje; en este sentido, la semántica que definimos no nos permite rescatar la validez del principio de *propiedad del conocimiento*[61]. Por ello, definiremos ahora una nueva semántica, que resulta de

[61] Ver la sección "El poder demostrativo del sistema K" en el capítulo anterior.

imponer una restricción a las relaciones de accesibilidad en nuestra noción de interpretación. Veamos cómo.

Definición de T-interpretación para $\mathcal{L}$-{K_1, ..., K_n}

Una T-interpretación ($\mathcal{E}_T$) para $\mathcal{L}$-{K_1, ..., K_n} es un tuplo:

$\mathcal{E}$= <S, π, R_1, ..., R_n>.

Donde:

1) S≠Ø, S es un conjunto no vacío (llamado el conjunto de estados s de $\mathcal{E}$).

2) π: L⟶P(S), π es una función que va de L al conjunto potencia de S, donde L es el conjunto de letras proposicionales de $\mathcal{L}$-{K_1, ..., K_n} ($\pi(p_i)\subseteq S$).

3) Para cada j=1, ..., n, $R_j\subseteq S^2$ es una relación reflexiva. Esto es: Para cualquier s∈S se tiene que <s, s>∈R_j (R_j se llama la relación de accesibilidad o posibilidad de j)

Nótese que esta noción de interpretación es idéntica a la ofrecida en el capítulo anterior, excepto por el hecho de que ahora se ha impuesto una restricción a las relaciones de accesibilidad (las R´s), a saber, que todas ellas sean relaciones reflexivas. Esto es así justo porque el axioma T exige, para su validez, que todos los mundos sean accesibles a sí mismos, de tal suerte que cualquier agente que conozca o ignore una proposición a partir de un mundo dado tenga acceso al valor de verdad de dicha proposición en ese mundo. Con ello, los agentes sólo conocerán lo que es verdadero en el mundo "real", es decir, en el mundo desde el cual conocen. Y es precisamente esto último lo que se declara con el axioma T. De hecho, en capítulo anterior, al mostrar que existía una estructura que hacía falsa a una instancia del axioma T, empleamos para ello una interpretación en donde la relación de accesibilidad no era reflexiva, ya que había un mundo o estado s∈S en la estructura que no tenía acceso a sí mismo, lo que hizo posible que la instancia del axioma fallara en la interpretación.

Definición de T-verdad para $\mathcal{L}$-{K_1, ..., K_n}

Nuestra noción de verdad para T es básicamente la misma que para el sistema K.

Sea $\mathcal{E}$ una T–interpretación cualquiera.

Definimos para cada $s \in S$ y cada *fbf* σ de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$ lo que significa:

$\mathcal{E}, s \vDash \sigma$ (se lee: la fórmula σ es verdadera (realizada o satisfecha) en el estado s bajo la interpretación (estructura) $\mathcal{E}$).

Definición de T-satisfacibilidad:

Cláusula 1: σ es una letra proposicional

Por lo tanto, $\sigma \in L$ (σ es una letra proposicional p_i)

$\mathcal{E}, s \vDash p_i$ sii $s \in \pi(p_i)$.

Cláusula 2: $\mathcal{E}, s \vDash \neg\sigma$ sii $\mathcal{E}, s \nvDash \sigma$.

Cláusula 3: $\mathcal{E}, s \vDash \alpha \rightarrow \beta$ sii $\mathcal{E}, s \nvDash \alpha$ o $\mathcal{E}, s \vDash \beta$.

Cláusula 4: Para $j=1, \ldots, n$, $\mathcal{E}, s \vDash K_j\sigma$ sii $\mathcal{E}, s' \vDash \sigma$ para todo $s' \in S$ tal que $\langle s,s' \rangle \in R_j$.

Sea σ cualquier *fbf* de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$.

Definición de T-verdad:

σ es verdadera en una T–interpretación $\mathcal{E}$ sii para todo estado s de $\mathcal{E}$ se tiene que $\mathcal{E}, s \vDash \sigma$.

Notación: $\mathcal{E} \vDash \sigma$

Definamos ahora lo que significa que σ sea una fórmula T-válida de nuestro lenguaje.

Definición de T-validez:

σ es T-válida sii para toda T–interpretación $\mathcal{E}$ se tiene que $\mathcal{E} \vDash \sigma$.

Notación: $\vDash^{T} \sigma$

Esta nueva semántica para $\mathcal{L}$-{K_1, ..., K_n}, *i. e.*, esta noción de interpretación y verdad[62], nos permite probar la validez del axioma T[63]:

$\vDash^T K_j\alpha\rightarrow\alpha$.

Prueba por reducción al absurdo.

Supongamos que existe una T-interpretación $\mathcal{E}$ tal que $\mathcal{E}\nvDash K_j\alpha\rightarrow\alpha$. Es decir, existe s de $\mathcal{E}$ en donde:

1) $\mathcal{E}$, s$\nvDash K_j\alpha\rightarrow\alpha$. De esta suposición se sigue que 2) $\mathcal{E}$, s$\vDash K_j\alpha$ y 3) $\mathcal{E}$, s$\nvDash\alpha$. Pero como R_j es reflexiva, tenemos que $<s, s>\in R_j$. De esto y de 3) se sigue, por cláusula 4 de nuestra definición de T-satisfacibilidad, que 4) $\mathcal{E}$, ¡s$\nvDash Kj\alpha$! Pero esto se contradice con 2). Por lo tanto, $\vDash^T K_j\alpha\rightarrow\alpha$.

La prueba anterior establece la validez de axioma T o, mejor dicho, que cualquiera de sus instancias es una fórmula T-válida.

3.3. Metateorema de corrección para T

Con esta semántica también es posible demostrar los metateoremas de corrección y completez para la teoría T. A continuación, probaremos el primero de ellos[64].

Nótese, en principio, que todos los axiomas de T son T-válidos: recién establecimos que el axioma T lo es, y el resto de los axiomas, al ser válidos (*i. e.*, verdaderos en cualquier interpretación), también son T-válidos (*i. e.*, verdaderos en cualquier interpretación cuyas relaciones de accesibilidad son reflexivas). Por otra parte, las reglas de inferencia de la teoría (MP y KG) preservan T-validez: la regla MP preserva verdad y, en consecuencia, también preserva T-validez; mientras que para la

[62] En esta "nueva" semántica, la definición de verdad es la misma que para el sistema K; lo que le confiere su novedad esa variación en la noción de interpretación, *viz.*, que ahora las relaciones de visibilidad son reflexivas.

[63] Nótese además que todas las fórmulas válidas son también T-válidas, porque al ser verdaderas en cualquier interpretación lo son de igual forma en todas las interpretaciones cuyas relaciones de accesibilidad son reflexivas.

[64] Para la prueba del metateorema de completez, para esta teoría, véase el apéndice II.

regla KG vale el siguiente razonamiento. Si a una fórmula, digamos α, que es T-válida, se le aplica la regla KG, se obtiene la fórmula $K_j\alpha$, que también es T-válida, pues de lo contrario habría una T-interpretación $\mathcal{E}$ y un estado s de $\mathcal{E}$ tales que $\mathcal{E}, s \nvDash \alpha$, en cuyo caso la fórmula original α no sería T- válida. Por consiguiente, la regla KG también preserva T-validez. Tomando en consideración todo esto, estableceremos el metateorema de corrección para el sistema T:

Metateorema de corrección para T

Sea φ una *fbf* cualquiera $\mathcal{L}$-{K_1, ..., K_n}.

Todo teorema de Tes T-válido. Es decir, si $\vdash^T \varphi$, entonces $\vDash^T \varphi$.

Demostración

Demostración por inducción matemática sobre el número m de pasos de la demostración de φ.

Base: m=1 (la demostración de φ tiene un renglón)

Entonces φ se demuestra en un paso. Por lo tanto, φ es un axioma de T (A1-T). Pero sabemos que todo axioma de T es una fórmula T-válida. Por lo tanto, $\vDash^T \varphi$.

m= n, para n>1 (la demostración de φ consta de n renglones, con n>1)

HI: El metateorema vale para todas las fórmulas que aparecen en los renglones anteriores a n.

P.D. El condicional vale para φ (que aparece en el renglón n).

Hay tres casos para φ: φ es axioma de T, es consecuencia de anteriores por MP, o bien, es consecuencia de anteriores por KG:

Caso 1: φ es axioma de T. Se procede igual que en la base.

Caso 2: φ es consecuencia de anteriores por MP. Entonces φ se obtuvo, por MP, de las fórmulas α→φ y α, que aparecen en renglones anteriores de la demostración. Por consiguiente, a estas dos fórmulas les podemos aplicar HI, de modo que $\vDash^{\mathbb{T}} \alpha \to \varphi$ y $\vDash^{\mathbb{T}} \alpha$. Como la regla de inferencia MP preserva validez, tenemos que $\vDash^{\mathbb{T}} \varphi$.

Caso 3: φ es consecuencia de anteriores por KG. Entonces φ es de la forma $K_j\alpha$ y se obtuvo, por KG, de la fórmula α, que aparece en un renglón anterior de la demostración. En consecuencia, a esta fórmula se le puede aplicar HI, de forma que tenemos que $\vDash^{\mathbb{T}} \alpha$. Como la regla de inferencia KG preserva validez, tenemos que$\vDash^{\mathbb{T}} \varphi$.

De la base y los casos 1-3, se sigue, por principio de inducción matemática completa (PIMC), que:

Si $\vdash^{\mathbb{T}} \varphi$, entonces $\vDash^{\mathbb{T}} \varphi$.

3.4. El poder demostrativo del sistema $\mathbb{T}$

La teoría $\mathbb{T}$ hereda el poder demostrativo de la teoría anterior, la teoría $\mathbb{K}$, y lo desborda, pues en $\mathbb{T}$ se pueden demostrar fórmulas de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$ que no son demostrables en $\mathbb{K}$, *e. g.*, los teoremas 16, 17 y 18, a los que se aludió anteriormente en este capítulo. En otras palabras, los teoremas del sistema $\mathbb{K}$ constituyen un *subconjunto propio* del conjunto de teoremas del sistema $\mathbb{T}$. Esto significa que el axioma característico de este sistema, el principio de propiedad del conocimiento, no es redundante, ya que agrega propiedades al saber que no se hallan supuestas por la simple propiedad de distribución o axioma K; una de estas nuevas propiedades sería, por ejemplo, la consistencia del conocimiento. Por otro lado, ahora que hemos establecido el metateorema de corrección para $\mathbb{T}$, podemos responder a la pregunta que planteamos con antelación, la cual guarda relación con sus alcances demostrativos: ¿en $\mathbb{T}$ se demuestra cualquier instancia de la tesis KK, *i. e.*, del esquema de fórmulas $K_j\alpha \to K_jK_j\alpha$? La respuesta es que no, y, para percatarse de ello, basta con reparar en dos cosas. Según el metateorema de corrección, cualquier teorema de $\mathbb{T}$ es $\mathbb{T}$-válido, de modo que la $\mathbb{T}$-validez resulta necesaria para ser demostrable en el sistema; sin embargo, algunas instancias del esquema en cuestión no satisfacen esta característica. Veámoslo con una instancia concreta, la fórmula $K_1p_1 \to K_1K_1p_1$.

Considérese la siguiente T-interpretación.

$\mathcal{E} = \langle S, \pi, R_1, \ldots, R_n \rangle$

$S=\{r, s, t\}$

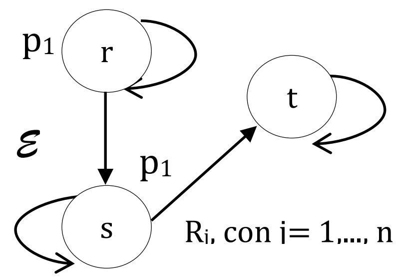
$\pi: L \longrightarrow P(S)$

$\pi(\psi)=\emptyset$, para toda letra proposicional ψ, excepto la letra proposicional p_1.

$\pi(p_1)=\{r, s\}$

$R_j= \{\langle r, r\rangle, \langle s, s\rangle, \langle t, t,\rangle, \langle r, s\rangle, \langle s, t\rangle\}$,

con $j= 1, \ldots, n$



Bajo esta interpretación se tiene que:

1) $\mathcal{E}, r \vDash p_1$, 2) $\mathcal{E}, s \vDash p_1$ y 3) $\mathcal{E}, t \nvDash p_1$. A partir de la relación de accesibilidad R_1 y la cláusula modal de nuestra definición de satisfacibilidad, se sigue, por (1) y (2), que 4) $\mathcal{E}, r \vDash K_1p_1$, mientras que por (3) se sigue que 5) $\mathcal{E}, s \nvDash K_1p_1$. Por otra parte, de (5), la relación R_1 y la misma cláusula modal, se tiene que 6) $\mathcal{E}, r \nvDash K_1K_1p_1$. Finalmente, de (4), (6) y la cláusula de la implicación, tenemos que 7) $\boldsymbol{\mathcal{E}, r \nvDash K_1p_1 \rightarrow K_1K_1p_1}$. Por consiguiente, esta fórmula no es T-válida debido a que no es verdadera en esta T-interpretación.

Lo anterior muestra que la fórmula $K_1p_1 \rightarrow K_1K_1p_1$ no es T-válida y, por tanto, dado el metateorema de corrección, que no toda instancia del esquema $K_j\alpha \rightarrow K_jK_j\alpha$ es teorema de T. En este sentido, el sistema T no contempla a la KK-tesis entre los principios o verdades lógicas acerca del conocimiento, y, por tanto, no reconoce la validez de la idea de que el conocimiento proposicional implica necesariamente un conocimiento proposicional de segundo orden. Sin embargo, no es claro si ello constituye una anomalía o una fortaleza de la teoría axiomática, debido a que la validez de la tesis KK es uno de los problemas más controversiales de la lógica epistémica. En el capítulo siguiente, al estudiar el sistema S4, abordaremos esta controversia, pero antes habrá que terminar el estudio del sistema T, y lo haremos examinando su supuesto epistémico fundamental, *viz.*, la idea de que la verdad es una condición necesaria para el conocimiento.

3.5. Consideraciones en torno al sistema T

Como señalé al comienzo de este capítulo, la Epistemología surgió con Platón, y a éste se le atribuye, al menos en Filosofía, la prestigiosa idea de que el conocimiento implica la verdad de lo conocido. Con todo, tal relación entre saber y verdad, la cual hemos capturado con el axioma T: $K_j\alpha \rightarrow \alpha$, no deja de tener sus inconvenientes. En esta sección, me dedicaré a señalar dos de ellos.

3.5.1. *La indistinguibilidad del saber*

En la *Metafísica*, Aristóteles ofrece la siguiente definición de “verdad”: << […] decir que lo que es es y que lo que no es no es, es verdadero>>[65]. En términos más contemporáneos: “p” es verdadera sii *p* [*i. e.*, sii *p* existe realmente], en donde “p” es una proposición arbitraria de un lenguaje, mientras que *p* es el hecho expresado por ella. Así definida, la verdad consiste en una relación objetiva entre dos componentes, *viz.*, lenguaje y realidad –específicamente, un enunciado y un hecho–. A esta relación nos referiremos con la expresión ‘verdad absoluta’, pues no requiere de ningún sujeto para existir; lo que tiene por consecuencia que una proposición puede ser verdadera aun si nadie lo sabe, más todavía, aun si nadie la cree o la juzga.

En el axioma T, supuesto fundamental de nuestro sistema axiomático T y de un número importante de análisis epistemológicos, la verdad implicada por el conocimiento es una verdad absoluta; por consiguiente, según este axioma, sólo las proposiciones verdaderas en sentido absoluto pueden ser conocidas. Una consecuencia de esto es que se vuelve problemático discernir las pretensiones de conocimiento legítimas de aquellas que no lo son. En otras palabras, si el conocimiento implica la verdad absoluta de lo conocido, entonces, dada una pretensión de conocimiento, nosotros (los seres humanos) no podremos distinguir si es legítima o sólo aparente, ya que somos incapaces de reconocer la verdad absoluta de las creencias o proposiciones.

[65] Aristóteles, *Metafísica* IV, 7, 26-29.

Para dar cuenta de esta conclusión, considérese el siguiente fragmento de la *Lógica*, de Immanuel Kant:

> Una perfección fundamental del conocimiento, condición sin duda esencial e inseparable de toda perfección del mismo, es la *verdad.* Verdad, dícese, consiste en la concordancia del conocimiento con el objeto. Según esta definición meramente nominal, mi conocimiento, para valer como verdadero, ha de concordar con el objeto. Ahora bien, sólo puedo comparar el objeto con mi conocimiento *en tanto que lo conozco.* Mi conocimiento debe ratificarse a sí mismo, lo cual, empero, no es ni con mucho suficiente para la verdad. Puesto que como quiera el objeto está fuera de mí y el conocimiento en mí, sólo puedo juzgar si mi conocimiento del objeto coincide con mi conocimiento del objeto[66].

En estas líneas, Kant critica la inclusión de la verdad absoluta como un componente insoslayable del saber: nunca sabemos si una pretensión de conocimiento satisface el requisito de la verdad, pues para ello habría que compararla con el objeto en sí (tal como éste es al margen de todo acto de conocimiento). Pero esto resulta imposible, porque para comparar las pretensiones de conocimiento con el objeto, este último debe ser conocido o aprehendido por nosotros. De modo que la comparación sólo es posible con el objeto de nuestra aprehensión, y no con el objeto en sí. Sin embargo, esto no es suficiente para determinar la verdad del pretendido conocimiento. Reconstruyamos esquemáticamente esta objeción.

El argumento de la indistinguibilidad

Definición de verdad
Una pretensión de conocimiento es verdadera si y sólo si concuerda con el objeto en sí mismo (con la cosa en sí).
Obs. 1: Si una pretensión de conocimiento es verdadera si y sólo si concuerda con el objeto en sí mismo, entonces, para poder determinar si las pretensiones de conocimiento son verdaderas, es necesario compararlas con el objeto en sí mismo.

---

[66] Kant, Immanuel, *Lógica. Acompañada de Una Selección de Reflexiones del Legado de Kant,* Akal (Ed.), Madrid, 2001, p. 111 (Introducción, § 7).

P1) Si no se comparan las pretensiones de conocimiento con el objeto en sí mismo, no se puede determinar si son verdaderas o no.

Obs. 2: Para comparar el objeto con las pretensiones de conocimiento es necesario conocerlo, aprehenderlo.

Obs. 3: Con lo único que es posible comparar las pretensiones de conocimiento es con el objeto conocido.

Obs. 4: El objeto conocido no es el objeto en sí mismo[67].

P2) Es imposible comparar las pretensiones de conocimiento con el objeto en sí mismo.

PI: No podemos determinar si las pretensiones de conocimiento son verdaderas o no [de P1) y P2), por MP]**.**

A partir de aquí, es relativamente sencillo llegar a la conclusión de que no somos capaces de discernir el saber auténtico del meramente aparente. Basta para ello con continuar el argumento de la manera siguiente:

Principio T: La verdad es una condición esencial para el conocimiento.

Obs. 5: Si la verdad es una condición esencial para el conocimiento, entonces, para saber si las pretensiones de conocimiento valen como conocimiento auténtico, es necesario determinar si éstas son verdaderas o no.

P3) Si no podemos discernir el valor de verdad de las pretensiones de conocimiento, entonces nunca sabemos si valen o no como conocimiento auténtico.

C: Nunca sabemos si las pretensiones de conocimiento valen como conocimiento auténtico o no [de PI) y P3), por MP]**.**

---

[67] Se descarta aquí toda forma de idealismo que identifique pensamiento y realidad. Una razón para ello es que al mantener la distinción entre el pensamiento y lo real se deja lugar a una posible desavenencia entre ellos, y es ésta, creo, la mejor explicación que tenemos del error.

En suma, si el saber implica la verdad absoluta de lo sabido, entonces los casos de conocimiento son indistinguibles de casos de no conocimiento[68].

Un corolario del resultado anterior es que nunca tenemos derecho a declarar que *sabemos*, sino simplemente que *creemos saber*, por lo que el concepto de *saber* sea prácticamente inaplicable o vacuo[69]. No obstante, se me ocurre que alguien podría cuestionar la relevancia de semejante objeción; después de todo, que nuestros poderes cognoscitivos no nos permitan determinar la autenticidad del saber no parece afectar en absoluto dicho saber: si una creencia –se dirá– cumple con las condiciones suficientes, incluida la verdad, entonces es un caso de conocimiento, y poco importa nuestra capacidad de discernimiento al respecto.

Quien así opina corre el riesgo de hacer de la Epistemología un juego puramente intelectual, uno en donde se busca forjar una noción perfecta de conocimiento, pero sin preocuparse en absoluto por sus condiciones de aplicación. Un punto de vista como el descrito olvida las raíces prácticas del acto de conocer: nos interesa distinguir, en la práctica, el conocimiento de la mera creencia porque, mientras que con la creencia se alcanza la verdad de manera azarosa (cuando se alcanza), con el saber se ata o sujeta, en la medida de lo posible, el pensamiento a lo real, *i. e.*, se tiene una garantía, aunque falible, de que la creencia alcanza la realidad y, por consiguiente, de que la acción se guía por ésta en caso de que se actúe conforme al saber. Por lo tanto, a la hora de actuar o decidir resulta de vital importancia el reconocimiento del saber.

Sin embargo, aún podría argüirse que dicha *atadura* en el conocimiento hace referencia a su justificación, y no directamente a la verdad. Por lo tanto, para guiar nuestras acciones basta con ser capaces de reconocer aquellas creencias con un grado

---

[68] Desde luego, hay casos de no conocimiento que podemos reconocer como tales, *v. g.*, cuando falta creencia o justificación. Sin embargo, lo que no podemos es reconocer que una creencia satisface, en efecto, la condición de verdad.

[69] Esta consecuencia de aceptar a la verdad absoluta como ingrediente del conocimiento ya la había señalado Luis Villoro en *Creer, saber, conocer*; específicamente, en el capítulo 8 ("Saber y verdad"), pp. 182-184.

de justificación adecuado al conocimiento –y conducirnos conforme a ellas, claro está-. En general, estoy de acuerdo con este punto de vista, pero entonces habrá que cuestionarse sobre la importancia real del conocimiento mismo, o bien, sobre la relevancia de incluir la condición de verdad en el concepto de *saber*. Parece como si los epistemólogos, a través de la condición de verdad, quisieran eliminar el error de la noción de *saber*; aun si con ello no se elimina la posibilidad de error en su aplicación, ni mucho menos la posibilidad de error a la hora de actuar con arreglo a aquello que creemos saber.

3.5.2. *La irrevocabilidad del saber*

Disponemos aún de una segunda objeción en contra del principio T: si la verdad absoluta se halla implicada en el conocimiento, entonces el conocimiento es irrevocable. Veamos por qué.

Desde muy temprano, el pensamiento filosófico reconoció que una característica fundamental de la verdad es que no puede devenir en falsedad; en otras palabras, lo verdadero lo es siempre sin importar qué. Es esto, me parece, a lo que se refiere Parménides con "el corazón imperturbable de la verdad bien redonda", la cual se trata en la vía de *lo que es siempre y carece de devenir*[70]. Así, por ejemplo, en caso de que fuera verdad que hay vida bajo la superficie congelada de Europa, luna de Júpiter, tal verdad permanecerá intacta sin importar lo que haya ocurrido antes y lo que sucederá después[71]. ¿Qué significado tiene esto en relación con el principio T? Consideremos un caso concreto de conocimiento.

[70] DK 28 B 1-8. Cf. G. S. Kirk, J. E. Raven y M. Schofield, *Los filósofos presocráticos*, Gredos (Ed.), Madrid, 2008.

[71] Se podría pensar, desde luego, que lo verdadero se muda en el tiempo, y que las verdades de hoy podrían no ser las verdades de mañana. En apariencia, un enunciado puede variar su valor veritativo en diferentes contextos. Así, un enunciado como 'el actual presidente de México tiene menos de 60 años', pronunciado en 2017, era verdadero; sin embargo, proferido ahora, en 2019, resulta falso. Existe más de una manera de explicar por qué esto no socava la invariabilidad de la verdad a través del tiempo, *v. g.*, distinguir entre proposiciones y enunciados, o entre declaraciones y oraciones; no obstante, recurriré

En la actualidad se sabe que la edad de la Tierra es de alrededor 4,500 millones de años, y no de apenas unos 6,000 años, como se pensaba antes[72]. En caso de aceptar el principio T, deberemos conceder que nuestra datación de la edad de la Tierra es una verdad absoluta. Es ésta la base de nuestra argumentación. Si tal datación es una verdad absoluta, entonces es completamente inmune a la refutación, ya que, en sentido estricto, no existe una auténtica refutación de algo verdadero, sino simplemente refutaciones aparentes de ello; la verdad, como diría Platón, no es difícil de refutar; <<más bien es imposible, pues la verdad jamás es refutada>>[73]. Por consiguiente, si concedemos la verdad absoluta de nuestra datación, lo conducente sería rechazar de antemano cualquier posible objeción futura, pues toda objeción en contra constituiría un intento de falsificar o poner en duda algo verdadero o, si se quiere, algo que *sabemos* que es verdadero. En este sentido, nuestro conocimiento resultaría simplemente irrevocable[74].

---

aquí al siguiente expediente: en nuestro ejemplo, se trata de enunciados distintos. Proferido en el 2017, el enunciado es en realidad 'quien ocupa el cargo de presidente en México, ahora, en el 2017, tiene a la fecha menos de 60 años'; proferido en el 2019, el enunciado es 'quien ocupa el cargo de presidente en México, ahora, en el 2019, tiene a la fecha menos de 60 años'. Estos enunciados no sólo son sintácticamente distintos, sino que también hablan acerca de personas distintas. En general, los enunciados aparecen en un contexto, el cual nos ayuda a precisarlos y, como en este caso, a distinguirlos. Como sea, nótese que, al margen de cualquier evento pasado o futuro, seguirá siendo verdad que quien ocupaba el cargo de presidente de México en el 2017 tenía menos de 60 años para entonces.

[72] Empleo la expresión 'se sabe' para indicar que se trata de un conocimiento bien establecido y con una aceptación generalizada. Por otra parte, respecto a la antigua datación de la edad de nuestro planeta, tenemos noticias de que, en el siglo XVII, el arzobispo James Ussher, empleando la genealogía bíblica a partir de Adán, calculó que el año de la creación había sido el 4004 a. C. Por su parte, el vicecanciller de la universidad de Cambridge, John Lightfoot, declaró que el hombre había sido creado el domingo 23 de octubre de ese mismo año, a las 9 am. *Cf.*, Bowler Peter J. (1989), *Evolution. The History of an Idea*, University of California Press, Berkeley, Los Angeles, London, cap. 1, p. 4.

[73] Platón, *Gorgias*, 473b.

[74] Hay aquí una doble irrevocabilidad. Una a nivel, digamos, ontológico, puesto que lo verdadero nunca se torna falso; una segunda a nivel epistémico, dado que los verdadero no puede jamás ser refutado. A esta segunda forma de irrevocabilidad le podemos llamar *incontrovertibilidad*.

El problema con esto es que resulta contrario a la tesis actual del conocimiento, al menos del conocimiento empírico, según la cual 1ro) tenemos conocimiento, y 2do) éste es falible y, por tanto, susceptible de refutación. Semejante tesis es consecuencia de la historia del conocimiento y se manifiesta, *v. g.*, en el hecho de que, al menos en ciencia y filosofía, la verdad se ha ido diluyendo como criterio de evaluación teórica, y ha sido sustituida por nociones tales como la *capacidad solución de problemas* o el *poder predictivo o explicativo* de una teoría[75]. Como el principio T, la falibilidad del saber también constituye una *éndoxa* u opinión reputada acerca del conocimiento. Por lo tanto, al ser incompatibles entre sí, lo que tenemos es un conflicto entre *éndoxa*[76].

Desde luego, no cualquier posición falibilista resulta inconsistente con la tesis o principio T; no necesariamente lo es un falibilismo epistémico que dependa exclusivamente de la condición de justificación del saber. Tal posición podemos expresarla como sigue:

Fabilismo de la Justificación: *Es posible, en principio, que un agente, en un momento determinado, crea justificadamente una proposición y la sepa en efecto, pero que posteriormente pierda su conocimiento al hallar, él mismo o los miembros de la comunidad epistémica pertinente, un socavador para su justificación inicial.*

Según FJ, el conocimiento es falible en el sentido de que es revocable, y su revocabilidad no le viene de la verdad, sino de la justificación; es ésta, y no la verdad, la que es susceptible de revocación. Este punto de vista implica, además, que la cláusula

[75] Otro campo en el que la noción de *verdad* ha ido perdiendo fuerza, aunque por razones un poco distintas, para ser remplazada por la de *consistencia* de una teoría, es la Matemática.

[76] Tomo el término *éndoxa* prestado de Aristóteles, para quien es un sinónimo de creencias u opiniones que suelen gozar de buena reputación, es decir, aquellas opiniones que tienen una aceptación total o general por parte de una comunidad, o que, al menos, tienen la aceptación de los más sabios o entendidos en un tema. Una manera más natural de entender el término *éndoxa* sería decir que involucra la manera general de ver (las opiniones de los grandes hombres o de personajes célebres) lo creído históricamente (incluso las creencias que se revelan en los ritos religiosos y los más antiguos mitos) y la experiencia colectiva de los hombres. En cierto sentido, hace referencia a las creencias que se establecen de un modo consuetudinario.

de justificación en el conocimiento no precisa, para ser satisfecha, de razones infalibles o inmunes a toda refutación futura, sino que le bastan simplemente las razones falibles, pero de una solidez considerable. No obstante, cabe preguntar si un falibilista adscrito a FJ aceptaría o no el principio T, ya que, sin ser inconsistente con ella, el principio T puede resultar problemático para la posición en cuestión. Veamos cómo.

Sea *j* un falibilista del tipo FJ, que suscribe el principio T. Supongamos que, en un momento t_1, *j* posee excelentes razones para su creencia en una proposición, digamos, *p* (*i. e.*, *j* está justificado en creer que *p*), y que sobre la base de estas razones él declara saber que *p* es verdadera. Concedamos también que su conocimiento es efectivo, que de hecho sabe que *p* es verdadera. Como *j* ha aceptado el principio T, deberá aceptar también que *p* es una verdad absoluta y, en consecuencia, invariable a través del tiempo. Avancemos en nuestro ejemplo y supongamos ahora que, más tarde, en un momento t_2, el conocimiento de *j* es revocado al encontrar un socavador para su justificación original. En tal caso, *j* no puede seguir sosteniendo legítimamente su conocimiento acerca de la verdad de *p*, pues carece de justificación; en t_1 sabía que *p* era verdadera, pero en t_2 ya no lo sabe. Sin embargo, dada su adscripción a FJ, su declaración de conocimiento en t_1 y su concesión del principio T, él está obligado a aceptar, en t_2, que *p* sigue siendo verdadera (y lo será siempre), aun cuando no lo sabe ni tiene razones para ello. Este resultado, por supuesto, tiene un aire problemático[77].

[77] Una solución sería asumir la indistinguibilidad del saber, *i. e.*, dar carta de ciudadanía al principio T y aceptar que los casos de conocimiento son indistinguibles de casos de no conocimiento. Entonces no sería posible formular el escenario planteado, pues nuestras razones no nos autorizarían para afirmar que sabemos, sino simplemente que creemos saber; así, cuando se desvanezca nuestra justificación para una pretensión de conocimiento dada, no nos veremos obligados a seguir sosteniendo su verdad, pues bien podría tratarse de un saber puramente aparente (*i. e.*, carente de verdad). Otra alternativa de solución –no aplicable en nuestro ejemplo dado su diseño, pero merecedora de exploración por sí misma– consiste en señalar que el socavador en t_2 de las razones de *j* muestra que éste nunca estuvo realmente justificado, sino sólo en apariencia, y que, por lo tanto, su saber en t_1 acerca de *p* era un falso saber. Bien visto, esto constituye otra forma de falibilismo no inconsistente con la tesis T (después de todo, si se trataba de un falso saber, no tenía por qué cumplir la condición de verdad); no obstante, no lo desarrollaré en este trabajo.

Otra forma de falibilismo no incompatible con el principio T consiste en afirmar que el conocimiento es contingente, de modo que la creencia que lo constituye podría haber sido falsa. Para precisar la posición, consideremos la siguiente formulación del *argumento de la posibilidad de error*, un argumento escéptico cuya presentación se encuentra al final del capítulo I[78]:

> P1) Para (casi) cualquier creencia que cualquier persona tiene acerca del mundo exterior, dicha creencia podría estar errada.
>
> P2) Si una creencia podría estar errada, entonces no es un caso de conocimiento.
>
> ∴
>
> Por lo tanto, (casi) cualquier creencia que cualquier persona tiene acerca del mundo exterior no es conocimiento (*i. e.*, nadie sabe nada, o casi nada, acerca del mundo exterior) (4-1), (4-2)[79].

Esta presentación del argumento es debida a Richard Feldman, en *Epistemology*. Allí mismo, en su respuesta al escepticismo, el autor hace notar que la premisa maestra del razonamiento anterior [*i. e.*, P2)] es lógicamente equivalente a:

> *P2*) Si j sabe p, entonces j no podría estar equivocado sobre p* [con *j* y *p* un agente y una proposición cualesquiera].

La declaración P2* es francamente ambigua, pero Feldman parece leerla como sigue: *si S sabe que p, entonces la creencia de S en p no podría ser falsa (i.e., dicha creencia es necesariamente verdadera).* En símbolos (de manera aproximativa):

$K_j p \rightarrow \Box p$[80].

[78] *Cf.*, La sección "Consideraciones en torno a la Semántica anterior", en capítulo I.

[79] Feldman, Richard, "Skepticism (I)", *Epistemology*, New Jersey, Prentice Hall, 2003, pp. 114-115.

[80] Aquí el símbolo '□' es el operador modal de necesidad de la lógica alética. Ahora, tal simbolización es sólo aproximativa porque con P2* se afirma que la creencia implicada en el conocimiento es necesaria, mientras que en la simbolización la necesidad está ligada directamente a la proposición creída. No obstante, que una creencia sea necesariamente verdadera no significa que la proposición creída lo sea:

Con esta interpretación o lectura de la afirmación P2*, Feldman se aventura a responder al argumento escéptico de la posibilidad de error mediante una forma de falibilismo que rechaza justo esta declaración y, consecuentemente, su equivalencia lógica P2. Es a esta forma de falibilismo a la que me he referido antes, y consiste en sostener que el conocimiento es compatible con la *posibilidad* de error;

> Ellos [los falibilistas] piensan que el saber es compatible con la *posibilidad* de error. Es importante dejar claro lo que están diciendo. Los falibilistas no están afirmando que el conocimiento sea compatible con el error actual [de *facto*]. No están declarando que tú puedas saber algo que no es verdadero. Tampoco están diciendo que tú puedas tener conocimiento cuando tienes razones positivas para creer que estás incurriendo en un error. Ellos afirman que el saber requiere de una justificación sólida, así como de la verdad. Por tanto, si crees algo sobre la base de excelentes razones, es verdadero, y no se trata de un caso Gettier, entonces posees conocimiento. Si, por el contrario, crees algo sobre la base de esas mismas razones, y resulta que ese algo es falso porque eres víctima de algún tipo de engaño o alucinación, entonces, en ese caso, no posees conocimiento[81].

En otros términos, según esta forma de falibilismo, nuestras creencias no requieren ser *necesariamente* verdaderas para constituir conocimiento: mucho de lo que sabemos es contingentemente verdadero, *i. e.*, es verdadero en efecto, pero podría haber sido falso. En símbolos (una vez más, de manera aproximativa):

$K_j p \rightarrow p \land \neg(K_j p \rightarrow \Box p)$.

Como se ve claramente, este tipo de falibilismo, lejos de ser inconsistente con el principio T, lo presupone. A este tipo de falibilismo le podemos denominar *falibilismo puramente potencial*, pues bajo él nuestro conocimiento nunca corre el riesgo real de ser erróneo o falso, sino que admite simplemente una posibilidad de error que no se actualizó y que, de hecho, nunca lo hará[82].

---

por ejemplo, la creencia de que hay creencias no podría ser falsa y, en ese sentido, es necesaria; pero la proposición 'hay creencias' es meramente contingente.

[81] Feldman, Richard, "Skepticism (I)", *Epistemology*, New Jersey, Prentice Hall, 2003, pp. 123-124.

[82] A diferencia de Feldman, hay quien considera que semejante postura no es falibilista, sino una consecuencia del falibilismo: si todas o algunas de nuestras creencias podrían estar erradas, entonces

¿Qué forma de falibilismo, entonces, resulta incompatible con el principio T? Aquélla según la cual algunos (cuando no cualquiera) de nuestros conocimientos actuales podrían ser refutados en el futuro a la luz de nueva evidencia, en cuyo caso llegaríamos a la conclusión de que nuestras razones ya no son suficientes para justificar nuestra creencia, o bien, que *las creencias o proposiciones involucradas, después de todo, eran falsas* (o se presentan como tales dada la nueva evidencia). Esta posición no equivale a sostener que los agentes pueden, consistentemente, pretender saber algo y considerarlo falso al mismo tiempo, pues con ello se concedería una contradicción: que un agente puede, simultáneamente, creer una proposición (en tanto que considera saberla) y no creerla (en la medida que la tiene por falsa). En lugar de esto, esta forma de falibilismo sostiene, como también lo hace el falibilismo puramente potencial, que el saber es compatible con la posibilidad de error; sin embargo, a diferencia del falibilismo potencial, no se trata aquí de una posibilidad meramente contrafáctica, sino de una que se puede concretar en los hechos mismo: nuestro conocimiento bien podría revelarse como falso en el futuro[83]. Una vez más, podemos esbozar una simbolización de esta posición mediante las herramientas de la lógica modal epistémica y alética:

$\Diamond(K_j p \wedge \neg p)$[84].

Nótese que la fórmula anterior es justo la negación de ésta otra: $\Box(K_j p \rightarrow p)$[85], que es, a mi juicio, la mejor expresión de la tesis T, pues tanto la Epistemología

éstas deben ser contingentes, pues, de lo contrario, sería mentira que pudieran resultar falsas. *Cf.*, Hurtado, Guillermo, *Por qué no soy falibilista*, CRÍTICA, Revista Hispanoamericana de Filosofía, Vol. XXXII, No. 96 (diciembre 2000): 59–97.

[83] En esta versión del falibilismo, como en las dos anteriores, se acepta que nuestras razones para saber no tienen por qué ser absolutamente contundentes. Más aún, en este tipo falibilismo se parte de la idea de que la justificación siempre deja abierta una posibilidad de error y, por ello, nuestro conocimiento es susceptible de corrección o refutación.

[84] El símbolo '◊' corresponde al operador de posibilidad de la Lógica Alética, y se define a partir del operador de necesidad al flanquear con dos negaciones (¬☐¬).

[85] Es ésta, por cierto, otra lectura posible de la afirmación P2*: *If j knows p, then j could not be mistaken about p*, en el argumento escéptico de la posibilidad de error.

contemporánea como la Lógica Epistémica la consideran un principio o axioma del saber y, en consecuencia, una *verdad necesaria* acerca del mismo. Esta fórmula muestra, por lo tanto, que la posición falibilista por ella esbozada es inconsistente con la idea de que el conocimiento implica la verdad de lo conocido.

Por último, hay que decir que semejante forma de falibilismo se halla profusamente desarrollada en el capítulo 8 del libro *Creer, Saber conocer.* En él, Luis Villoro establece que las razones para saber, *i. e.*, razones objetivamente suficientes, son falibles pero incontrovertibles: deben ser tales que nos permitan inferir la ausencia, al interior de nuestra comunidad epistémica, de razones suplementarias o adicionales que puedan socavar nuestra pretensión de conocimiento. Esto liga el conocimiento a un horizonte espacio-temporal (o comunidad epistémica) en donde cierto conjunto de evidencia tiene valor justificatorio y es suficiente para saber; sin embargo, ello no implica que en otro horizonte espacio-temporal, con más evidencia disponible, dicho conjunto conserve su valor probatorio o garantice nuestro conocimiento. Así, podría ocurrir que un agente *i* , en un horizonte o comunidad epistémica C_1, diga la verdad al aseverar que *i* (él mismo) sabe una proposición *p*, ya que su evidencia le permite inferir la ausencia, en C_1, de razones suplementarias que socaven su saber; mientras que un agente *j*, situado en el horizonte C_2, en donde se tiene evidencia suplementaria no disponible para *i* desde C_1, tiene razón en afirmar que *i* no sabe *p*, sino que *sólo cree* saberlo, debido a que *j* dispone de razones suplementarias que socavan el conocimiento de *i* , a las que, como se dijo, éste no puede tener acceso desde su propio horizonte o comunidad epistémica. Todo esto es posible, como el propio Villoro aclara, porque el conocimiento no implica la verdad de lo conocido: <<Los enunciados “S sabe que *p*” y “*p* es falso” son inconsistentes aseverados por la misma persona en el mismo momento, pero no lo son, afirmados por personas que pertenecen a diferentes comunidades epistémicas>>[86].

[86] Villoro, Luis, *Creer, saber, conocer,* Siglo XXI (Ed.), México, 1989, p. 194.

4. Sistemas de Lógica Epistémica (los sistemas S4 y S5)

4.1. El sistema S4

El tercero y penúltimo sistema axiomático que analizaremos aquí es el sistema S4, cuya formulación original (para la Lógica Modal Alética) es debida a C. I. Lewis, mientras que su interpretación epistémica es atribuida usualmente a Jaakko Hintikka[87]. Su axioma fundamental, *viz.*, el principio de introspección positiva o KK-tesis, declara que nuestro conocimiento proposicional implica siempre un conocimiento proposicional de segundo orden, *i. e.*, un conocimiento acerca de nuestro propio conocimiento. Puesto de otro modo, con este axioma se afirma que no conocemos proposición alguna a menos que sepamos que conocemos dicha proposición. En el sistema S4, esto se refleja con el esquema axiomático A4: $K_j\alpha\rightarrow K_jK_j\alpha$. Al igual que sucede con el sistema T respecto a la teoría K, S4 también es una extensión de la teoría axiomática anterior, la teoría T, pues se obtiene de ésta al agregar simplemente el axioma A4, razón por la cual preserva su poder demostrativo (todo lo demostrable en T será igualmente demostrable en S4), así como las tres tesis epistemológicas que hemos analizado con antelación, *i. e.*, los dos supuestos del sistema K y el supuesto adicional de T. De forma más esquemática, éstos serían los axiomas y reglas de inferencia de S4:

Axiomas de S4

Sean α, β y γ *wff*'s cualesquiera de $\mathcal{L}$-$\{K_1,\ldots, K_n\}$ y sea j=1, …, n.

A1: $\alpha\rightarrow(\beta\rightarrow\alpha)$,

A2: $(\alpha\rightarrow(\beta\rightarrow\gamma))\rightarrow((\alpha\rightarrow\beta)\rightarrow(\alpha\rightarrow\gamma))$,

A3: $(\neg\alpha\rightarrow\neg\beta)\rightarrow((\neg\alpha\rightarrow\beta)\rightarrow\alpha)$,

K: $K_j(\alpha\rightarrow\beta)\rightarrow(K_j\alpha\rightarrow K_j\beta)$ [Propiedad de distribución],

T: $K_j\alpha\rightarrow\alpha$ [Propiedad de conocimiento],

A4: $K_j\alpha\rightarrow K_jK_j\alpha$ [Propiedad de introspección positiva].

---

[87] En realidad, la contribución de Hintikka a este respecto consiste en un desarrollo semántico para una lógica bimodal de conocimiento y creencia, una parte del cual corresponde a lo aquí será referido como la semántica adecuada para el sistema S4.

Reglas de inferencia de S4

Sean α y β *fbf's* cualesquiera de $\mathcal{L}$-$\{K_1,\ldots, K_n\}$.

MP: Si $\vdash^{S4} \alpha$ Y $\vdash^{S4} \alpha\rightarrow\beta$, entonces $\vdash^{S4} \beta$ [*Modus ponens*] y

KG: Si $\vdash^{S4} \alpha$, entonces $\vdash^{S4} K_j\alpha$ [Generalización del conocimiento].

Entre los teoremas más importantes de S4 que no se demuestran en los sistemas anteriores (K y T) se encuentra un teorema de equivalencia, según el cual nuestro conocimiento proposicional de primer orden [el agente *i* sabe que α] es equivalente a un conocimiento proposicional de segundo orden [el agente *i* sabe que *i* sabe que α][88]. Así, si aplicamos este teorema, por ejemplo, al caso que presentamos hacia el final del capítulo I, en donde Alice y Bob reciben cada uno un número natural distinto y, al cabo de un breve conversación, terminan averiguando el número del otro, entonces afirmar que *Bob sabe que él mismo sabe que posee el número 3* sería simplemente una forma rebuscada de afirmar que *Bob sabe que posee el número 3*, ya que estas afirmaciones serían equivalentes y, en consecuencia, estarían expresando lo mismo[89]. En general, este teorema de equivalencia declara que todo conocimiento proposicional, sin importar de qué orden sea, es equivalente a un conocimiento proposicional de primer orden. Podemos capturar esta equivalencia con el esquema T19: $K_j\alpha\leftrightarrow K_jK_j\alpha$, y el hecho de que cualquier instancia de éste sea demostrable en S4 nos dice justamente que el sistema autoriza o reconoce tal equivalencia como uno de los principios del conocimiento proposicional[90].

Otro resultado importante que se obtiene en S4, y que no está presente en los sistemas anteriores, es la regla de inferencia que se muestra a continuación:

KK: Si $\vdash^{S4} K_j\alpha\rightarrow\beta$, entonces $\vdash^{S4} K_j\alpha\rightarrow K_j\beta$.

---

[88] Nótese que en este teorema, así como en el axioma A4, el conocimiento de primer orden y el conocimiento de segundo orden hacen referencia al mismo individuo: el agente *i*.

[89] *Cf.*, la sección "Consideraciones en torno a la Semántica anterior", en el capítulo I.

[90] La demostración de este teorema en S4 es sumamente sencilla y dejo al lector el trabajo de realizarla por su cuenta.

Con esta regla se afirma que si nuestro conocimiento proposicional (digamos, nuestro conocimiento de α) implica o conlleva la verdad de una proposición (*exempli gratia*, de β), entonces carecemos de dicho conocimiento proposicional a menos que sepamos que, en efecto, la proposición implicada es verdadera. De hecho, podemos establecer algunas relaciones significativas entre esta regla y el principio de introspección positiva. Por ejemplo, se puede probar el siguiente par de afirmaciones:

> Sea S cualquier sistema cuya fuerza demostrativa es exactamente la misma que la del sistema K ($\vdash^{S} \alpha$ si y sólo si $\vdash^{K} \alpha$, con α cualquier *fbf* de $\mathcal{L}$-{K_1, ..., K_n}).
>
> I) Si a S se le agrega el axioma A4, entonces en el sistema resultante (S+A4) se obtiene la regla de inferencia KK.
>
> II) Si a S se le agrega la regla de inferencia KK, entonces en el sistema resultante (S+KK) se obtiene como teorema a A4.

Veamos primero la demostración de la afirmación I).

> Supongamos que $\vdash^{S+A4} K_j\alpha\rightarrow\beta$. Entonces existe una demostración en S+A4 que termina en la fórmula '$K_j\alpha\rightarrow\beta$'. Podemos entonces continuar la demostración como sigue y llegar a la fórmula '$K_j\alpha\rightarrow K_j\beta$':
>
> --------
>
> m. $K_j\alpha\rightarrow\beta$
>
> m+1. $K_jK_j\alpha\rightarrow K_j\beta$ *m, RD1*
>
> m+2. $K_j\alpha\rightarrow K_jK_j\alpha$ *A4*
>
> m+3. $K_j\alpha\rightarrow K_j\beta$ *m+2, m+1 Transitividad*

Veamos ahora la demostración de la afirmación II).

> $\vdash^{S+KK} K_j\alpha\rightarrow K_jK_j\alpha$
>
> 1. $K_j\alpha\rightarrow K_j\alpha$ *Identidad*
> 2. $K_j\alpha\rightarrow K_jK_j\alpha$ *1, KK*

En conjunto, este par de afirmaciones muestra que el principio de introspección negativa (la tesis KK) y la regla inferencia KK son, en cierto modo, equivalentes[91], pues a partir de cualquiera de ellos se puede llegar a la demostración del otro. Esto último, así como el teorema de equivalencia (T19) que mencionamos hace un momento, será de gran importancia más adelante, cuando se discuta la validez de la tesis KK, por ello pido al lector que lo conserve en la memoria hasta entonces.

De manera análoga a como procedimos con los sistemas anteriores, indagaremos ahora sobre los límites demostrativos de la teoría S4. Una cuestión de interés a este respecto es si en S4, con la introducción del axioma de introspección positiva, se demuestra el así llamado *principio de introspección negativa*, *i. e.*, el principio según el cual los agentes epistémicos no son ignorantes de su propia ignorancia, pues siempre que un agente no sabe una proposición (la ignora), sabe que no la sabe. Esto queda capturado por el esquema "$\neg K_j\alpha \rightarrow K_j\neg K_j\alpha$". Desde luego, cualquier instancia del esquema converso a éste ($K_j\neg K_j\alpha \rightarrow \neg K_j\alpha$) es teorema de S4, ya que se trata simplemente de instancias del axioma T. No obstante, ¿cualquier instancia del esquema $\neg K_j\alpha \rightarrow K_j\neg K_j\alpha$ es demostrable en ese sistema? Tal como hicimos con las teorías anteriores, responderemos a esta interrogante acerca de los límites de S4 con ayuda del metateorema de corrección; y para la formulación y prueba de éste, hallaremos menester definir una semántica que se adecue a nuestra teoría.

4.2. Semántica y metateorema de corrección para S4

Previamente, al explorar el poder demostrativo del sistema T, establecimos que el principio de introspección positiva ($K_j\alpha \rightarrow K_jK_j\alpha$) no es T-válido, pues mostramos una T-interpretación que hacía falsa a una de sus instancias[92]. Esto significa que la semántica

---

[91] A este respecto, *cf.,* Hintikka, Jaakko, "Knowing that One Knows' Reviewed", *Synthese* 21, 141-162, 1970.

[92] Ver la sección "El poder demostrativo del sistema T", en este mismo capítulo. Por otra parte, nótese que el principio de introspección positiva tampoco es válido *a secas*, pues toda T-interpretación es una interpretación así sin más (*i. e.*, una estructura sin ninguna restricción sobre las relaciones de accesibilidad).

que definimos, y que resultó adecuada para el sistema T, no rescata la validez de dicho principio. Lo que haremos ahora será definir una nueva semántica para nuestro lenguaje formal $\mathcal{L}$-{K_1,..., K_n}, la cual nos permitirá capturar la validez del principio de introspección positiva, así como probar el metateorema de corrección para S4. Veamos, pues, esta nueva semántica.

Definición de S4-interpretación para $\mathcal{L}$-{K_1, ..., K_n}

Una S4-interpretación ($\mathcal{E}_{S4}$) para $\mathcal{L}$-{K_1, ..., K_n} es un tuplo:

$\mathcal{E}$= <S, π, R_1,..., R_n>.

Donde:

1) S≠Ø S es un conjunto no vacío (llamado el conjunto de estados s de $\mathcal{E}$).

2) π: L ⟶ P(S) π es una función que va de L al conjunto potencia de S, donde L es el conjunto de letras proposicionales de $\mathcal{L}$-{K_1, ..., K_n} ($\pi(p_i)\subseteq S$).

3) Para cada j=1, ..., n, $R_j\subseteq S^2$ es una relación reflexiva [*i. e.*, para cualquier s∈S se tiene que $<s, s>\in R_j$] y transitiva [*i. e.*, para cualesquiera s, t, u∈S se tiene que si $<s, t>\in R_j$ y $<t, u>\in R_j$, entonces $<s, u>\in R_j$] (R_j se llama la relación de accesibilidad o posibilidad de j).

Definición de S4-verdad para $\mathcal{L}$-{K_1, ..., K_n}

Nuestra noción de satisfacibilidad para S4 es esencialmente la misma que para T; por ello, omitiremos la repetición de las cláusulas correspondientes[93].

Definición de S4-verdad:

σ es verdadera en una S4–interpretación $\mathcal{E}$ sii para todo estado s de $\mathcal{E}$ se tiene que $\mathcal{E}$, s⊨σ.

[93] *Cf.* , la sección "Semántica formal para el sistema T", en este mismo capítulo.

Notación: $\mathcal{E}\vDash\sigma$

Definamos ahora lo que significa que σ sea una fórmula S4-válida de nuestro lenguaje.

Definición de S4-validez:

σ es S4-válida sii para toda S4–interpretación $\mathcal{E}$ se tiene que $\mathcal{E}\vDash\sigma$.

Notación: $\vDash^{S4}\sigma$

La semántica anterior es casi un equivalente de aquella que definimos para el sistema T, con la salvedad de que varía en la noción de interpretación, pues además de la reflexividad, ahora se exige también que las relaciones de accesibilidad entre los mundos sean relaciones transitivas, es decir, *si un mundo dado tiene acceso a algún mundo y éste, a su vez, tiene acceso a un tercer mundo, entonces el primero también tiene acceso a este último.* Pero ¿qué conlleva esta nueva restricción? Si un agente epistémico *j* se halla situado en un estado o mundo s, entonces, por esta nueva restricción de transitividad, toda la información disponible desde cualquier alternativa epistémica s' de s (*i. e.*, los estados accesibles para *j* desde s) no es más que un subconjunto de la información disponible para *j* desde s. Esto significa que ya en s el agente *j* tiene acceso a toda la información que tendría si se situara en cualquiera de las alternativas epistémicas s'. Con ello se consigue lo siguiente: si *j* conoce una proposición desde el mundo "real" s, entonces también conocería dicha proposición desde cualquiera de las alternativas epistémicas s' de s. En otras palabras, la restricción de transitividad sobre las relaciones de accesibilidad garantiza que el conocimiento de los agentes se preserve a través de las alternativas o mundo posibles que les son concebibles con base en la información de la que disponen. Si leemos el axioma A4: $K_j\alpha\rightarrow K_jK_j\alpha$ a la luz de la cláusula modal (cláusula 4) de nuestra definición de satisfacibilidad, veremos que es precisamente esto lo que se declara con él. Más aún, si atendemos a la demostración que ofrecimos anteriormente, con la que se estableció que el principio de introspección positiva no es T-válido, veremos que la interpretación que hace falsa a una de sus instancias no es transitiva; por ello, el principio pudo fallar[94].

En contraste con lo anterior, con esta nueva semántica probaremos la validez del axioma A4 o, dicho de otro modo, probaremos que cualquier instancia del esquema $K_j\alpha\rightarrow K_jK_j\alpha$ resulta S4-válida.

$\vDash^{S4} K_j\alpha\rightarrow K_jK_j\alpha$.

Demostración por reducción al absurdo:

> Supongamos que $\nvDash^{S4} K_j\alpha\rightarrow K_jK_j\alpha$. Es decir, que existe una S4-interpretación $\mathcal{E}$ y un estado s de $\mathcal{E}$ tales que:
>
> $\mathcal{E}$, s$\vDash K_j\alpha$ y 2) $\mathcal{E}$, s$\nvDash K_jK_j\alpha$.
>
> De 2) y cláusula (4) de nuestra definición de satisfacibilidad, se sigue que:
>
> 3) $\mathcal{E}$, t$\nvDash K_j\alpha$, con t algún estado de $\mathcal{E}$ tal que $<s, t>\in R_j$.
>
> Ahora, de la misma cláusula y 3), se tiene que:
>
> 4) $\mathcal{E}$, u$\nvDash\alpha$, con u un estado de $\mathcal{E}$ tal que $<t, u>\in R_j$.
>
> Como R_j es transitiva, de $<s, t>\in R_j$ y $<t, u>\in R_j$, se tiene que $<s, u>\in R_j$.
>
> De esto y la misma cláusula que empleamos con antelación, se sigue que:
>
> 5) $\mathcal{E}$, s$\nvDash K_j\alpha$**!** Pero esto se contradice con 1).
>
> $\therefore \vDash^{S4} K_j\alpha\rightarrow K_jK_j\alpha$.

Con esta semántica, además, se puede probar el metateorema de corrección para S4. En realidad, su demostración es completamente análoga a la que desarrollamos para la teoría T (establecer que todos los axiomas de la teoría son S4-válidos, que las reglas de inferencia preservan S4-validez, etcétera); razón por la que dejaré en manos del lector su elaboración y me limitaré simplemente a enunciar este metateorema para el sistema S4.

---

[94] Una vez más, *cf.*, la sección “El poder demostrativo del sistema T.

Metateorema de corrección para S4

Sea φ una *fbf* cualquiera $\mathcal{L}$-$\{K_1, \ldots, K_n\}$.

Todo teorema de S4 es S4-válido. En otras palabras, si $\vdash^{S4} \varphi$, entonces $\vDash^{S4} \varphi$.

4.3. El poder demostrativo del sistema S4

Ahora sabemos, por el metateorema de corrección, que todos los teoremas de S4 comparten una propiedad: son todos S4-válidos. Con ello responderemos a la interrogante acerca de si en S4 se demuestra cualquier instancia del esquema de fórmulas $\neg K_j\alpha \rightarrow K_j\neg K_j\alpha$. Bastará para ello con establecer que alguna de sus instancias no posee esta propiedad; si resulta que éste es el caso, entonces tal instancia no podría ser teorema de S4 (no sería demostrable en el sistema). Consideremos entonces la fórmula $\neg K_3p_1 \rightarrow K_3\neg K_3p_1$, que es una instancia del esquema, y calculemos su valor de verdad en la siguiente S4-interpretación.

$\mathcal{E} = \langle S, \pi, R_1, \ldots, R_n \rangle$

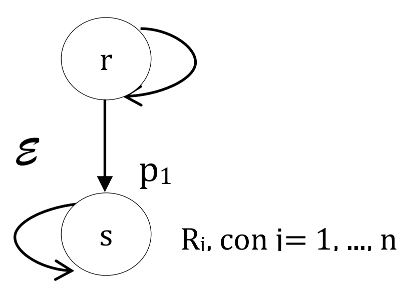
$S = \{r, s\}$

$\pi: L \longrightarrow P(S)$

$\pi(\psi) = \emptyset$, para toda letra proposicional ψ, excepto la letra proposicional p_1.

$\pi(p_1) = \{s\}$

$R_j = \{\langle r, r \rangle, \langle s, s \rangle, \langle r, s \rangle\}$, con $j = 1, \ldots, n$



Bajo esta S4-interpretación se tiene que:

1) $\mathcal{E}$, $r \nvDash p_1$, 2) $\mathcal{E}$, $s \vDash p_1$. A partir de la relación de accesibilidad R_3 y la cláusula modal de nuestra definición de satisfacibilidad, se sigue, por (1) y (2), respectivamente, que 3) $\mathcal{E}$,

$r\nvDash K_3p_1$ y 4) $\mathcal{E}, s\vDash K_3p_1$. De la cláusula de la negación, por (3) y (4), respectivamente, se obtiene 5) $\mathcal{E}, r\vDash\neg K_3p_1$ y 6) $\mathcal{E}, s\nvDash\neg K_3p_1$. Por otra parte de (6), la relación R_3 y la misma cláusula modal, se tiene que 7) $\mathcal{E}, r\nvDash K_3\neg K_3p_1$. Finalmente, de (5), (7) y la cláusula de la implicación, tenemos que 7) $\mathcal{E}, r\nvDash\neg K_3p_1\rightarrow K_3\neg K_3p_1$. Por consiguiente, esta fórmula no es S4-válida debido a que no es verdadera en esta S4-interpretación ($\boldsymbol{\mathcal{E}\nvDash\neg K_3p_1\rightarrow K_3\neg K_3p_1}$).

Con lo anterior queda claro que existe una S4-interpretación que hace falsa a la fórmula $\neg K_3p_1\rightarrow K_3\neg K_3p_1$ y, por consiguiente, que no toda instancia de $\neg K_j\alpha\rightarrow K_j\neg K_j\alpha$ [principio de introspección negativa] es una fórmula S4-válida. De esto, por el metateorema de corrección, se sigue que no toda instancia de este principio es teorema de la teoría S4, lo que es tanto como decir que la teoría no reconoce a la introspección positiva entre los principios del conocimiento. Sin embargo, esto no parece involucrar ningún problema para la teoría, pues, a diferencia de lo que acontece con el axioma de introspección positiva, la introspección negativa no es controversial sino francamente contraintuitiva[95]; es, por decirlo así, contraria al sentido común y a la tradición filosófica. El primero admite abiertamente que hay quienes creen saber lo que no saben, y que justo por ello están lejos de saber que nada saben en absoluto. La filosofía, por su parte, ofrece múltiples ejemplos en los que se transgrede el principio de introspección negativa. Así, la filosofía escéptica, a través de la acusación de dogmatismo, sugiere que los hombres se engañan en sus pretensiones de conocimiento; con frecuencia el escéptico ha tenido razón, y ha hecho falta su intervención para persuadirnos de nuestra ignorancia. También el arte de Sócrates sirve a este propósito, pues una de las tareas de la mayéutica consiste en llevar al autoconocimiento, específicamente, al conocimiento de la propia ignorancia; con ello se abre camino para la auténtica búsqueda de la verdad[96]. Si el principio de introspección negativa fuese verdadero, la

[95] Por ejemplo, en *Knowledge and Belief…*, Hintikka acepta la validez del principio de introspección positiva, pero rechaza la validez del principio de introspección negativa. *Cf.*, el capítulo V, "Knowing that one knows", del libro en cuestión.

[96] Para esta interpretación de la mayéutica socrática, ver W.K.C. Guthrie, *Historia de la filosofía griega: introducción a Aristóteles* (vol. VI), Gredos (Ed.), Madrid, 1981; específicamente, página 169.

actividad o misión socrática se tornaría, en gran medida, innecesaria, ya que todos los hombres estarían al tanto de su ignorancia. Todo esto hace que, lejos de una anomalía, la indemostrabilidad de $\neg K_j\alpha \rightarrow K_j\neg K_j\alpha$ en S4 pase por un acierto[97].

Antes de emprender el análisis de una última teoría axiomática, aquélla que parte justo de la introspección negativa cono axioma, habrá que examinar, tal como hicimos con K y T, el supuesto epistémico fundamental del sistema S4, *i. e.*, la tesis KK.

4.4. Consideraciones en torno al sistema S4

La tesis KK consiste en declarar que el conocimiento proposicional de orden n implica un conocimiento proposicional reflexivo de orden n+1: si un agente sabe una proposición, entonces él sabe que la sabe ($K_j\alpha \rightarrow K_jK_j\alpha$). En esta sección, abriré una discusión respecto a ella, la cual se nutre, en lo fundamental, de dos argumentos; un argumento en contra, así como un argumento a favor de esta tesis.

4.4.1. *El argumento del regreso epistémico*

El argumento del regreso epistémico funciona con cualquier análisis del saber que acepte a la justificación como una nota insoslayable del mismo, y la idea es relativamente sencilla:

La tesis KK tiene las siguientes consecuencias:

> *I) El conocimiento proposicional requiere conocer una sucesión infinita de proposiciones, cada una de las cuales es de un orden mayor a las anteriores*[98].

---

[97] En la siguiente sección, discutiremos algunas objeciones a la tesis KK; no obstante, éstas también sirven para criticar el principio de introspección negativa.

[98] En realidad, la tesis KK exige lo siguiente:

Definimos $K_j^n\alpha$ como $\underbrace{K_jK_j \ldots K_j}_{\text{n-veces}}\alpha$, para $n \geq 0$.

Si $K_j^n\alpha$, entonces $K_j^{n+1}\alpha$, para cualquier $n \in \mathbb{N}$.

II) Dada la condición de justificación, saber una proposición requiere, a su vez, que se tenga una justificación para cada una de las proposiciones de esta sucesión infinita.
No obstante, las consecuencias I) y II), además de imponer exigencias insatisfacibles al saber, son falsas. Por lo tanto, también lo es la tesis KK.

Para comprender mejor el argumento, recurriremos a un caso concreto de conocimiento proposicional. Imaginemos a Simone, una pequeña niña quien aprendió en la escuela que Cristóbal Colón llegó a América el 12 de octubre de 1492. Intrigada por el asunto, al llegar a su casa, Simone complementó la información vista en clase con datos obtenidos de una enciclopedia, así como de fuentes en internet. En todas las fuentes que consultó, ella encontró la misma fecha para la llegada a nuestro continente: el 12 de octubre de 1492. Preguntémonos ahora, ¿Simone sabe que Cristóbal Colón arribó a América en 1492? Me parece que la respuesta más natural sería responder que sí. No obstante, de acuerdo con la tesis KK, si ella sabe esto (K_Sp), entonces debe poseer también un conocimiento reflexivo, pero ahora de segundo orden: Simone sabe que *ella misma sabe que Cristóbal Colón llegó a América en 1492* (K_SK_Sp); lo que implicaría, a su vez, un conocimiento reflexivo de tercer orden, y así *ad infinitum*. De modo que terminaremos con una sucesión infinita de proposiciones conocidas por ella, cada una de un orden superior a las anteriores. Por otra parte, como el saber requiere de justificación, la sucesión anterior supondría que Simone posee una justificación para todas y cada una de las proposiciones que en ella aparecen. Por ejemplo, Simone está justificada en creer que Cristóbal Colón llegó a nuestro continente en el año de 1492, ya que lo escuchó en clase, además de leerlo en internet y en una enciclopedia. Sin embargo, según la tesis KK, para saberlo no le bastaría con ello, sino que aún requeriría de una justificación para su creencia de segundo orden (Simone sabe que Cristóbal Colón…), así como de órdenes superiores. En resumen, la tesis KK nos obliga a conceder que Simone no sabe que Cristóbal Colón arribó a América en 1492, a menos que conozca

---

En este sentido, la tesis KK nos pide conocer infinitas proposiciones sólo de manera potencial, y no real o actual. Sin embargo, por economía verbal, expresaré esta idea diciendo simplemente que nos exige conocer un número infinito de proposiciones.

una infinidad de proposiciones y, en consecuencia, disponga de una justificación para cada una de ellas.

Ahora bien, la idea del argumento del regreso epistémico es justo que tal regreso o sucesión infinita es innecesaria para el conocimiento proposicional, pues en nuestro ejemplo, lo más natural y razonable sería declarar que Simone cree justificadamente, y sabe en efecto, que 1492 es la fecha de llegada de Colón a América, y que para ello no precisa de una serie infinita de saberes ni de justificación. Además, en caso de aceptar las consecuencias de la tesis KK, *i. e.*, I) y II), el conocimiento proposicional se tornaría imposible o, por lo menos, muy escaso: en muy pocos casos, cuando no en ninguno, podemos dar una justificación para cada una de las proposiciones de la serie infinita requerida por la tesis KK. Por consiguiente —concluiría el proponente del argumento del regreso epistémico—, la tesis en sí misma resulta falsa, además de absurda, y debemos rechazarla como un principio válido para el conocimiento proposicional[99].

El argumento del regreso epistémico tiene la característica de concordar con nuestra intuición; la idea de que el conocimiento precisa de una serie infinita de saberes y justificaciones se nos presenta como una *paradoxa*, *i. e.*, como una idea contraria al sentido común o que difícilmente alguien sostendría con seriedad. Ello se debe, en parte, a la dificultad asociada a la idea misma de poder justificar un número infinito de proposiciones para cada instancia de conocimiento proposicional. Hagamos un breve comentario de esta dificultad.

No es un secreto que el infinito supone una dificultad, no sólo para el sentido común, sino para el pensamiento en general. Por ejemplo, en filosofía, las paradojas de Zenón y el trilema de Agripa se han nutrido, en términos generales, del problema del infinito. Desde luego, existen ámbitos en donde el pensamiento ha logrado domeñar el infinito; *v. g.*, nosotros recién probamos que todos los teoremas del sistema S4, infinitos en número, comparten la propiedad de ser S4-válidos. ¿Cómo hicimos esto? A través de

[99] Por supuesto, asoma aquí también la posición escéptica, desde la cual ni Simone ni nadie, al parecer, tiene conocimiento proposicional —por muy intuitivo que nos parezca lo contrario—; y la tesis KK simplemente nos señalaría por qué.

una inducción matemática, *i. e.*, a través de un principio racional, según el cual, si en un conjunto infinito bien ordenado [con el mismo orden que los números naturales], el primero de sus elementos tiene una propiedad, y dado un elemento cualquiera del conjunto, siempre que todos los elementos anteriores a él tienen dicha propiedad, él también la tiene, entonces todos los elementos del conjunto tienen la propiedad en cuestión[100].

Cabe preguntar si la sucesión infinita de saberes requerida para el conocimiento proposicional por la tesis KK es susceptible de una inducción matemática, de manera tal que, a través de ella, se obtenga una justificación para cada una de las proposiciones de la sucesión. En caso de que querer practicar una inducción matemática en este caso, *v. g.,* sobre el número de operadores epistémicos que aparecen en la proposición, habrá que proceder del modo siguiente:

> Sean j un sujeto arbitrario y p cualquier proposición pretendidamente sabida por j.

Afirmación por demostrar:

> j está justificado en creer '$K_j^n p$', para toda $n \geq 0$.

Deberemos establecer, primero, nuestra base inductiva, *i. e.*, que j está justificado en creer que p [éste es el caso cuando $m=0$, con m el número de operadores epistémicos a la izquierda de p]; y luego establecer que siempre que j cree justificadamente la proposición $K_j^i p$, también cree justificadamente la proposición $K_j^{i+1} p$ [para $m=i$, con $i \geq 0$]. Quien ensaye esta vía enfrentará dos dificultades. Primero, una lectura bastante natural de la tesis KK sería comprender que ésta exige la justificación de '$K_j^n p$', para

---

[100] El anterior es el principio de inducción matemática completa. También existe un principio de inducción matemática simple, el cual se enuncia como sigue: dado un conjunto infinito con un buen orden y una propiedad cualquiera, digamos λ, si su primer elemento tiene la propiedad λ, y para cualquiera de sus elementos, si él tiene λ, el elemento siguiente también la tiene, entonces todos los elementos del conjunto tienen la propiedad λ. Es éste, y no el principio de inducción completa, el que se hallará presente en la discusión que vendrá a continuación.

cualquier n, como una condición necesaria para la justificación de *p*; en cuyo caso la inducción matemática resultaría completamente estéril para nuestro propósito, pues tan sólo para establecer el primer paso de ésta, *i. e.*, para probar la base inductiva (que *j* cree justificadamente *p*), se requeriría justificar todas y cada una de las proposiciones, infinitas en número, supuesta por KK; sin embargo, es justo esto último lo que intentamos obtener mediante la inducción. En otras palabras, si esta lectura de KK es correcta, entonces la inducción matemática, lejos de resolver nuestro problema, exige su solución desde el inicio.

Por otro lado, aún si damos por sentado que se puede justificar *p* al margen de la justificación de cualquier otra proposición supuesta por KK, la inducción matemática nos encara con una nueva dificultad, la cual, a mi parecer, constituye el corazón del argumento del regreso epistémico, así como de la problemática en torno a la tesis KK: para llevar adelante nuestra inducción, habrá que probar que *si el agente j tiene una justificación para la proposición* $K_j^i p$ [*i. e.*, con i operadores epistémicos], *entonces también la tiene para la proposición* $K_j^{i+1} p$. ¿Cómo probar semejante afirmación? Comencemos por notar que para algunas afirmaciones es posible probar este condicional. Si *j* prueba, *v. g.*, mediante inducción matemática, que todos los números naturales tienen la propiedad φ, entonces estará justificado en creer, y sabrá, la proposición 'todos los números naturales tienen φ'. Sin embargo, dada la naturaleza o solidez de su fundamento (una demostración por inducción matemática), estará también justificado para creer, y sabrá, la proposición '*j* sabe que todos los números naturales tienen φ', así como todas las proposiciones de órdenes superiores implicadas por la tesis KK. En este caso, justificar (y conocer) la primera proposición de la serie infinita supuesta por KK equivale a justificar (y conocer) cada una de las proposiciones de la serie. En otras palabras, en nuestro ejemplo se cumple que, si un agente cree justificadamente una proposición $K_j^i p$, también cree justificadamente la proposición de orden superior $K_j^{i+1} p$; más todavía, se cumple el condicional $K_j p \rightarrow K_j K_j p$, el cual es una instancia del teorema KK. Y todo ello no a partir de una serie infinita de justificaciones —algo, por lo demás, imposible de proporcionar—, sino a partir de una única justificación que vale, o se mueve, a través de todos los niveles de conocimiento supuestos por KK.

Aunque no siempre es posible dotar a nuestras pretensiones de conocimiento de una justificación que valga para toda esta cadena infinita de niveles de saber. Con objeto de mostrar lo anterior, recurriré a una segunda objeción en contra del axioma KK (la primera es esta que discutimos ahora, *i. e.*, el argumento del regreso epistémico), la cual fue formulada en 1978, por W. Lenzen. Esta objeción descansa en la lectura de KK como un axioma de *introspección*: <<A diferencia de la creencia, el saber no es puramente subjetivo en carácter, sino que requiere de al menos una marca objetiva, *viz.*, la verdad. Por tanto, no podemos —por mera introspección— averiguar si sabemos que p>>[101]. La crítica de Lenzen va dirigida originalmente al axioma A5: $\neg K_j\alpha \rightarrow K_j\neg K_j\alpha$ [axioma de introspección negativa], pero, dada su naturaleza, también es aplicable a la tesis KK [principio de introspección positiva].

A primera vista, Lenzen parece estar equivocado; existen proposiciones que podemos conocer por simple introspección. Se trata de una clase de proposiciones a las que denominaré *proposiciones de apariencia*, y son de la forma “me parece que *h”* con *h* cualquier hecho susceptible de ser expresado mediante una proposición. Así, si un agente dice de sí mismo ‘me parece que siento miedo’, le bastará con un mero ejercicio de introspección (*i. e.*, con inspeccionar detenidamente sus propios contenidos metales) para saber el valor de verdad de dicha proposición; ello es así porque tanto la verdad como la justificación de las proposiciones de apariencia dependen únicamente de la vida mental del agente, de sus apariencias[102]. Ahora bien, una lectura más

[101] Wolfgang, Lenzen, *Recent work in epistemic logic*, Acta Philosophica Fennica, 30:5–219, 1978, p. 79.

[102] Nótese que el enunciado ‘me parece que siento miedo’ no es equivalente al enunciado ‘siento miedo’, pues mientras que éste último puede ser refutado por una observación o análisis posteriores en donde se establezca que el agente en cuestión no experimentaba miedo en el momento de proferir este enunciado , sino que confundió, digamos, la angustia con el miedo, tal observación, aun si está en lo correcto, no refuta el enunciado de apariencia, ya que seguirá siendo verdad que al agente le parecía sentir miedo cuando pronunció el enunciado. Otra característica de las proposiciones de apariencia es que no tienen, digámoslo así, una validez intersubjetiva, ni van más allá del momento presente.

cuidadosa de la objeción planteada por Lenzen revela que lo anterior no refuta su posición, puesto que, después de todo, él hace referencia al conocimiento de segundo orden, no al de primer orden: <<Por tanto **no podemos —por mera introspección— averiguar si sabemos que p**>>. Quizá podamos saber una proposición mediante introspección; pero no podemos, a través de medios introspectivos, saber que la sabemos. Veamos.

Aun si un agente puede conocer por introspección una proposición de apariencia, no puede conocer de la misma manera la proposición epistémica, la cual es de la forma "sé que *p*", *v. g.*, 'sé que me parece que siento miedo'. Quien esto profiere no se limita a declarar sus apariencias, *i. e.*, que le parece saber algo; sino que afirma con ello que en efecto sabe algo, que su conocimiento es real. En otros términos, las proposiciones de la forma "sé que *p*" no son proposiciones de apariencia, y no pueden ser conocidas por mera introspección: a través de ésta, el agente podrá determinar si le *parece* o no saber una proposición, pero no podrá discernir si su conocimiento es sólo aparente o si, en cambio, se trata de un conocimiento efectivo; no podrá distinguir, como dice Sócrates, si su pretensión de conocimiento es un falso fantasma o algo dotado de vida y verdad. De modo que la introspección no lo justifica para creer que sabe una proposición y, por consiguiente, no le proporciona conocimiento de segundo orden[103].

Por su puesto, la objeción de Lenzen, como ya señalé, descansa en una interpretación introspectiva de la tesis KK, y nada nos obliga a sujetarnos a semejante interpretación –la cual, dicho sea de paso– se halla contraindicada en la literatura[104].

Más que la crítica de Lenzen en sí misma, me interesan los resultados alcanzados en relación con las proposiciones de apariencia: nuestra justificación para ellas, *i. e.*, nuestras apariencias verificadas por introspección, no es suficiente para obtener

[103] Una razón para ello es, como señala Lenzen, que el saber no se agota en las apariencias o creencias de los sujetos, sino que involucra características objetivas, como lo serían la verdad o la justificación objetiva.

[104] *Cf.*, El capítulo cinco de *Knowledge and Belief…,* de J. Hintikka: "Knowing that one knows"; específicamente, la sección 5.3.

conocimiento de segundo orden al respecto (para saber que sabemos estas proposiciones), ni para ningún conocimiento de orden superior[105]. Esto significa que, a diferencia de lo que ocurre con el conocimiento obtenido a través de una inducción matemática, en algunos casos la justificaciones o pruebas disponibles no son válidas para diferentes órdenes de conocimiento; en estos casos, no cabe esperar que una única prueba permita establecer el condicional 'si creemos justificadamente una proposición, entonces también creemos justificadamente que la sabemos', el cual, como señalé, constituye el corazón de la tesis KK. Esto significa que las condiciones requeridas para la justificación y conocimiento en un primer nivel no son siempre las mismas que se requieren en niveles superiores.

Cerraré mi comentario del argumento del regreso epistémico con este señalamiento: este argumento conforma una sólida objeción en contra del pretendido principio KK, pero no para el sistema S4 que lo contiene. ¿Por qué? Nuestro sistema trabaja con el axioma T: $K_j\alpha \rightarrow \alpha$, *i. e.*, con la idea de que el saber implica la verdad de lo sabido; con este axioma y la tesis KK: $K_j\alpha \rightarrow K_jK_j\alpha$, se demuestra en el sistema la fórmula '$K_j\alpha \leftrightarrow K_jK_j\alpha$'. Esto significa que la teoría S4 reconoce como equivalentes al conocimiento de primer orden (K_jp, con p cualquier fórmula o enunciado sin operadores epistémicos) y el conocimiento de órdenes superiores ($K_j^n p$, con $n>1$). En otras palabras, puesta en el contexto del sistema S4, la tesis KK no nos obliga a conceder una cadena infinita de órdenes o niveles de conocimiento como requisito para el conocimiento proposicional, ya que el sistema, gracias al axioma T, permite reducir cualquier conocimiento de orden superior a un conocimiento de primer orden[106]. De manera general, cualquier agente que comparta el principio T puede aceptar KK sin caer en el problema del regreso

[105] Estoy presuponiendo que nuestras apariencias (o, si se quiere, el acceso directo que tenemos a éstas) aun cuando no son la única justificación disponible para las proposiciones de apariencia, constituyen la mejor de ellas.

[106] Es claro que el sistema sólo garantiza la reducción cuando el agente epistémico de primer orden y de órdenes superiores es el mismo. En consecuencia, si un agente sabe que *otro* agente conoce una proposición, el saber del primero, aun siendo de segundo orden, no tiene por qué ser reductible, en el sistema, a un saber de primer orden.

epistémico. Además, la equivalencia '$K_j\alpha \leftrightarrow K_jK_j\alpha$', demostrable en S4, armoniza con una tradición nada despreciable en filosofía, de acuerdo con la cual nuestro conocimiento de que sabemos una proposición es idéntico o equivalente a nuestro conocimiento de esa proposición: <<tu saber que sabes sólo difiere en palabras de tu saber [...] 'Yo sé que sé' no significa nada más que 'Yo sé' [...]>>[107]. Esta coincidencia entre sistema y tradición puede contribuir a la aceptación de S4 y, a través de éste, a la aceptación de la tesis KK.

4.4.2. *Un argumento de J. Hintikka a favor de la tesis KK*

En 1987, vio la luz el volumen número 8 de una serie internacional de filósofos y lógicos contemporáneos, la cual estuvo a cargo de Radu J. Bogdan de la universidad de Tulane. Este volumen, dedicado al filósofo Jaakko Hintikka, incluye, además los trabajos acostumbrados acerca de las contribuciones del autor, una sección de respuestas y comentarios del propio Hintikka a cada uno de los artículos que conforman el volumen. En sus respuestas al artículo sobre lógica epistémica, "Hintikka's epistemic logic", debido a Kennet Collier, Hintikka ofrece una de sus últimas defensas de la tesis KK[108]. Cito *in extenso* su argumento:

> [...] Recientemente ha sido propuesto que el conocimiento puede ser definido como conocimiento *no socavable*. Sin embargo, las formulaciones usuales de esta idea en la

[107] Schopenhauer, Arthur, *Two Essays*, (tr. Mme Karl Hillebrand), London, 1891, p. 166; citado en Hintikka J., *Knowledge and Belief...*, p. 108. Allí mismo, Hintikka proporciona al lector referencias suficientes para rastrear la tradición antes aludida. No obstante, nosotros acabamos de ver un contraejemplo en el que el conocimiento de primer orden no es de ningún modo equivalente al conocimiento de órdenes superiores: el caso de las proposiciones de apariencia. Desde luego, el argumento que hemos presentado, que parte de la crítica de Lenzen, no nos impide identificar el saber de segundo orden con el de órdenes superiores; por lo que si bien las proposiciones de apariencia socavan la generalidad o validez de la equivalencia $K_j\alpha \leftrightarrow K_jK_j\alpha$, no socavan que a través de las tesis T y KK se detenga el regreso epistémico (pues seguiría siendo válida la equivalencia $K_jK_j\alpha \leftrightarrow K_jK_jK_j\alpha$).

[108] Anteriormente, Hintikka había abordado el problema de la tesis KK. Primero, en 1962, en *Knowledge and Belief...*; años más tarde, 1970, en un artículo titulado "Knowing that One Knows' Reviewed", *Synthese* **21**, 141-162.

literatura son todas defectuosas, pues hablan de la existencia o inexistencia de enunciados o proposiciones. El significado de esta cuantificación de orden superior queda sin especificar hasta que no se diga si se está presuponiendo una interpretación estándar o no estándar de los cuantificadores. Si se presupone una semántica estándar (puramente extensional), es fácil obtener resultados que son obviamente incompatibles con las intenciones de los filósofos que cuantifican sobre proposiciones; y si se asume una semántica no estándar, queda completamente abierto cuál interpretación no estándar, de un inmenso número de ellas, está siendo utilizada.

No obstante, la tesis KK puede ser considerada como una major realización de la ausencia de socavadores. Pues, ¿qué nos dice esta idea? Supóngase que α sabe que *p*, pero que su saber deja abierta la posibilidad de que *q*. O, mejor, supóngase que es éste un estado consistente de cosas. En breve, asumamos que

(1) $K_\alpha p \,\&\, {\sim}K_\alpha{\sim}q$

es consistente. Entonces, si *q* implicara que $\sim K_\alpha p$, el saber de α de que *p* podría ser socavado, *viz.*, al mostrar que *q* es el caso. En consecuencia, de acuerdo con la ausencia de socavadores, si (1) es consistente, entonces también debe serlo

(2) $K_\alpha p \,\&\, q$

Al traducir esto, tenemos la siguiente regla de inferencia:

$$(***)\ \frac{K_\alpha p \supset {\sim}q}{K_\alpha p \supset K_\alpha{\sim}q}$$

Pero si reemplazamos aquí $\sim q$ por *q*, obtenemos precisamente la regla de inferencia (**), la cual caracteriza la tesis KK. Por consiguiente, el sentido fuerte de conocimiento en el que (**) es válido es, con mucho, una mejor realización de la idea de la ausencia de socavadores que cualquier formulación presente en la literatura[109].

El argumento anterior depende esencialmente de la regla derivada de inferencia (KK): Si $\vdash^{S4} K_j\alpha\rightarrow\beta$, entonces $\vdash^{S4} K_j\alpha\rightarrow K_j\beta$ o, mejor dicho, de la *equivalencia* entre esta regla y

[109] Radu J. Bogdan (ed.), '*Jaakko Hintikka'*, D. Reidel Publishing Company, 1987, pp. 311-312.

la tesis KK: nosotros ya establecimos que prácticamente cualquier sistema axiomático con idéntico poder demostrativo que el sistema $\mathbb{K}$ permite probar la regla derivada (KK) a partir de la tesis KK, y viceversa[110]. La regla (KK) captura la idea de que, si nuestro conocimiento proposicional implica la verdad de alguna proposición, entonces, para que nuestro conocimiento sea efectivo, debemos saber que la proposición implicada es verdadera. El alegato de Hintikka es que la tesis KK, a través de (KK) [*stricto sensu*, a través de una de sus instancias], conduce a una mejor formulación, respecto a aquellas formulaciones "defectuosas" presentes en la literatura filosófica, del análisis del concepto de *saber* conocido como la *teoría de la ausencia de socavadores.* Para tener un referente de esta teoría, así como de los defectos aludidos por Hintikka, reproduciré aquí la versión que de ella hace Richard Feldman, en *Epistemology*[111]:

> [La teoría de la ausencia de socavadores]
> ND. S sabe p = df. (i) S cree p; (ii) p es verdadera; (iii) S está justificado en creer p; (iv) No hay una proposición verdadera t tal que, si S estuviera justificado en creer t, entonces S no estaría justificado en creer p. (Ninguna verdad socava la justificación de S para p).

Esta teoría es una modificación del análisis tradicional del saber, ya que, a las tres condiciones de éste, *viz.*, creencia, justificación y verdad, se agrega una cuarta condición: que no exista una proposición verdadera que socave la justificación del agente para la proposición que pretende conocer. La motivación de este nuevo análisis del saber en términos de ausencia de socavadores se encuentra en los contraejemplos tipo Gettier al análisis tradicional. Éstos muestran que las tres condiciones del análisis tradicional no bastan para saber; pero al agregar una cuarta cláusula, la de la ausencia de socavadores, se busca remediar esta falla, *i. e.*, se busca obtener un conjunto de condiciones necesarias y suficientes para el conocimiento proposicional. Sin embargo,

[110] Véase la "Introducción" a este capítulo. A diferencia de Hintikka, nosotros tenemos la regla KK indexada a un sistema formal; de modo que lo que nos dice es que, si una fórmula es teorema del sistema, otra fórmula también lo es. Sin embargo, para discutir el argumento de Hintikka, interpretaremos la regla tal como él lo hace, y para ello es necesario desligarla de cualquier sistema. Con ello, la regla quedará de la siguiente manera: Si $K_j\alpha\rightarrow\beta$, entonces $K_j\alpha\rightarrow K_j\beta$ [o, mejor aún: Si $K_j\alpha\vDash\beta$, entonces $K_j\alpha\vDash K_j\beta$].

[111] Ya hemos citado esta versión al comienzo del capítulo III.

esta versión de la teoría, a la que llamaremos la versión ND, está abierta a la objeción de Hintikka, según la cual en las formulaciones presentes en la literatura filosófica se suele cuantificar sobre proposiciones, mas sin especificar siquiera cómo se están interpretando los cuantificadores (de una manera estándar o no estándar). Además, también es objeto de contraejemplos con los que se muestra que la conjunción de las cuatro cláusulas mencionadas nos es necesaria para saber, pues existen casos patentes de conocimiento proposicional en los que no se cumple justo la cuarta cláusula. Veamos uno de estos contraejemplos[112].

Imaginemos que René, una mujer tranquila y chapada a la antigua, asiste un fin de semana a su cabaña en el bosque, lejos de la ciudad, para no ser molestada por nadie. En ella, no hay ningún servicio de telefonía ni mucho menos conexión a internet; con lo único que cuenta, además de unos muebles del siglo XVII, es con un viejo radio. La tarde del sábado, René se dispone a escribir frente a su chimenea, y decide apagar y desconectar la radio para evitar cualquier interferencia. Al mismo tiempo que ella redacta, la estación de radio 69.G está transmitiendo un concierto completo de Marin Mersenne, dedicado a su hija Francine. Evidentemente, René no tiene la menor idea de ello.

Notemos lo siguiente. Primero, René cree justificadamente (y sabe) que la radio se encuentra apagada mientras escribe, ya que ella misma la desconectó. Segundo, dado el contexto de René, la única forma disponible de enterarse de la transmisión del concierto es si la radio estuviese encendida y sintonizada en la estación correcta.

Es éste un contraejemplo a la teoría de la ausencia de socavadores, pues en él, René posee conocimiento, pese a que no se cumple la cuarta cláusula propuesta. René sabe, en efecto, que la radio está apagada; sin embargo, existe una proposición verdadera tal que si ella la creyera justificadamente, no estaría justificada en creer que la radio se encuentra apagada: si ella supiera o creyera justificadamente que el concierto de Mersenne está siendo transmitido por la estación 69.G, entonces, dado el

[112] Nuestro contraejemplo ha sido tomado, con ligeras variaciones, de Richard Feldman. *Cf.*, pp. 34-5 en *Epistemology*.

contexto, la radio tendría que estar encendida y sonando [por ser el único medio por el que podría enterarse]; pero entonces, no estaría justificada en creer que la radio se encuentra apagada. Nuestro contraejemplo se vale del hecho de que la cuarta cláusula se formula mediante un condicional subjuntivo o contrafáctico [*si una cosa fuera verdadera, entonces también lo sería otra*], por lo que es posible, en principio, que un agente tenga conocimiento efectivo y actual, mientras que existe, al mismo tiempo, una situación posible en la que, si las cosas fuesen diferentes a como de hecho son, el agente no sabría lo que en efecto sabe.

Como señalé antes, esto ilustra la teoría o idea epistémica de la ausencia de socavadores, así como algunas de las dificultades en su formulación, ambas aludidas por Hintikka[113].

El argumento de Hintikka a favor de la tesis KK será, pues, que ésta, mediante la regla (KK): Si $K_j\alpha \rightarrow \beta$, entonces $K_j\alpha \rightarrow K_j\beta$, permite una mejor aproximación a esta idea. Consideremos una instancia específica de esta regla:

> (KK'): Si $K_j\alpha \rightarrow \neg\beta$, entonces $K_j\alpha \rightarrow Kj\neg\beta$ [que se obtiene de (KK) al sustituir 'β' por '¬β'].

(KK') es la regla clave en la argumentación de Hintikka, quien la interpreta del modo siguiente: cuando nuestro pretendido conocimiento proposicional implica la falsedad de alguna proposición, ésta se convierte en un socavador potencial de nuestro saber, pues si ella fuera verdadera, entonces nuestro saber no lo sería. Por consiguiente, a menos que descartemos este posible socavador, *i. e.*, a menos que sepamos que la proposición en cuestión es de hecho falsa, no tendremos conocimiento proposicional. En resumen, la versión de la teoría de ausencia de socavadores que obtiene nuestro autor con la regla (KK') es la siguiente:

---

[113] El contraejemplo de René muestra que la cuarta cláusula, tal como se formuló, es muy fuerte o exige demasiado al conocimiento proposicional, pues exhibe que podemos saber una proposición aun cuando aquélla no se cumple.

Sean α y β proposiciones cualesquiera; y sea *j* un agente cualquiera.

HND: Si el conocimiento de j acerca de α ($K_j\alpha$) *implica la falsedad de β* ($\neg\beta$), *entonces, a menos que j sepa que β es falsa* ($Kj\neg\beta$), *no sabrá α* ($\neg K_j\alpha$).

No obstante, ¿esta versión (HND) es mejor que las formulaciones presentes en la literatura? ¿Es mejor, por ejemplo, que la formulación ND, que presentamos hace un momento? Reparemos en que la proposición 'la estación de radio 69.G transmite un concierto completo de Marine Mersenne', a diferencia de lo que sucede con ND, no constituye un socavador para el conocimiento de René en la formulación HND, ya que su falsedad no se haya implicada por el conocimiento de ésta[114]. De modo que esta nueva versión de la idea de la ausencia de socavadores es inmune a nuestro contraejemplo. Sin embargo, HND no logra esquivar la objeción general que se encuentra detrás del ejemplo de René, *i. e.*, HND también exige demasiado al conocimiento, pues existen casos de conocimiento en los que no se cumplen las condiciones impuestas por ella. Con el objeto de mostrar semejante crítica, recurriré a uno de los acertijos matemáticos más famosos de la historia, la *paradoja de Monty Hall*. Ésta figuró por vez primera en febrero de 1975, en la revista académica *American Statistician* —específicamente, en una carta al editor debida al matemático Steve Selvin,

[114] Aun si suponemos que René se encuentra en la situación descrita, y que la única forma de enterarse de la transmisión del concierto es a través de la radio, no obtendremos una objeción para esta versión de la teoría –que no incluye un condicional contrafáctico en su formulación–: René sabe que la radio está apagada, ella misma la desconectó. La proposición cuya falsedad se halla implicada por su conocimiento no es acerca de la transmisión del concierto, sino ésta otra: ' la radio se encuentra encendida'. Dado el conocimiento de René, ella sabe que esta proposición es falsa. Por lo tanto, la versión HDN de la teoría da el resultado correcto en este caso. Imaginemos ahora una situación contrafáctica en la que la radio está encendida y René escucha la transmisión del concierto. En esta situación, ella sabe que la radio está encendida y, por consiguiente, no sabe que está apagada [*i. e.*, ella no sabe que es falsa la proposición 'la radio está encendida' y, en consecuencia, no sabe que la radio está apagada]. Una vez más, la nueva versión de la teoría da el resultado correcto; sin embargo, lo importarte aquí es que, en esta versión, lo que sucede en una situación contrafáctica no afecta el saber actual o real del agente.

de la Universidad de California—[115]; quince años más tarde, el 9 de septiembre de 1990, reapareció en la revista *Parade*, en la columna de preguntas y respuestas de Marilyn vos Savant, y, a partir de entonces, dejó su marca en la comunidad matemática[116]. Aunque existen diferentes versiones, la forma canónica o clásica de la paradoja es la siguiente:

> La versión canónica
> Se te muestran tres puertas idénticas. Detrás de una de ellas hay un automóvil. Las otras dos ocultan cabras —una en cada puerta—. Se te pide escoger, sin abrirla, una de las puertas. Después de hacerlo, Monty, quien sabe dónde está el automóvil, abre una de las dos puertas restantes. Él siempre abre una puerta que sabe que es errónea —i.e., que esconde una cabra—, y elige azarosamente cuál puerta abrirá cuando tiene más de una opción (lo cual sucede en aquellas ocasiones en las que tu opción inicial oculta el auto). Después de abrir una puerta incorrecta, Monty te da la opción de cambiar a la otra puerta que permanece cerrada o quedarte con tu elección original. Entonces recibirás lo que sea que esté detrás de la puerta que has escogido. ¿Qué deberías hacer?[117].

Un gran número de personas, incluido quien esto escribe, responde que, sin importar nuestra decisión, la probabilidad de que el automóvil se encuentre detrás de la puerta que hemos elegido o detrás de la puerta restante es exactamente la misma, 1/2, por lo que da igual si cambiamos de puerta o nos aferramos a nuestra elección original, nuestra probabilidad de ganar —o perder— para cada una de ellas es la misma, 50/50. No obstante, en este punto la intuición de la mayoría está errada (lo que, dicho sea de paso, le confiere su carácter paradójico al problema de Monty Hall): en realidad, al cambiar nuestra opción inicial duplicamos nuestras oportunidades de ganar el automóvil, pues la probabilidad de que el automóvil se encuentre detrás de la puerta que elegimos en un principio es de 1/3, mientras que la probabilidad de que esté tras

---

[115] Steve Selvin, "A Problem in Probability" (letter to the editor), American Statistician, Vol. 29, No. I, 1975, p. 67. Ese mismo año, en agosto, apareció una segunda carta, en donde se continúa el tratamiento del problema y se le da el nombre de *Monty Hall problem:* Steve Selvin, "On the Monty Hall Problem" (letter to the editor), American Statistician, Vol. 29, No. 3, 1975, p. 134.

[116] Marilyn vos Savant, *The Power of Logical Thinking*, St. Martin's Press, New York, 1996.

[117] *Cf.*, Rosenhouse, Jason, *The Monty Hall Problem: The Remarkable Story of Math's Most Contentious Brainteaser*, Oxford University Press, New York, 2009, pp. 35-36.

la otra puerta restante es de 2/3. Por consiguiente, lo racional sería cambiar de puerta. El acertijo de Monty Hall se puede modelar empleando cálculo de probabilidad clásico, o bien, a través de cálculo Bayesiano; con ambos desarrollos se puede demostrar este resultado que, si bien es contraintuitivo, es cierto[118]. Yo optaré aquí por una prueba mucho más sencilla[119].

Empecemos por distinguir las dos cabras que aparecen en el acertijo, y las designaremos *cabra número uno* y *cabra número dos*. Ahora, representaremos el contenido de las diferentes puertas del modo siguiente: A, C_1 o C_2 (A para automóvil, cabra uno y cabra dos, respectivamente). Tenemos entonces seis combinaciones posibles:

Ubicaciones posibles del auto y las cabras

Puerta 1	Puerta 2	Puerta 3
A	C_1	C_2
A	C_2	C_1
C_1	A	C_2
C_2	A	C_1
C_1	C_2	A
C_2	C_1	A

[118] Para la resolución de este problema empleando probabilidad clásica, así como probabilidad Bayesiana, ver Rosenhouse, J, *The Monty Hall Problem…*, caps. 2 y 3.

[119] La solución que presento a continuación fue formulada por Steven Krantz en *Techniques of Problem Solving*, American Mathematical Society, Providence, 1997. Sin embargo, yo tuve conocimiento de ella a través del capítulo 2 del libro de J. Rosenhouse al que aludo en la nota anterior (*cf.*, específicamente, p. 54), y cuya exposición de la solución de Krantz sigo aquí casi al pie de la letra.

Todas estas combinaciones son igualmente probables, de modo que la probabilidad para cada una de ellas es de 1/6. Supongamos, sin perder generalidad, que escogemos inicialmente la puerta número uno. Si nos encontramos en la primera línea de nuestra tabla, entonces Monty abrirá la puerta dos o la puerta tres; en cualquier caso, perderemos el auto si cambiamos de puerta. Lo mismo ocurre con la segunda línea de la tabla. Si nos hallamos en la tercera línea o combinación, entonces Monty descubrirá necesariamente la cabra que está detrás de la puerta tres; en este caso, ganaremos el auto si cambiamos de puerta. La misma situación ocurre en la cuarta línea de la tabla. Algo completamente análogo sucede en las líneas quinta y sexta, con la diferencia de que en ellas Monty se ve forzado a abrir la puerta dos en lugar de la tres. Por lo tanto, en cuatro de seis posibilidades ganaremos al cambiar de puerta, lo que significa que la probabilidad de ganar al hacerlo es de 2/3.

Sin duda, es éste un acertijo de lo más interesante, pero ¿en qué sentido nos permite objetar la teoría epistemológica HND? Como ya mencioné, un gran número de personas yerran al responder que la probabilidad de ganar con cada una de las puertas restantes —la que han elegido inicialmente y la otra puerta que permanece cerrada— es de 1/2. A este respecto, haremos dos observaciones: I) Muchos de los individuos que equivocan la respuesta tiene instrucción en ciencias empíricas o formales (específicamente, en Matemáticas); II) dentro de este último grupo, existen personas que se rehúsan a aceptar que la probabilidad de ganar se duplica con el cambio de puerta, incluso después de haber recibido alguna explicación o prueba[120]. Partamos de aquí en la formulación de nuestra objeción contra HND. Consideremos la formulación clásica del problema de Monty Hall, la cual presentamos hace un momento, y designemos con C^{mh} a la conjunción de todas las proposiciones que la articulan (*i. e.*, C^{mh}=se te muestran tres puertas idénticas y detrás de una de ellas hay un automóvil. Además, las otras dos ocultan cabras —una en cada puerta—…). Para cualquier agente estándar, las proposiciones que articulan esta conjunción son comprendidas y

[120] En relación con estas dos observaciones, *cf.*, las secciones 1.10 [*L'Affaire* Parade] y 1.11 [The *American Statistician* Exchange, así como la sección 2.8 [Final Comments], en los capítulos 1 y 2, respectivamente, de *The Monty Hall Problem…* (J. Rosenhouse).

conocidas como ciertas tan pronto como recibe un planteamiento adecuado del problema[121]. Por consiguiente, cabe decir que un gran número de agentes contemplados en II) —cuando no todos— conocen cada uno de los conyuntos de C^{mh} y, en consecuencia, saben que es verdad C^{mh}[122]. Más aún, es probable que algunos de ellos, dada su formación, contaran con conocimientos de cálculo de probabilidades suficientes como para determinar de manera correcta las probabilidades al interior del acertijo. Con todo, ninguno de ellos sabe que la probabilidad de ganar no es la misma para ambas puertas, sino que ésta se duplica al cambiar de puerta. En otras palabras, existen agentes que conocen la conjunción C^{mh} (así como herramientas probabilísticas que les permitirían calcular probabilidades de manera adecuada), pero que ignoran una consecuencia lógica de su conocimiento o, si se quiere, de la conjunción por ellos conocida, *viz.*, la falsedad de la proposición 'la probabilidad de ganar para cada una de las puertas que permanece cerrada es de 1/2'[123]. Pero esto es un contraejemplo a la teoría epistemológica HND.

Para ver lo anterior con mayor claridad, consideremos el diagnóstico que de la situación planteada haría la teoría HND.

Sea *j* un agente cualquiera y sea *p* esta proposición:

P: la probabilidad de ganar para cada puerta que permanece cerrada es de 1/2.

Tenemos entonces la siguiente instancia de la teoría HND:

---

[121] Aquí la situación es análoga a la del caso de Alice y Bob: los agentes saben cada una de las proposiciones contenidas en C^{mh} por "construcción del problema". Así, un agente sabe que es verdad la proposición 'el automóvil está detrás de una de las tres puertas' porque ello es una presuposición necesaria para la construcción y resolución del acertijo.

[122] Estoy empleando aquí el principio epistémico: un agente sabe una conjunción sii sabe cada uno de los conyuntos $(K_j(\alpha\wedge\beta)\leftrightarrow(K_j\alpha\wedge K_j\beta)$ [es éste, por cierto, el teorema 2 del sistema $\mathbb{K}$]).

[123] Decimos que estos agentes ignoran la falsedad de la proposición en cuestión, no simplemente debido a que no la creen, sino porque carecen de una demostración de su falsedad, o bien, porque no comprenden o no aceptan aquellas pruebas que se les han ofrecido.

Si el conocimiento de j acerca de C^{mh} (K_jC^{mh}) implica la falsedad de p ($\neg p$), entonces, a menos j sepa que P es falsa ($K_j\neg p$), no sabrá C^{mh} ($\neg K_jC^{mh}$).

Nótese que el antecedente de esta proposición se cumple cabalmente, pues del conocimiento de *j* [K_jC^{mh}] se sigue lógicamente que la proposición p es falsa [$\neg p$][124].

Por lo tanto –la teoría HND nos llevaría a concluir–, el agente *j* no sabe la conjunción C^{mh} [$\neg K_jC^{mh}$] a menos que sepa que la proposición p es falsa [$\neg p$]. La cuestión aquí es si semejante conclusión es o no cierta. Veamos, la conjunción C^{mh}, como señalé en una nota anterior, se encuentra articulada únicamente por enunciados que forman parte de la construcción del problema y, en consecuencia, son puntos de partida obligatorios para la resolución del mismo. En cuanto a las bases probabilísticas del problema, es bien sabido que más de uno de los agentes que han errado en la solución, y permanecido en el error por un tiempo nada despreciable, eran matemáticos y, en consecuencia, habían recibido alguna instrucción en el área de la probabilidad[125]. Tomando en consideración esto, ¿cuál es un mejor diagnóstico de la situación de estos agentes?, ¿su falta de conocimiento acerca de las verdaderas probabilidades en el problema de Monty Hall es una prueba de su ignorancia respecto a la conjunción C^{mh} —así como de la probabilidad asociada al problema—, o bien, conocen, en efecto, la conjunción y las bases probabilísticas en cuestión, pero fallaron en percatarse de una de las consecuencias lógicas de su conocimiento? Al menos para mí, esta última descripción es más adecuada y, de estar en lo correcto, éste sería un contraejemplo para HND.

He escogido este contraejemplo en el ámbito de las matemáticas por dos razones principalmente. Primero, me permite mostrar que el concepto de saber capturado por

[124] Empleo aquí, de manera implícita, la idea de que el saber implica verdad (principio T): si *j* sabe que C^{mh} es verdadera, entonces C^{mh} de hecho lo es [$K_jC^{mh} \rightarrow C^{mh}$]; si C^{mh} es verdadera, entonces p es falsa [$C^{mh} \rightarrow \neg p$]. Por lo tanto, si *j* sabe que C^{mh} es verdadera, entonces p es falsa [$K_jC^{mh} \rightarrow \neg p$].

[125] *Cf.*, Marilyn vos Savant, *The Power of Logical Thinking*, St. Martin's Press, New York, 1996.

la teoría HND es demasiado fuerte incluso para los estándares de esta ciencia formal[126]. Tal observación cobra relevancia cuando se la pone en contexto. Por ejemplo, respecto a esta noción de *saber*, Jaakko Hintikka ha escrito: << [...] la semántica de la tesis KK muestra que sirve para capturar un sentido fuerte de saber. Sospecho que éste puede ser el sentido en el que los filósofos han tratado de emplear el concepto de conocimiento. La pregunta verdaderamente interesante aquí es si esta sospecha es correcta>>[127]. Pues bien, sea o no producto de la Filosofía, nuestro ejemplo exhibe que tal concepto de "saber en sentido fuerte" es demasiado para prácticamente cualquier conocimiento humano, lo que lo convierte en un concepto vacío o inaplicable. Segundo, la tesis KK, como otros elementos de Lógica Epistémica (la regla de inferencia K, por ejemplo), conlleva una concepción ideal de los agentes, pues exige de éstos, a través de la regla que lleva el mismo nombre, que conozcan cualquier proposición, digamos β, implicada por su saber $[K_j\alpha \rightarrow K_j\beta]$, al margen de si conocen o no esta relación de implicación $[K_j\alpha \rightarrow \beta]$. En contraste, la paradoja de Monty Hall permite alumbrar las condiciones reales a las que se hallan sujetos los agentes epistémicos, al menos los seres humanos en el ámbito del razonamiento probabilístico. A este respecto, Charles S. Peirce observó alguna vez que, en teoría de la probabilidad, como en ningún otro campo en Matemáticas, es fácil tropezar incluso si se es un experto en la materia; y la historia del pensamiento parece darle la razón: <<Leibniz pensó que, al lanzar un par de dados, era igualmente fácil tirar un 12 que un 11. Jean Ie Rond d' Alembert, el gran matemático frances del siglo XVIII, no fue capaz de ver que los resultados de lanzar una moneda tres veces eran los mismos que al lanzar tres monedas una vez, y creyó (como muchos jugadores amateurs insisten en creer) que tras un gran número de caras, una cruz es

[126] Nótese que los usos de la noción de *saber*, en el contexto de la paradoja de Monty Hall, se ajustan a los estándares de la ciencia matemática: por ejemplo, decimos que los agentes *conocen* la conjunción C^{mh} por definición o construcción del ejemplo; que *conocen* determinadas herramientas probabilísticas por definición, como axiomas o como resultado de alguna demostración; mientras que declaramos su ignorancia (su ausencia de saber) respecto a la probabilidad correcta porque no la creen ni tienen, *stricto sensu*, una demostración para ella.

[127] Radu J. Bogdan (ed.), '*Jaakko Hintikka*', D. Reidel Publishing Company, 1987, 310.

más probable>>[128]. Tarde o temprano, la Lógica Epistémica tendrá que tomar en cuenta estas condiciones si aspira a alguna vez modelar realmente el conocimiento humano.

4.5. El sistema S5

Como remate para este capítulo, examinaré con brevedad un último sistema axiomático estándar, *viz.*, el sistema S5. De todos los sistemas que hemos analizado hasta ahora, éste es el de mayor poder demostrativo, pues entre sus teoremas están contenidos todos los de los tres sistemas anteriores (K, T y S4), más algunos teoremas adicionales. Al igual que S4, S5 es una extensión de la teoría T, pues se obtiene de ella al agregar el esquema A5: $\neg K_j\alpha \rightarrow K_j\neg K_j\alpha$ [propiedad de introspección negativa], que es el axioma o principio epistémico que lo caracteriza[129]. De modo que los axiomas y reglas de inferencia de nuestro sistema son:

Axiomas de S5

Sean α, β y γ *fbf ´s* cualesquiera de $\mathcal{L}$-{K_1, ..., K_n} y sea j=1, ..., n.

A1: $\alpha\rightarrow(\beta\rightarrow\alpha)$,

A2: $(\alpha\rightarrow(\beta\rightarrow\gamma))\rightarrow((\alpha\rightarrow\beta)\rightarrow(\alpha\rightarrow\gamma))$,

A3: $(\neg\alpha\rightarrow\neg\beta)\rightarrow((\neg\alpha\rightarrow\beta)\rightarrow\alpha)$,

K: $K_j(\alpha\rightarrow\beta) \rightarrow(K_j\alpha\rightarrow K_j\beta)$ [Propiedad de distribución],

T: $K_j\alpha\rightarrow\alpha$ [Propiedad de conocimiento]

A5: $\neg K_j\alpha\rightarrow K_j\neg K_j\alpha$ [Propiedad de introspección negativa].

Reglas de inferencia de S5

Sean α y β *fbf ´s* cualesquiera de $\mathcal{L}$-{K_1, ..., K_n}.

MP: Si $\vdash^{S4} \alpha$ Y $\vdash^{S4} \alpha\rightarrow\beta$, entonces $\vdash^{S4} \beta$ [*Modus ponens*] y

---

[128] Martin Gardner, "Problems Involving Questions of Probability and Ambiguity'' Scientific American, Vol. 201, No. 4, April 1959, pp. 174-82.

[129] Recuérdese que, de manera intuitiva, este principio se lee: *si agente ignora una proposición, entonces sabe que la ignora.*

KG: Si $\vdash^{S4} \alpha$, entonces $\vdash^{S4} K_j\alpha$ [Generalización del conocimiento].

Como acabo de mencionar, a partir de este nuevo axioma se pueden derivar teoremas nuevos, que no se obtenían en los sistemas precedentes, por ejemplo:

Teoremas de S5

T20 $\vdash^{S5} \neg K_j\neg K_j\alpha \rightarrow K_j\alpha$

T21 $\vdash^{S5} K_j\alpha \leftrightarrow \neg K_j\neg K_j\alpha$

T22 $\vdash^{S5} \neg K_j\alpha \leftrightarrow K_j\neg K_j\alpha$

T23 $\vdash^{S5} \alpha \rightarrow K_j\neg K_j\neg\alpha$ **[Propiedad brouweriana]** [130]

Todos estos teoremas no sólo evidencian el poder demostrativo de esta teoría, sino justo que los alcances demostrativos de S5 van más allá de la frontera de lo que nos parece aceptable, y, en este sentido, podríamos decir que en ella se demuestra más de lo que quisiéramos. Al analizar el sistema S4, señalamos ya que el principio de introspección negativa (A5) es controversial, pues resulta contrario al sentido común y a la tradición filosófica al afirmar que los agentes conocen siempre su propia ignorancia; razón por la que el teorema 22 —cuya lectura intuitiva sería que ignoramos una proposición siempre y cuando sepamos que la ignoramos—, enfrenta la misma dificultad[131]. El teorema 20 también es problemático; nos dice que, si un agente no sabe que ignora una proposición, entonces sabe dicha proposición. Así, este teorema nos llevaría a concluir que, en el contexto de nuestra discusión anterior acerca de la paradoja de Monty Hall, todos aquellos quienes desconocen su propia ignorancia respecto al aumento de probabilidad al cambiar de puerta *saben*, de hecho, esto último —de modo que no nos explicamos el porqué de su empeño en sostener justo lo contrario, *i. e.*, que no existe ventaja en el cambio de puerta—; lo que, desde luego, es completamente absurdo. Podríamos imputar, por la misma razón, el teorema 21[132]. Por

---

[130] Para la demostración de estos teoremas, consúltese el apéndice I.

[131] El teorema 22 se obtiene directamente de A5 y una instancia del axioma T.

[132] El teorema 21 se demuestra a partir del teorema 20 y una instancia del teorema 16. Ver la lista teoremas del sistema T al comienzo del capítulo III.

su parte, con el teorema 23, conocido como *propiedad brouweriana*, se afirma que, si una proposición es verdadera, entonces sabemos que no conocemos (que ignoramos) su negación[133]. Esto, en principio, es falso; por ejemplo, un terraplanista está lejos de saber que ignora la siguiente falsedad: *la Tierra no es semiesférica* (porque, según él, es plana).

Entre los teoremas de S5, se cuenta la tesis KK, cuya demostración sería la siguiente:

T24$\vdash^{S5} K_j\alpha \rightarrow K_jK_j\alpha$

1. $\neg K_j\neg K_j\alpha \rightarrow K_j\alpha$	*Teorema 20*
2. $K_j\neg K_j\neg K_j\alpha \rightarrow K_jK_j\alpha$	*1, RD1*
3. $K_j\alpha \rightarrow K_j\neg K_j\neg K_j\alpha$	*Teorema 23 (α/ $K_j\alpha$)*
4. $K_j\alpha \rightarrow K_jK_j\alpha$	*2, 3, SH*

Podemos preguntarnos ahora sobre los alcances demostrativos de esta teoría, y averiguar, por ejemplo, si en ella se demuestra cualquier instancia del esquema '$\alpha \rightarrow K_j\alpha$'. De ser así, dado que se tiene también el axioma T, en S5 se borraría la distinción entre conocimiento y verdad, al poder demostrar el bicondicional '$\alpha \leftrightarrow K_j\alpha$'. Por grande que sea el poder demostrativo de una teoría, es éste un resultado que no se desea obtener, y ello por dos razones: primero, este bicondicional no dice que el sistema reconoce como equivalentes la verdad y el conocimiento, lo que no sólo es contrario al sentido común, sino que es falso; segundo, la Lógica Epistémica estaría de más, pues toda fórmula con operadores epistémicos sería equivalente a una fórmula sin dichos operadores, *i. e.*, a una fórmula de la lógica proposicional estándar. Así, para determinar si el esquema '$\alpha \rightarrow K_j\alpha$' es demostrable en la teoría o no, definiremos una semántica "adecuada" para este sistema, con la que obtendremos el metateorema de corrección. Con este último, obtendremos la respuesta que buscamos.

[133] Este teorema, o principio B, sirve para caracterizar al sistema axiomático B, el cual guarda relación con la lógica de corte intuicionista. Quizá en ésta, en donde la verdad vale a demostrabilidad, el teorema cobre plausibilidad.

4.6. Semántica y metateorema de corrección para S5

Definiremos ahora una semántica que nos permita recatar la validez del axioma de introspección negativa. Al igual que para los sistemas anteriores, esta semántica consta de una noción de interpretación, así como de una noción de verdad.

Definición de S5-interpretación para $\mathcal{L}$-$\{K_1, \ldots, K_n\}$

Una S5-interpretación ($\mathcal{E}_{S4}$) para $\mathcal{L}$-$\{K_1, \ldots, K_n\}$ es un tuplo:

$\mathcal{E} = \langle S, \pi, R_1, \ldots, R_n \rangle$.

Donde:

1) $S \neq \emptyset$ S es un conjunto no vacío (llamado el conjunto de estados s de $\mathcal{E}$).

2) $\pi: L \longrightarrow P(S)$ π es una función que va de L al conjunto potencia de S, donde L es el conjunto de letras proposicionales de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$ ($\pi(p_i) \subseteq S$).

3) Para cada j=1, ..., n, $R_j \subseteq S^2$ es una relación reflexiva [*i. e.*, para cualquier $s \in S$ se tiene que $\langle s, s \rangle \in R_j$] y euclidiana [*i. e.*, para cualesquiera s, t, $u \in S$ se tiene que si $\langle s, t \rangle \in R_j$ y $\langle s, u \rangle \in R_j$, entonces $\langle t, u \rangle \in R_j$] ($R_j$ se llama la relación de accesibilidad o posibilidad de j).

Definición de S5-verdad para $\mathcal{L}$-$\{K_1, \ldots, K_n\}$

Nuestra noción de satisfacibilidad para S5 es esencialmente la misma que para T y S4[134].

Definición de S5-verdad:

σ es verdadera en una S5–interpretación $\mathcal{E}$ sii para todo estado s de $\mathcal{E}$ se tiene que $\mathcal{E}$, $s \vDash \sigma$.

Notación: $\mathcal{E} \vDash \sigma$

[134] Para las cláusulas de satisfacibilidad, ver el apartado "Semántica formal para el sistema T", en el capítulo III.

Definamos ahora lo que significa que σ sea una fórmula S5-válida de nuestro lenguaje.

Definición de S5-validez:

σ es S5-válida sii para toda S5–interpretación $\mathcal{E}$ se tiene que $\mathcal{E} \vDash \sigma$.

Notación: $\vDash^{S5} \sigma$

En esta nueva semántica, se mantiene la restricción de reflexividad en la interpretación para salvaguardar la validez del principio T, pero se exige ahora que las relaciones R_j sean también relaciones euclidianas, *i. e.*, que, *si un mundo o estado determinado ve a otros dos mundos, entonces estos dos mundos deben verse entre sí*. Esta restricción equivale a pedir que la información accesible a un agente *j* desde un mundo o estado s sea un subconjunto de la información disponible para él desde cualquier alternativa epistémica s' (*i. e.*, los mundos s' accesibles para *j* desde s). Con ello se garantiza que la ignorancia de *j* se preserve a través de todas estas alternativas epistémicas: en caso de que *j* ignore una proposición α en s ($\mathcal{E}$, s$\nvDash\alpha$), también la ignorará en todo s' visible para él desde s ($\mathcal{E}$, s'$\nvDash\alpha$, con s' cualquier s$\in$S tal que $\langle s, s' \rangle \in R_j$); un momento de reflexión le mostrará al lector que esto es justo lo que se requiere para asegurar la validez del axioma A5: $\neg K_j\alpha \rightarrow K_j\neg K_j\alpha$.

Que en esta semántica se garantiza la validez de A5 (*i. e.*, A5 es S5-válido), lo probaremos del modo siguiente:

$\vDash^{S5} \neg K_j\alpha \rightarrow K_j\neg K_j\alpha$.

Demostración por reducción al absurdo

Supongamos que $\nvDash^{S5} \neg K_j\alpha \rightarrow K_j\neg K_j\alpha$. Es decir, que existe una S5-interpretación $\mathcal{E}$ y un estado s de $\mathcal{E}$ tales que:

1) $\mathcal{E}$, s$\vDash\neg K_j\alpha$ y 2) $\mathcal{E}$, s$\nvDash K_j\neg K_j\alpha$.

De 2) y cláusula (K_j) de nuestra definición de satisfacibilidad, se sigue que:

3) $\mathcal{E}$, t$\nvDash\neg K_j\alpha$, con t algún estado de $\mathcal{E}$ tal que $\langle s, t \rangle \in R_j$.

Ahora, de 1), 3) y cláusula ($\neg$), se tiene, respectivamente, que:

4) $\mathcal{E}$, s$\not\models K_j\alpha$ y 5) $\mathcal{E}$, t$\models K_j\alpha$.

De 4) y cláusula (K_j), tenemos que:

6) $\mathcal{E}$, u$\not\models\alpha$, con u algún estado de $\mathcal{E}$ tal que $\langle s, u\rangle \in R_j$.

Como R_j es euclidiana, de $\langle s, t\rangle \in R_j$ y $\langle s, u\rangle \in R_j$, se tiene que $\langle t, u\rangle \in R_j$.
De esto y la misma cláusula que empleamos con antelación, se sigue que:

5) $\mathcal{E}$, t$\not\models K_j\alpha$**!** Pero esto se contradice con 5).

$\therefore \models^{S5} \neg K_j\alpha \rightarrow K_j\neg K_j\alpha$.

Como era de esperarse, esta semántica permite probar el metateorema de corrección para S5, y su demostración, una vez más, la dejaré en manos del lector, ya que es completamente análoga a la de los sistemas anteriores. Enunciemos simplemente, entonces, el metateorema de corrección.

Metateorema de corrección para S5

Sea φ una *fbf* cualquiera $\mathcal{L}$-$\{K_1, \ldots, K_n\}$.

Todo teorema de S5 es S5-válido. Es decir, si $\vdash^{S5} \varphi$, entonces $\models^{S5} \varphi$.

4.7. El poder demostrativo del sistema S5

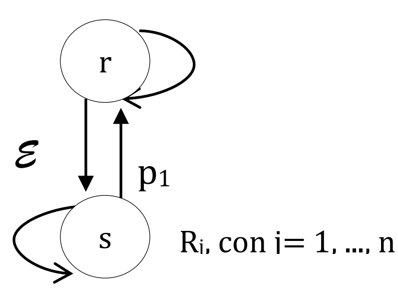
Podemos ahora, con el metateorema de corrección, responder a la pregunta de si en S5 se demuestra cualquier instancia del esquema de fórmulas '$\alpha \rightarrow K_j\alpha$'. El metateorema de corrección nos deja saber que todo teorema de S5 tiene la propiedad de ser S5-válido; por ello, si logramos mostrar que alguna instancia del esquema anterior no satisface dicha propiedad, habremos respondido nuestra pregunta, pues habremos probado que ella no es demostrable en nuestro sistema.

Consideremos la fórmula '$p_1 \rightarrow K_2p_1$', así como la siguiente S5-interpretación:

$\mathcal{E} = \langle S, \pi, R_1, \ldots, R_n\rangle$; S={r, s}

π: L $\longrightarrow$ P(S)

$\pi(\psi) = \emptyset$, para toda letra proposicional ψ,



excepto la letra proposicional p_1.

$\pi(p_1)=\{s\}$

$R_j= \{<r, r>, <s, s>, <r, s>, <s, r>\}$, con j= 1, .., n

Bajo esta S5-interpretación se tiene que:

1) $\mathcal{E}$, $r\nvDash p_1$, 2) $\mathcal{E}$, $s\vDash p_1$. A partir de la relación de accesibilidad R_2 y la cláusula modal de nuestra definición de satisfacibilidad, se sigue, por (1), que 3) $\mathcal{E}$, $s\nvDash K_2p_1$. Finalmente, de (2), (3) y la cláusula de la implicación, tenemos que 4) $\mathcal{E}$, $s\nvDash p_1\rightarrow K_2p_1$. Por consiguiente, esta fórmula no es S5-válida debido a que no es verdadera en esta S5-interpretación ($\boldsymbol{\mathcal{E}\nvDash p_1\rightarrow K_2p_1}$).

De lo anterior se desprende que, si bien el poder demostrativo de la teoría S5 va más allá de lo deseable, no llega al límite de demostrar cualquier instancia de '$\alpha\rightarrow K_j\alpha$'; de modo que tampoco borra la distinción básica entre el conocimiento proposicional y la verdad.

4.8. Consideraciones en torno al sistema S5

A diferencia de como procedí con los sistemas K, T y S4, respecto a S5 me limitaré a mencionar algunas consideraciones generales sobre su estatus epistemológico.

Como ya mencioné en repetidas ocasiones a lo largo de este capítulo, el axioma característico de este sistema ($\neg K_j\alpha\rightarrow K_j\neg K_j\alpha$ [axioma de introspección negativa]), o la idea de que los agentes siempre saben cuándo ignoran una proposición, es inaceptable o simplemente falso. Por otra parte, las dos objeciones que presentamos para la tesis KK, *i. e.*, la crítica de Lenzen, así como la objeción del regreso epistémico, son aplicables, casi sin modificación alguna, para este principio[135]. Por último, hay que señalar que, en el último capítulo de este trabajo, dedicado al análisis de la Lógica Epistémica

[135] Ver la primera parte de la sección "Consideraciones epistémicas en torno al sistema S5, en este mismo capítulo.

Multimodal, se plantea una última crítica a este principio a partir de un problema conocido como la paradoja del creyente perfecto[136].

[136] *Cf.*, la parte final del apartado "El sistema KB^{PG}, en el capítulo VI.

5. Lógica Epistémica como lógica de primer orden

5.1. Lógica Epistémica como lógica de primer orden

Las *fbf's* de un lenguaje epistémico proposicional de *n*-agentes pueden ser traducidas a *fbf's* de un lenguaje de primer orden en una forma diáfana[137].

Definimos $\mathcal{L}$-{K_1, ..., K_n} (el lenguaje epistémico proposicional cuyo conjunto de operadores epistémicos es {K_1, ..., K_n}, para algún número natural n).

Símbolos de $\mathcal{L}$-{K_1, ..., K_n}:

I) Un conjunto infinito numerable de letras proposicionales $\{p_i: i \in \mathbb{N}\}$

$p_1, p_2, \ldots, p_m, \ldots$

II) $\neg, \rightarrow$

III) (,)

IV) Un conjunto de n operadores epistémicos

$K_1, \ldots, K_n$

Fbf's de $\mathcal{L}$-{K_1, ..., K_n}:

I) Toda letra proposicional es una *fbf*.

II) Si α y β son *fbf 's,* también lo son:

$(\neg\alpha)$, $(\alpha\rightarrow\beta)$, $(K_j\alpha)$ con j=1, 2, ..., n.

III) No hay más *fbf 's.*

A partir de este lenguaje, definimos un lenguaje de primer orden $\mathcal{I}$ como sigue.

---

[137] Lo que se presenta en este quinto capítulo es una suerte de reducción de la Lógica Modal a la lógica de primer orden. Le agradezco al Dr. Max Fernández de Castro por haberme hecho notar que tal reducción forma parte de un trabajo más extenso, y cuyo desarrollo debemos al lógico J. van Benthem. Al respecto, *cf.,* el capítulo 6 de su *Modal Logic for Open Minds*.

$\mathcal{L}$ consta de los siguientes símbolos:

Símbolos lógicos

I) Un conjunto numerable de variables individuales

$x_1, x_2, x_3, \ldots$

II) Conectivos lógicos

$\neg, \rightarrow$

III) Cuantificador universal

$\forall$

IV) Paréntesis

(,)

Símbolos no lógicos

V) Un conjunto infinito numerable de símbolos de predicados de aridad 1 $\{P_i : i \in \mathbb{N}\}$

$P_1, P_2, \ldots, P_m, \ldots$

VI) Un conjunto de n símbolos de predicados de aridad 2

$Q_1, \ldots, Q_n$[138].

Reglas de formación de *fbf*`s de $\mathcal{L}$

Sean v y w variables individuales cualesquiera.

I) $P_i v$ es *fbf*, para $i \in \mathbb{N}$

II) $Q_j vw$ también es *fbf*, $j=1, \ldots, n$

[138] Los conjuntos de subíndices para los símbolos de predicados unarios y binarios de $\mathcal{L}$ serán exactamente los conjuntos de subíndices para las letras proposicionales y los operadores epistémicos de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$, respectivamente.

III) Si α y β son *fbf ´s*, también lo son:

$(\neg\alpha)$, $(\alpha\rightarrow\beta)$.

IV) Si α es una *fbf*, también es *fbf*:

$\forall v\alpha$.

V) No hay más *fbf ´s*.

Una vez definidos estos lenguajes, a cada fórmula φ del lenguaje epistémico proposicional $\mathcal{L}$-$\{K_1, \ldots, K_n\}$ vamos a asociarle una fórmula φ^* del lenguaje de predicados $\mathcal{L}$ con una variable libre (su traducción a este lenguaje).

Definimos recursivamente cómo traducir *fbf´s* de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$ en *fbf´s* de $\mathcal{L}$.

Sea v una variable fija.

1) Si $\varphi=p_i$, entonces $\varphi^*= P_i v$.

2) Si $\varphi=\neg\alpha$, entonces $\varphi^*=\neg\alpha^*_{(v)}$. Y si $\varphi=\alpha\rightarrow\beta$, entonces $\varphi^*= \alpha^*_{(v)}\rightarrow\beta^*_{(v)}$.

3) Si $\varphi=K_j\alpha$ (con $j=1, \ldots, n$), entonces $\varphi^*=\forall w(Q_j vw\rightarrow \alpha^*_{[v/w]})$, con w una variable nueva (que no aparezca en α^*). Donde '$\alpha^*_{[v/w]}$' es la fórmula que se obtiene al sustituir todas las ocurrencias libres de v en 'α^*' por w[139].

Veamos algunos ejemplos de traducciones:

Fbf´s de $\mathcal{L}$-{K1, …, Kn}	Traducción a $\mathcal{L}$
p_2	P_2x_1
$\neg(p_1\rightarrow p_2)\rightarrow\neg\neg p_3$	$\neg(P_1x_1\rightarrow P_2x_1)\rightarrow\neg\neg P_3x_1$
$K_1(p_1\rightarrow p_5)$	$\forall x_2(Q_1x_1x_2\rightarrow(P_1x_2\rightarrow P_5x_2))$
$K_1K_2\neg p_{10}$	$\forall x_2(Q_1x_1x_2\rightarrow\forall x_3(Q_2x_2x_3\rightarrow\neg P_{10}x_3))$

[139] Nótese que las reglas 1) - 3) de traducción llevan a fórmulas en donde únicamente aparece libre la variable v.

Esta traducción de las fórmulas de $\mathcal{L}$-{K_1, ..., K_n} en fórmulas de $\mathcal{I}$ es complementada con una correspondencia semántica entre estos lenguajes. Para ello, definimos primero una interpretación, así como una noción de verdad, para cada uno de nuestros lenguajes.

Una estructura de Kripke, *i. e.*, una interpretación, para $\mathcal{L}$-{K_1, ..., K_n} es un tuplo:

$\mathcal{E}$= <S, π, R_1, …, R_n>[140].

Donde:

1) S≠Ø S es un conjunto no vacío (llamado el conjunto de estados s de $\mathcal{E}$).

2) π: $L \longrightarrow P(S)$ π es una función que va de L al conjunto potencia de S, donde L es el conjunto de letras proposicionales de $\mathcal{L}$-{K_1, ..., K_n} ($\pi(p_i)\subseteq S$).

3) Para cada j=1, …, n, $R_j\subseteq S^2$ (R se llama la relación de accesibilidad o posibilidad de j).

Definimos recursivamente, para cada s∈S y cada *fbf* σ de $\mathcal{L}$-{K_1, ..., K_n}, lo que significa:

$\mathcal{E}$, s⊨σ (se lee: la fórmula σ es verdadera (realizada o satisfecha) en el estado s bajo la interpretación (estructura) $\mathcal{E}$).

Cláusula 1: σ es una letra proposicional

Por lo tanto, σ∈L (σ es una letra proposicional p_i)

$\mathcal{E}$, s⊨p_i sii s∈π(p_i).

Cláusula 2: $\mathcal{E}$, s⊨¬σ sii $\mathcal{E}$, s⊭σ.

Cláusula 3: $\mathcal{E}$, s⊨α→β sii $\mathcal{E}$, s⊭α o $\mathcal{E}$, s⊨β.

Cláusula 4: Para j=1, ..., n, $\mathcal{E}$, s⊨K_jσ sii $\mathcal{E}$, s'⊨σ para todo s' tal que <s, s'>∈R_j.

[140] El número de las R's en la estructura corresponde exactamente al número de operadores en el lenguaje (en este caso, las K's). Es decir, el conjunto de índices para las relaciones de accesibilidad y para los operadores de $\mathcal{L}$-{K_1, ..., K_n} será el mismo.

Definición de verdad epistémica:

σ es verdadera en una estructura $\mathcal{E}$ sii para todo estado s de $\mathcal{E}$ se tiene que $\mathcal{E}$, s$\vDash\sigma$.

Notación: $\mathcal{E}\vDash\sigma$

Definición de validez epistémica:

σ es una fórmula válida sii para toda estructura $\mathcal{E}$ se tiene que $\mathcal{E}\vDash\sigma$.

Notación: $\vDash\sigma$

Una estructura o interpretación I para $\mathcal{L}$ consta de lo siguiente:

1) A≠Ø A es el dominio (base o universo) de la estructura.
2) Para cada predicado P_i de $\mathcal{L}$ de aridad 1, una relación unaria en A ($P_i^I \subseteq A$).
3) Para cada predicado Q_j de $\mathcal{L}$ de aridad 2, una relación binaria en A ($Q_i^I \subseteq A^2$).

Considérese una sucesión ā de elementos de A:

$\bar{a}=<a_1, a_2, \ldots, a_n, \ldots>$

A esta sucesión la podemos pensar como una asignación de elementos de A a las variables del lenguaje.

Observación: a_i es el elemento del dominio de la estructura denotado (temporalmente) por la variable x_i.

Definimos recursivamente, para cada *fbf* σ de $\mathcal{L}$, lo que significa:

$I\vDash\sigma[\bar{a}]$ (se lee: la fórmula σ es verdadera (realizada o satisfecha) en la sucesión ā bajo la interpretación I).

Cláusula 1: σ es una fórmula atómica del tipo P_ix_n.

$I\vDash P_ix_n[\bar{a}]$ sii $a_n \in P_i^I$.

Cláusula 2: σ es una fórmula atómica del tipo Qx_mx_n.

$I\vDash Q_ix_mx_n[\bar{a}]$ sii $<a_m,a_n> \in Q_j^I$.

Cláusula 3: $\sigma= \neg\alpha$.

$I\vDash\neg\alpha[\bar{a}]$ sii $I\nvDash\alpha[\bar{a}]$.

Cláusula 4: $\sigma= \alpha\rightarrow\beta$

$I \vDash \alpha \rightarrow \beta[\bar{a}]$ sii $I \nvDash \alpha[\bar{a}]$ o $I \vDash \beta[\bar{a}]$

Cláusula 5: $\sigma = \forall x_n \alpha$.

$I \vDash \forall x_n \alpha[\bar{a}]$ sii $I \vDash \alpha[\bar{a}(n/b)]$, para toda $b \in A$ (donde $\bar{a}(n/b)$ es la sucesión que coincide con $\bar{a}$ en todas las entradas, excepto que en la n-ésima entrada aparece b en lugar de a_n).

Observación 2:

A veces, en lugar de escribir $I \vDash \sigma[\bar{a}]$, se escriben únicamente los elementos del dominio que interpretan a las variables libres de la fórmula. Por ejemplo, $I \vDash P_i x_n[a]$ se lee 'la fórmula $P_i x_n$ es verdadera bajo la interpretación I cuando la variable "x_n" se interpreta como "a"', mientras que $I \vDash Q_j x_m x_n[a,b,]$ se lee 'la fórmula $I \vDash Q_j x_m x_n$ es verdadera bajo la interpretación I cuando las variables "x" y "y" se interpretan como "a" y "b", respectivamente'.

Definición de verdad:

σ es verdadera en una estructura I sii para toda sucesión $\bar{a}$ se tiene que $I \vDash \sigma[\bar{a}]$.

Notación: $I \vDash \sigma$

Definición de validez universal:

σ es una fórmula universalmente válida sii para toda estructura I se tiene que $I \vDash \sigma$.

Notación: $\vDash \sigma$

Ahora bien, dada una estructura $\mathcal{E} = \langle S, \pi, R_1, \ldots, R_n \rangle$ para $\mathcal{L}$-$\{K_1, \ldots, K_n\}$, definimos $\mathcal{E}^*$, una interpretación para $\mathcal{L}$ como sigue:

1) El dominio de $\mathcal{E}^*$ es el conjunto no vacío S de estados de $\mathcal{E}$.

2) A cada predicado P_i de $\mathcal{L}$, de aridad 1, se le asigna una relación unaria en S ($P_i^{E^*} \subseteq S$), tal que para todo $s \in S$ se tiene que $s \in P_i^{E^*}$ sii $s \in \pi(p_i)$.

3) A cada predicado Q_j (con j= 1, ..., n) de L, de aridad 2, se le asigna una relación binaria en S ($Q_j^{E^*} \subseteq S^2$), tal que para cualesquiera $s,s' \in S$ se tiene que $\langle s,s' \rangle \in Q_j^{E^*}$ sii $\langle s,s' \rangle \in R_j$.

Sean $\mathcal{E}$ y φ una interpretación y una fórmula arbitrarias de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$, respectivamente. Y sea $s \in S$ un estado cualquiera.

Entonces:

Afirmación ⋆*:* $\mathcal{E}, s \vDash \varphi$ sii $\mathcal{E}^* \vDash \varphi^*_{(x)}[s]$.

Demostremos esta afirmación empleando para ello inducción matemática.

Demostración:

Por inducción matemática sobre el número m de conectivos y cuantificadores en φ.

Base: $m=0$.

Entonces φ es una letra proposicional p_i. Entonces φ^* es $P_i x$.

$\mathcal{E}, s \vDash p_i$ sii $s \in \pi(p_i)$. Por definición de $\mathcal{E}^*$, tenemos que $s \in \pi(p_i)$ sii $s \in P_i^{E^*}$; pero $s \in P_i^{E^*}$ sii $\mathcal{E}^* \vDash P_i x[s]$.

Por lo tanto, la afirmación vale para φ atómica.

Sea φ cualquier fórmula con m cuantificadores y conectivos, para $m \geq 1$ arbitrario.

HI: La afirmación vale para cualquier fórmula con menos de m conectivos y cuantificadores.

P.D. La afirmación vale para φ.

Hay tres casos para φ: $\varphi = \neg\alpha$, $\varphi = \alpha \to \beta$ y $\varphi = K_j\alpha$ (con $j = 1, \ldots, n$).

Caso 1: $\varphi = \neg\alpha$. Entonces $\varphi^* = \neg\alpha^*_{(x)}$.

$\mathcal{E}, s \vDash \neg\alpha$ sii $\mathcal{E}, s \nvDash \alpha$. Como α es de menor complejidad que φ, podemos aplicarle HI. Por HI, tenemos que $\mathcal{E}, s \nvDash \alpha$ sii $\mathcal{E}^* \nvDash \alpha^*_{(x)}[s]$. Pero por cláusula 3 de nuestra definición de verdad, tenemos que $\mathcal{E}^* \nvDash \alpha^*_{(x)}[s]$ sii $\mathcal{E}^* \vDash \neg\alpha^*_{(x)}[s]$.

Caso 2: $\varphi = \alpha \to \beta$. Entonces $\varphi^* = \alpha^*_{(x)} \to \beta^*_{(x)}$.

$\mathcal{E}, s \nvDash \alpha \to \beta$ sii $\mathcal{E}, s \vDash \alpha$ y $\mathcal{E}, s \nvDash \beta$. Por HI, tenemos que $\mathcal{E}, s \vDash \alpha$ y $\mathcal{E}, s \nvDash \beta$ sii $\mathcal{E}^* \vDash \alpha^*_{(x)}[s]$ y $\mathcal{E}^* \nvDash \beta^*_{(x)}[s]$. Pero $\mathcal{E}^* \vDash \alpha^*_{(x)}[s]$ y $\mathcal{E}^* \nvDash \beta^*_{(x)}[s]$ sii $\mathcal{E}^* \nvDash \alpha^*_{(x)} \to \beta^*_{(x)}[s]$.

Caso 3: $\varphi = K_j\alpha$ (con j=1, ..., n). Entonces $\varphi^* = \forall y(Q_jxy \rightarrow \alpha^*_{[x/y]})$, con y variable nueva[141].

Supongamos que $\mathcal{E}^* \nvDash \forall y(Q_jxy \rightarrow \alpha^*_{[x/y]})[s]$. Lo anterior es verdadero sii $\mathcal{E}^* \nvDash Q_jxy \rightarrow \alpha^*_{[x/y]}[s, s']$, para algún $s' \in S$. Pero $\mathcal{E}^* \nvDash Q_jxy \rightarrow \alpha^*_{[x/y]}[s, s']$ sii $\mathcal{E}^* \vDash Q_jxy[s, s']$ y $\mathcal{E}^* \nvDash \alpha^*_{[x/y]}[s']$. Como α es de menor complejidad que φ, podemos aplicarle HI. Entonces, por HI, tenemos que $\mathcal{E}^* \nvDash \alpha^*_{[x/y]}[s']$ sii $\mathcal{E}, s' \nvDash \alpha$. Ahora, $\mathcal{E}^* \vDash Q_jxy[s, s']$ sii $\langle s, s'\rangle \in Q_j^{E^*}$, sin embargo, por definición de $\mathcal{E}^*$, tenemos que $\langle s,s'\rangle \in Q_j^{E^*}$ sii $\langle s, s'\rangle \in R_j$. Pero $\mathcal{E}, s' \nvDash \alpha$ y $\langle s, s'\rangle \in R_j$ sii $\mathcal{E}, s \nvDash K_j\alpha$.

De la base y los casos 1-3, se sigue, por Principio de Inducción Matemática Completa (PIMC), que:

$\mathcal{E}, s \vDash \varphi$ sii $\mathcal{E}^* \vDash \varphi^*_{(x)}[s]$, para cualquier estado $s \in S$, así como cualquier $\mathcal{E}$ estructura y cualquier fórmula φ de $\mathcal{L}$-$\{K_1, ..., K_n\}$.

A partir de esta afirmación tenemos el corolario siguiente:

Corolario:

Si $\nvDash \varphi$, entonces $\nvDash \varphi^*_{(x)}$ (si φ no es válida, entonces $\varphi^*_{(x)}$ no es universalmente válida).

Veamos la prueba de este corolario:

Si $\nvDash \varphi$, entonces existen $\mathcal{E}$ y $s \in \mathcal{E}$ tales que $\mathcal{E}, s \nvDash \varphi$. Por consiguiente, de Afirmación ⋆ tenemos que $\mathcal{E}^* \nvDash \varphi^*_{(x)}[s]$. Por lo tanto, $\nvDash \varphi^*_{(x)}$.

Considérense los siguientes esquemas de *fbf's* de $\mathcal{L}$-$\{K_1, ..., K_n\}$.

K: $K_j(\alpha \rightarrow \beta) \rightarrow (K_j\alpha \rightarrow K_j\beta)$,
T: $K_j\alpha \rightarrow \alpha$,
A4: $K_j\alpha \rightarrow K_jK_j\alpha$,
A5: $\neg K_j\alpha \rightarrow K_j\neg K_j\alpha$.

Sus respectivas traducciones a $\mathcal{L}$ son:

K^*: $\forall y(Q_jxy \rightarrow (\alpha^* \rightarrow \beta^*)_{[x/y]}) \rightarrow (\forall y(Q_jxy \rightarrow \alpha^*_{[x/y]}) \rightarrow \forall y(Q_jxy \rightarrow \beta^*_{[x/y]}))$, con 'y' una variable nueva.
T^*: $\forall y(Q_jxy \rightarrow \alpha^*_{[x/y]}) \rightarrow \alpha^*_{(x)}$, con 'y' una variable nueva.
$A4^*$: $\forall y(Q_jxy \rightarrow \alpha^*_{[x/y]}) \rightarrow \forall y(Q_jxy \rightarrow \forall z(Q_jyz \rightarrow \alpha^*_{[y/z]}))$, con 'y' y 'z' variables nuevas.

[141] Para probar este último caso, emplearemos transposición.

A5*: $\neg\forall y(Q_jxy\rightarrow\alpha^*_{[x/y]})\rightarrow \forall y(Q_jxy\rightarrow\neg\forall z(Q_jyz\rightarrow\alpha^*_{[y/z]}))$, con ‘y’ y ‘z’ variables nuevas.

Demostremos que cualquier instancia de K* es una fórmula universalmente válida.

Demostración por reducción al absurdo.

Supongamos lo contrario, es decir, que existe una interpretación I de $\mathcal{L}$ y un elemento a del dominio de I tales que $I\not\vDash K^*[a]$. De esta suposición y cláusula de ‘→’, se sigue que:

1) $I\vDash\forall y(Qjxy\rightarrow(\alpha^*\rightarrow\beta^*)_{[x/y]})[a]$ y

2) $I\not\vDash(\forall y(Qjxy\rightarrow\alpha^*_{[x/y]})\rightarrow\forall y(Qjxy\rightarrow\beta^*_{[x/y]})[a]$.

De 2) y cláusula de ‘→’, tenemos que:

3) $I\vDash\forall y(Qjxy\rightarrow\alpha^*_{[x/y]})[a]$ y

4) $I\not\vDash\forall y(Qjxy\rightarrow\beta^*_{[x/y]})[a]$.

De 4) y cláusula de ‘∀’, se sigue que:

5) $I\not\vDash Qjxy\rightarrow\beta^*_{[x/y]}[a,a']$, para algún a’ elemento del dominio de I.

De 5) y cláusula de ‘→’, tenemos:

6) $I\vDash Qjxy[a,a']$ y

7) $I\not\vDash\beta^*_{[x/y]}[a']$.

De 3) y cláusula de ‘∀’, se sigue que:

8) $I\vDash Qjxy\rightarrow\alpha^*_{[x/y]}[a,a']$.

De 6), 8) y *MP*, tenemos que:

9) $I\vDash\alpha^*_{[x/y]}[a']$.

De 7), 9) y cláusula de ‘→’, se sigue que:

10) $I\not\vDash(\alpha^*\rightarrow\beta^*)_{[x/y]}[a']$.

De 6), 10) y cláusula de ‘→’, tenemos:

11) $I\not\vDash Qjxy\rightarrow(\alpha^*\rightarrow\beta^*)_{[x/y]})[a,a']$.

De 11) y cláusula de ‘∀’, se sigue que:

12) $I\not\vDash\forall y(Qjxy\rightarrow(\alpha^*\rightarrow\beta^*)_{[x/y]})[a]$. Pero esto se contradice con 2)**!**

Por lo tanto, tal I y tal a no pueden existir. Por lo tanto:

Cualquier instancia de K* es una fórmula universalmente válida.

Probemos ahora que no toda instancia de T*, A4* y A5*, es una fórmula universalmente válida.

Definimos la siguiente interpretación I para $\mathscr{L}$:

$A= \{a_1, a_2, a_3\}$

$P_1^I= \{a_1, a_2\}$

$P_i^I= \emptyset$, con $i\neq 1$

$Q_j^I= \{<a_1,a_2>, <a_2, a_3>\}$, con $j= 1, \ldots, n$.

Bajo esta interpretación consideremos las fórmulas siguientes:

1) $\forall y(Q_1xy\rightarrow P_1y)\rightarrow P_1x$ (instancia de T*)

Note que, por definición de Q_j^I, $I\nvDash Q_1xy[a_3, a]$ para todo $a\in A\therefore I\vDash Q_1xy\rightarrow P_1y[a_3, a]$ para todo $a\in A\therefore I$, $t\vDash\forall y(Q_1xy\rightarrow P_1y)[a_3]$. Por otro lado, $a_3\notin P_1^I$ $\therefore I\nvDash P_1x$. Por lo tanto, I, $t\nvDash\forall y(Q_1xy\rightarrow P_1y)\rightarrow P_1x[a_3]$. Como esta fórmula es instancia de T*, hemos probado que no toda instancia de T* es una fórmula universalmente válida.

2) $\forall y(Q_1xy\rightarrow P_1y)\rightarrow\forall y(Q_1xy\rightarrow\forall z(Q_1yz\rightarrow P_1z))$ (instancia de A4*)

Note que $I\vDash Q_1xy\rightarrow P_1y[a_1, a]$ para todo $a\in A\therefore I\vDash\forall y(Q_1xy\rightarrow P_1y)[a_1]$. Ahora bien, $<a_1, a_2>\in Q_1$, $<a_2, a_3>\in Q_1$ y $a_3\notin P_1^I\therefore I\vDash Q_1xy[a_1, a_2]$, $I\vDash Q_1yz[a_2, a_3]$ y $I\nvDash P_1z[a_3]\therefore I\nvDash Q_1yz\rightarrow P_1z[a_2, a_3]\therefore I\nvDash\forall z(Q_1yz\rightarrow P_1z)[a_2]$. Como $I\vDash Q_1xy[a_1, a_2]$ y $I,\nvDash\forall z(Q_1yz\rightarrow P_1z)[a_2]$, tenemos que $I\nvDash Q_1xy\rightarrow\forall z(Q_1yz\rightarrow P_1z)[a_1, a_2]\therefore I\nvDash\forall y(Q_1xy\rightarrow\forall z(Q_1yz\rightarrow P_1z))[a_1]$. De lo anterior y de que $I,\vDash\forall y(Q_1xy\rightarrow P_1y)[a_1]$, tenemos que $I,\nvDash\forall y(Q_1xy\rightarrow P_1y)\rightarrow\forall y(Q_1xy\rightarrow\forall z(Q_1yz\rightarrow P_1z))[a_1]$. Dado que tal fórmula es instancia de A4*, hemos probado que no toda instancias de A4* es una fórmula universalmente válida. 127

3) $\neg\forall y(Q_1xy\rightarrow Py)\rightarrow\forall y(Q_1xy\rightarrow\neg\forall z(Q_1yz\rightarrow Pz))$ (Instancia de A5*)

Como $<s_2, s_3>\in Q_1$ y $s_3\notin P_1^I$, tenemos que $I\vDash Q_1xy[a_2, a_3]$ y $I\nvDash P_1y[a_3]\therefore I\nvDash Q_1xy\rightarrow Py[a_2, a_3]$ $\therefore I\nvDash\forall y(Q_1xy\rightarrow Py)[a_2]\therefore I\vDash\neg\forall y(Q_1xy\rightarrow Py)[a_2]$. Por otro lado, $I\nvDash Q_1yz[a_3, a]$ para cualquier $a\in A\therefore I\vDash Q_1yz\rightarrow Pz[a_3, a]$, para cualquier $a\in A\therefore I\vDash\forall z(Q_1yz\rightarrow Pz)[a_3]\therefore$ $I\nvDash\neg\forall z(Q_1yz\rightarrow Pz)[a_3]$. De lo anterior y de que $I\vDash Q_1xy[a_2, a_3]$, se sigue que $I\nvDash Q_1xy\rightarrow\neg\forall z(Q_1yz\rightarrow Pz)[a_2, a_3]\therefore I\nvDash\forall y(Q_1xy\rightarrow\neg\forall z(Q_1yz\rightarrow Pz))[a_2]$. De lo anterior y de que $I\models\neg\forall y(Q_1xy\rightarrow Py)[a_2]$, se sigue que $I\not\models\neg\forall y(Q_1xy\rightarrow Py)\rightarrow\forall y(Q_1xy\rightarrow\neg\forall z(Q_1yz\rightarrow Pz))[a_2]$. Como la fórmula anterior es instancia de $A5^*$, hemos probado que no toda instancia de $A5^*$ es una fórmula universalmente válida.

No obstante lo anterior, las instancias de los esquemas T^*, $A4^*$ y $A5^*$, son verdaderas en todos los modelos de ciertos grupos de fórmulas o enunciados[142]. Tenemos las afirmaciones siguientes.

Afirmación 1: toda instancia de T^* es verdadera en cualquier modelo del conjunto de fórmulas $\{\forall xQ_jxx: j=1, \ldots, n\}$.

Afirmación 2: toda instancia de $A4^*$ es verdadera en cualquier modelo del conjunto de fórmulas $\{\forall x\forall y\forall z(Q_jxy\rightarrow(Q_jyz\rightarrow Q_jxz)): j=1, \ldots, n\}$.

Afirmación 3: toda instancia de $A5^*$ es verdadera en cualquier modelo del conjunto de fórmulas $\{\forall x\forall y\forall z(Q_jxy\rightarrow(Q_jxz\rightarrow Q_jyz)): j=1, \ldots, n\}$.

Probemos, a modo de ejemplo, la primera de estas afirmaciones.

Demostración de Afirmación 1 por reducción al absurdo.

Supongamos lo contrario, es decir, que existe un modelo, digamos I, de la fórmula $\forall xQ_jxx$, y que $I\not\models T^*$ (esto es, $I\not\models\forall y(Q_jxy\rightarrow\alpha^*_{[x/y]})\rightarrow\alpha^*_{(x)})$. De esta suposición se sigue que:

1) $I\models Q_jxx[a]$, con a cualquier elemento del dominio de I, y

2) $I\not\models\forall y(Q_jxy\rightarrow\alpha^*_{[x/y]})\rightarrow\alpha^*_{(x)})[b]$, con b algún elemento del domino de I.

De 2) y cláusula de '→', tenemos que:

3) $I\models\forall y(Q_jxy\rightarrow\alpha^*_{[x/y]})[b]$ y

4) $I\not\models\alpha^*_{(x)}[b]$.

De 3) y cláusula '∀', tenemos que $I\models Q_jxy\rightarrow\alpha^*_{[x/y]})[b,a]$, con a cualquier elemento del dominio de I. En particular tenemos que:

5) $Q_jxy\rightarrow\alpha^*_{[x/y]}[b,b]$.

[142] En ocasiones, el término ´modelo' se emplea como sinónimo de 'interpretación'; no obstante, es mucho más común emplearlo para designar una interpretación que hace verdadera a una fórmula o grupo de axiomas (teoría). Aquí se emplea la expresión 'modelo' en este último sentido.

De 1) se sigue que:

6) $I \vDash Q_j xx[b]$.

De 6), por cláusula 2 de nuestra definición de verdad, tenemos que:

7) $<b,b> \in Q_j^I$.

De 7) se sigue que:

8) $I \vDash Q_j xy[b,b]$

De 5), 8) y cláusula '→', tenemos que:

9) $\alpha^*_{[x/y]}[b]$

Pero de 4) y principio de reemplazo se sigue que[143]:

10) $I \nvDash \alpha^*_{[x/y]}[b]$. Pero esto se contradice con (9)**!**

Por tanto, el modelo I del que partimos no puede existir. Por lo tanto, toda instancia de T^* es verdadera en cualquier modelo de la fórmula $\forall x Q_j xx$.

Sea σ cualquier *fbf* de $\mathcal{L}$-{$K_1, \ldots, K_n$}.

Teorema:

$\vDash\sigma$ sii $\vDash\sigma^*_{(x)}$ (σ es válida si y sólo si $\sigma^*_{(x)}$ es universalmente válida).

El teorema anterior se puede expresar con estos dos condicionales:

I) Si $\nvDash\sigma$, entonces $\nvDash\sigma^*_{(x)}$ y

II) Si $\vDash\sigma$, entonces $\vDash\sigma^*_{(x)}$.

Ya hemos dado la prueba del condicional I) como un corolario de la afirmación $\mathcal{E},s\vDash\varphi$ sii $\mathcal{E}^*\vDash\varphi^*_{(x)}[s]$, la cual demostramos anteriormente.

[143] El principio de reemplazo se puede enunciar como sigue:

Sea φ una *fbf* de $\mathcal{L}$, I una interpretación y t una sucesión de elementos del dominio de I.

Sea t* la sucesión que coincide con t en todas las entradas, excepto que t*(y)=s(x), es decir, t* le asigna a la variable 'y' el mismo elemento que t le asigna a la variable 'x'. Entonces: I, $t\vDash\varphi$ si y sólo si I, $t^*\vDash\varphi_{[x/y]}$. Cuando en φ únicamente aparece libre la variable 'x', tenemos: $I\vDash\varphi_{(x)}[a]$ si y sólo si $I\vDash\varphi_{[x/y]}[a]$, con a cualquier elemento del domino de I.

Demostremos ahora el segundo de estos condicionales (Si $\vDash\sigma$, entonces $\vDash\sigma^*_{(x)}$). Para ello emplearemos fundamentalmente lo que sabemos hasta ahora del sistema de lógica epistémica $\mathbb{K}$: partiremos del hecho de que los axiomas de $\mathbb{K}$, al traducirlos a fórmulas de nuestro lenguaje de predicados $\mathcal{L}$, sus traducciones resultan en esquemas de *fbf's* cuyas instancias son universalmente válidas. Después, probaremos que las reglas de inferencia de $\mathbb{K}$ preservan validez universal. A partir de lo anterior, demostraremos que todo teorema de $\mathbb{K}$, al traducirlo a una fórmula de $\mathcal{L}$, su traducción resulta en un esquema de fórmula cuyas instancias son universalmente válidas. Con ello, dado que el conjunto de teoremas de $\mathbb{K}$ y el conjunto de fórmulas válidas son equivalentes, habremos alcanzado una demostración de la afirmación $\vDash\sigma$ sii $\vDash\sigma^*_{(x)}$.

Recordemos los axiomas y las reglas de inferencia para este sistema.

Axiomas de $\mathbb{K}$

Sean α, β y γ *fbf's* cualesquiera de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$.

A1. $\alpha\rightarrow(\beta\rightarrow\alpha)$,

A2. $(\alpha\rightarrow(\beta\rightarrow\gamma))\rightarrow((\alpha\rightarrow\beta)\rightarrow(\alpha\rightarrow\gamma))$,

A3. $(\neg\alpha\rightarrow\beta)\rightarrow((\neg\alpha\rightarrow\neg\beta)\rightarrow\alpha)$,

K. $K_j(\alpha\rightarrow\beta)\rightarrow(K_j\alpha\rightarrow K_j\beta)$.

Reglas de inferencia de $\mathbb{K}$

Sean α y β *fbf's* cualesquiera de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$.

MP. Si $\vdash^{\mathcal{K}}\alpha$ y $\vdash^{\mathcal{K}}\alpha\rightarrow\beta$, entonces $\vdash^{\mathcal{K}}\beta$

KG. Si $\vdash^{\mathcal{K}}\alpha$, entonces $\vdash^{\mathcal{K}}K_j\alpha$.

Las respectivas traducciones de A1-K a $\mathcal{L}$ son:

A1*. $\alpha^*_{(x)}\rightarrow(\beta^*_{(x)}\rightarrow\alpha^*_{(x)})$,

A2*. $(\alpha^*_{(x)}\rightarrow(\beta^*_{(x)}\rightarrow\gamma^*_{(x)}))\rightarrow((\alpha^*_{(x)}\rightarrow\beta^*_{(x)})\rightarrow(\alpha^*_{(x)}\rightarrow\gamma^*_{(x)}))$,

A3*. $(\neg\alpha^*_{(x)}\rightarrow\beta^*_{(x)})\rightarrow((\neg\alpha^*_{(x)}\rightarrow\neg\beta^*_{(x)})\rightarrow\alpha^*_{(x)})$,

K*. $\forall y(Q_jxy\rightarrow(\alpha^*\rightarrow\beta^*)_{[x/y]})\rightarrow(\forall y(Q_jxy\rightarrow\alpha^*_{[x/y]})\rightarrow\forall y(Q_jxy\rightarrow\beta^*_{[x/y]}))$, con 'y' una variable nueva.

Sabemos que toda instancia de A1*-K* es una fórmula universalmente válida. Probemos ahora que las reglas de inferencia de $\mathbb{K}$ (MP y KG) preservan validez universal.

Desde un punto de vista semántico, la regla de inferencia MP constituye una forma de razonamiento correcto y, por lo tanto, preserva validez universal. En cuanto a la regla de inferencia KG, considérese lo siguiente:

Sea α una *fbf* cualquiera de $\mathcal{L}$-{K_1, …,K_n}.

Supongamos que $\vDash\alpha^*_{(x)}$ ($\alpha^*_{(x)}$ es una fórmula universalmente válida). Si aplicamos la regla KG a α, obtendremos la fórmula $K_j\alpha$ (j= 1, …, n), cuya traducción a $\mathcal{L}$-{Q1, …, Qn} es: $\forall y(Q_jxy\rightarrow\alpha^*_{[x/y]})$, con 'y' una variable nueva. Como $\vDash\alpha^*_{(x)}$, tenemos que $\vDash\alpha^*_{[x/y]}$; por consiguiente, $\vDash Q_jxy\rightarrow\alpha^*_{[x/y]}$ y, por tanto, $\vDash\forall y(Q_jxy\rightarrow\alpha^*_{[x/y]})$. Por lo tanto, la regla de inferencia KG también preserva validez universal.

Demostremos ahora que para cualquier fórmula σ de $\mathcal{L}$-{K_1, …, K_n} vale el siguiente condicional: si $\vdash^{\mathbb{K}}\sigma$, entonces $\vDash\sigma^*_{(x)}$.

Demostración por inducción matemática sobre el número m de pasos de la demostración de σ.

Base: m=1 (la demostración de σ tiene un renglón)

Entonces σ se demuestra en un paso. Por lo tanto, σ es un axioma de $\mathbb{K}$ (A1-K). Pero sabemos que toda instancia de A1*-K* es universalmente válida. Por lo tanto, $\vDash\sigma^*_{(x)}$.

m= n, para n>1 (la demostración de σ consta de n renglones, con n>1)

HI: El condicional vale para todas las fórmulas que aparecen en los renglones anteriores a n.

P.D. El condicional vale para σ (que aparece en el renglón n).

Hay tres casos para σ: σ es axioma $\mathbb{K}$, es consecuencia de anteriores por MP, o bien, es consecuencia de anteriores por KG.

Caso 1: σ es axioma de $\mathbb{K}$. Se procede igual que en la base.

Caso 2: σ es consecuencia de anteriores por MP. Entonces σ se obtuvo, por MP, de las fórmulas $\alpha\rightarrow\sigma$ y α, que aparecen en renglones anteriores de la demostración. Por consiguiente, a estas dos fórmulas les podemos aplicar HI, de modo que $\vDash\alpha^{*}_{(x)}\rightarrow\sigma^{*}_{(x)}$ y $\vDash\alpha^{*}_{(x)}$. Como la regla de inferencia MP preserva validez universal, tenemos que $\vDash\sigma^{*}_{(x)}$.

Caso 3: σ es consecuencia de anteriores por KG. Entonces σ es de la forma $K_j\alpha$ y se obtuvo, por KG, de la fórmula α, que aparece en un renglón anterior de la demostración. En consecuencia, a esta fórmula se le puede aplicar HI, de forma que tenemos que $\vDash\alpha^{*}_{(x)}$. Como la regla de inferencia KG preserva validez universal, tenemos que $\vDash\forall y(Q_jxy\rightarrow\alpha^{*}_{[x/y]})$ (recuérdese que esta fórmula es la traducción de $K_j\alpha$ a $\mathcal{L}$-{Q1, …, Qn}). Por lo tanto, $\vDash\sigma^{*}_{(x)}$.

De la base y los casos 1-3, se sigue, por *principio de inducción matemática completa*, que:

Si $\vdash^{K}\sigma$, entonces $\vDash\sigma^{*}_{(x)}$.

Ahora bien, para el sistema de lógica epistémica $\mathbb{K}$ valen los metateoremas de corrección y de completez, los cuales, en conjunto, nos dicen que el conjunto de los teoremas de $\mathbb{K}$ y el conjunto de las fórmulas válidas de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$ son idénticos, esto es, que $\vdash^{K}\sigma$ sii $\vDash\sigma$, con σ cualquier *fbf* de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$. De esto, y del condicional anterior, podemos concluir que:

Si $\vDash\sigma$, entonces $\vDash\sigma^{*}_{(x)}$.

5.2. Consideraciones acerca de la reducción anterior

¿Cuál es el significado de esta reducción? En primer término, constituye un excelente ejemplo de la capacidad expresiva de los lenguajes de primer orden: todo lo expresable en lenguajes modales se puede traducir y, por tanto, expresar en un lenguaje de primer orden. Por consiguiente, al introducir los primeros no crecemos en expresividad con respecto a los segundos, al menos no extensionalmente, de manera que, si quisiéramos, podríamos hacer Lógica Modal empleando para ello únicamente los lenguajes de primer orden. Además, esta reducción muestra que las estructuras de la Lógica Modal son susceptibles de trabajarse como estructuras de primer orden, y que las verdades modales constituyen un subconjunto de las fórmulas o proposiciones universalmente válidas, *i. e.*, de las verdades de la lógica de primer orden. Por lo tanto, esta reducción sugiere que podríamos, en principio, prescindir de la maquinaria formal que hemos desarrollado y emprender el estudio de los conceptos modales a partir de los mismos recursos con los que se han estudiado los cuantificadores.

No obstante lo anterior, de lo que se trata aquí es de herramientas formales que permitan capturar adecuadamente las propiedades y relaciones lógicas concernientes a determinados conceptos. Consideremos entonces a la Lógica Modal y a la lógica de primer orden como recursos alternativos para el estudio formal de los conceptos modales. La cuestión es si disponemos de razones para decantarnos por alguna de ellas o si, por el contrario, se nos presentan como realizaciones completamente equivalentes. A este respecto, sabemos que cualquier verdad lógica o principio de razonamiento modal aprehensible con la Lógica Modal es también aprehensible mediante la lógica de predicados; mas, si invocamos criterios cualitativos en nuestra valoración, *v. g.*, lo intuitivo de las traducciones de proposiciones modales a un lenguaje o a otro, entonces la Lógica Modal se nos presenta como una mejor alternativa en el estudio formal de las modalidades. Como ejemplo, considérese el principio epistémico de la *propiedad de conocimiento*, *i. e.*, la idea de que el conocimiento implica verdad. Traducido al lenguaje epistémico $\mathcal{L}$-{K_1, ..., K_n}, tal principio se expresaría como sigue: "$K_j\alpha\rightarrow\alpha$", con j=1, ..., n y α cualquier fórmula, mientras que su traducción al lenguaje de predicados $\mathcal{L}$ sería ésta: "$\forall y(Q_jxy\rightarrow\alpha^*_{[x/y]})\rightarrow\alpha^*_{(x)}$", con 'y' una variable nueva. Está claro que la primera traducción refleja con mayor nitidez el contenido original del principio en cuestión; constituye, digamos, una traducción más intuitiva que la segunda[144]; esto, dicho sea de

[144] De hecho, las fórmulas φ* en un lenguaje de primer orden, *i. e.*, las traducciones de las fórmulas de un lenguaje modal a un lenguaje de primer orden, no reflejan tanto las proposiciones en lenguaje natural acerca de la necesidad o el conocimiento, por ejemplo, como sí las condiciones bajo las cuales dichas proposiciones resultan verdaderas, esto es, las restricciones que hemos impuesto a las estructuras que garantizan su verdad. Así, la propiedad de conocimiento es verdadera bajo cualquier estructura en donde

paso, no resulta sorprendente, ya que los lenguajes epistémicos han sido desarrollados justo con el objeto de capturar el discurso en torno al conocimiento, mientras que los lenguajes de primer orden encuentran su motivación en otra parte. Además, la semántica de la Lógica Modal o semántica de mundos posibles concuerda mucho mejor, respecto a la semántica de primer orden, con la idea de las modalidades, en donde se evalúa una proposición a la luz de diferentes escenarios o alternativas relacionadas, y no desde un único mundo posible. Por consiguiente, aunque la lógica de primer orden permite rescatar tantas verdades modales como la Lógica Modal misma, esta última captura mucho mejor el contenido o intencionalidad de los conceptos modales. Y si ha de importar en lógica no sólo la formalización, sino también la adecuación o ajuste de la misma a los conceptos que se pretende formalizar, entonces habrá que preferir, según parece, a la Lógica Modal por sobre la lógica de primer orden como herramienta para el análisis lógico de las modalidades.

Como nota final a este capítulo, sólo queda agregar que si bien la traducción o reducción anterior constituye una larga digresión, la juzgo tan interesante y poco conocida que considero que su inclusión en un trabajo como éste, cuya finalidad es introducir a uno de los campos más importantes de la lógica modal, se encuentra plenamente justificada. Además, sirve de punto de arranque para una reflexión más honda acerca de la naturaleza de la Lógica Modal, así como acerca de los lenguajes empleados por ella.

la relación de accesibilidad es reflexiva (todos los individuos se relacionan consigo mismos), y es esto justamente los que refleja la fórmula “$\forall y(Q_jxy \rightarrow \alpha^*_{[x/y]}) \rightarrow \alpha^*_{(x)}$” de primer orden.

6. Sistemas de Lógica Epistémica Multimodal

En los capítulos precedentes, hemos analizado sistemas epistémicos cuyo lenguaje incluye únicamente operadores de conocimiento, *i. e.*, el conjunto de operadores {K_1, ..., Kn}; en este capítulo, analizaremos sistemas formales que están definidos en un lenguaje que consta de dos tipos de operadores epistémicos, a saber, el conjunto de operadores de conocimiento recién mencionado, así como un conjunto de operadores para representar creencia, *i. e.*, el conjunto {B_1, ..., B_n}. A estos sistemas se le da el nombre de *sistemas epistémicos multimodales*.

6.1. Relación entre conocimiento y creencia en Filosofía

La relación entre conocimiento y creencia ha sido objeto de investigación filosófica desde los orígenes de la disciplina[145]. Así, Platón, hacia el final del "Menón", distingue el conocimiento de la creencia verdadera de la siguiente manera: <<Una vez que están sujetas, [las creencias verdaderas] se convierten, en primer lugar, en fragmentos de conocimientos y, en segundo lugar, se hacen estables. Por eso, precisamente, el conocimiento es de mayor valor que la recta opinión y, además, difiere aquél de ésta por su vínculo>>[146]. Según estas líneas, el conocimiento, a diferencia de la creencia, se encuentra atado a la realidad; tal atadura se consigue mediante un *aitias logismos*[147]. Este anclaje en lo real redunda en una segunda diferencia: el conocimiento permanece estable en el alma del agente mientras que la creencia tiende a ser fugitiva. Más tarde, en el "Teetetes", Platón retomará la relación entre conocimiento y creencia. En este diálogo, después de refutar dos posibles definiciones para el conocimiento, Platón llega

[145] La relación de distinción entre el conocimiento y la creencia parece ser de sentido común, pero en filosofía se retrotrae a épocas muy tempranas de la misma, al menos a los fundadores de la escuela eleática, Jenófanes y Parménides. Una característica de la tradición eleática es la devaluación de la creencia frente al conocimiento. *Cf.*, Waterfield, Robin, *Meno and other dialogues*, Oxford University Press, USA, 2005, p. 187 (nota a 97b de la paginación canónica).

[146] Platón, *Menón*, 98a. Trad. J. Calonge, Gredos (Ed.), Madrid, 1983.

[147] *Aitias logismos* suele traducirse con las expresiones 'razonamiento causal', 'discurso explicativo' o 'consideración del fundamento'.

a la conclusión de que el conocimiento es creencia verdadera a la que se le agrega un *logos*, aunque la naturaleza de este *logos* queda sin determinar.

Similar a la noción de conocimiento de Platón, encontramos, en 1956, una definición de conocimiento elaborada por A. J. Ayer en *The Problem of Knowledge*, conocida como el análisis estándar o tradicional del conocimiento: <<las condiciones necesarias y suficientes para saber que algo es el caso son, primero, que lo que uno dice saber sea verdadero; segundo, que se esté seguro de ello [que uno lo crea]; y tercero, que uno tenga el derecho de estar seguro [que uno esté justificado en creer]>>[148]. Nótese que tanto en la definición de Ayer como en la de Platón el conocimiento implica creencia, *i. e.*, "S sabe que α" implica "S cree que α".

Además de esta implicación que va del saber al creer, la investigación filosófica ha dado cuenta de algunas otras relaciones entre estas actitudes proposicionales, como se pone de manifiesto en lo que se ha dado en llamar *La paradoja de Moore*, en honor a G.E. Moore, quien fuera el primero en tematizar la cuestión[149]. La paradoja se puede plantear como sigue: en principio, no parece haber nada problemático en afirmar "α, pero S no sabe que α". No obstante, en el caso particular de que S sea justo quien afirma lo anterior (S= quien afirma), tal declaración resulta contradictoria y, por lo tanto, problemática. Esto es así ya que nuestras afirmaciones categóricas (planas o no cualificadas) comportan una pretensión de conocimiento: cuando hacemos una afirmación categórica de α, pretendemos, en cierto sentido, saber la verdad sobre el asunto, de modo que afirmar o declarar α implica, al mismo tiempo, afirmar o declarar que sabemos que α. Por lo tanto, al agregar "pero no sé que α" a nuestras declaraciones no cualificadas de α incurrimos en una contradicción, pues afirmamos al mismo tiempo que sabemos y no sabemos α[150]. Ahora bien, nótese que afirmar categóricamente una

---

[148] Alfred Jules Ayer, *The Problem of Knowledge*, Macmillan, London, 1956, p. 34. Los corchetes son mios.

[149] G. E. Moore, *Philosophical Studies,* Reoutledge & Kegan Paul, London, 1922.

[150] Habrá que decir que, en la paradoja de Moore, la palabra 'saber' no significa *creencia verdadera y justificada*, ya que en tal caso sería perfectamente coherente decir "Creo que α, estoy seguro de ello, aunque, *stricto sensu,* no sé que α (pues no estoy plenamente justificado)". En el caso de Moore, 'saber'

proposición equivale a creer dicha proposición. Así, por ejemplo, cuando Protágoras afirma “el hombre es la medida de todas las cosas”, nos está diciendo con ello que él *cree* que el hombre es la medida de todas las cosas, que el acepta como verdadero esto. Por lo tanto, al mostrarnos que las declaraciones no cualificadas implican una pretensión de conocimiento, la paradoja de Moore nos muestra que creer en una proposición implica creer que se sabe dicha proposición, *i. e.*, “S cree que α” implica “S cree que S sabe que α”. He aquí otra relación entre conocimiento y creencia.

Por último, a partir de las consideraciones anteriores, cabe señalar otra relación entre el conocimiento y la creencia que también ha sido de interés para la Filosofía[151]. Tanto en la noción de conocimiento de Platón y Ayer como en lo que toca a la paradoja de Moore, creer una proposición significa simplemente *aceptar dicha proposición como verdadera, estar convencido de su verdad*. Por otra parte, solemos escuchar expresiones de la forma “yo no creo α, lo sé” o “creo que α, pero no lo sé”. En muchos casos como éstos, sería erróneo decir que tales afirmaciones son contradictorias pues el conocimiento implica la creencia y al creer algo queremos significar con ello que lo sabemos. En expresiones como las anteriores, la palabra ‘creer’ tiene un sentido restringido: <<quiere decir tener algo por verdadero pero sin estar seguro de ello, ni contar con pruebas suficientes>>[152]. Aquí, ‘creer’ tiene el sentido de *conjeturar, suponer*, consiste en una modalidad de afirmación en la que se declara la verdad de α pero con un tono de voz sustancialmente menos confiado y firme. En este sentido, el conocimiento no implica la creencia, sino que la excluye, además de que ésta no involucra una pretensión de saber, antes bien la niega, es por ello que el poeta puede decir sin problema <<*Yo no lo sé de cierto, pero supongo* [creo] *[…]*>>. En suma, en este caso, tenemos la siguiente relación entre conocimiento y creencia: “S cree (supone, conjetura) que α” implica “S no sabe que α”.

---

posee un significado intuitivo cuyas notas específicas quedan aquí sin determinar (salvo, quizá, la certeza o seguridad).

[151] La relación entre conocimiento y creencia que presentaré a continuación es señalada en la introducción de Luis Villoro as su *Creer, saber, conocer*, Siglo XXI (Ed.), México, 1989, p. 15.

[152] *Ibíd.*

6.2. Lógica Epistémica Multimodal

Hasta este momento, he hecho referencia únicamente a la aplicación del análisis conceptual en Filosofía en torno a la relación entre el conocimiento y la creencia. Una vía alternativa consiste en emplear las herramientas de la Lógica Formal para el estudio de las múltiples relaciones entre estas dos actitudes proposicionales, *viz.,* saber y creer: éste es justo el objetivo de la LEM. Ésta, como sucede con la LE estándar, emplea lenguajes formales que son extensiones de los lenguajes proposicionales, de primer orden y de órdenes superiores; sin embargo, en el caso multimodal, los lenguajes constan de dos conjuntos de operadores modales: un conjunto de operadores para conocimiento (operadores epistémicos) así como un conjunto de operadores para creencia (operadores doxásticos). A continuación, definiremos el lenguaje proposicional epistémico multimodal $\mathcal{L}_{KB}$[153].

Definición de $\mathcal{L}_{KB}$

Símbolos de $\mathcal{L}_{KB}$:

I) Un conjunto infinito numerable de letras proposicionales
$p_1, p_2, p_3, \ldots$

II) $\neg, \rightarrow$

III) (,)

IV) Un conjunto de n operadores de conocimiento
$K_1, \ldots, K_n$

V) Un conjunto de n operadores de creencia
$B_1, \ldots, B_n$[154]

[153] Aquí me ceñiré únicamente a la Lógica Proposicional Epistémica Multimodal.

[154] El conjunto de subíndices para ambos tipos de operadores modales será exactamente el mismo.

Fbf´s de $\mathcal{L}_{KB}$:

I) Toda letra proposicional es *fbf* de $\mathcal{L}_{KB}$.

II) Si α y β son *fbf ´s*, también lo son:

$(\neg\alpha)$, $(\alpha\rightarrow\beta)$, $(K_j\alpha)$, $(B_j\alpha)$ con j=1, 2, …, n.

III) No hay más *fbf ´s*.

Definiciones

$\alpha\wedge\beta \overset{def}{=} \neg(\alpha\rightarrow\neg\beta)$

$\alpha\vee\beta \overset{def}{=} (\neg\alpha\rightarrow\beta)$

$\alpha\leftrightarrow\beta \overset{def}{=} (\alpha\rightarrow\beta)\wedge(\beta\rightarrow\alpha)$

En este lenguaje, como se verá, se pueden definir diversas teorías axiomáticas que resultan útiles para dar cuenta de algunas consecuencias de asumir ciertos postulados respecto al conocimiento, la creencia y la relación entre ambos.

6.3. El sistema axiomático KB

Definamos nuestro primer sistema axiomático para la LE multimodal, el sistema KB. Se trata, en realidad, de un subsistema del sistema bimodal KB_{CD} que S. Kraus y D. Lehmann presentaron, en 1988, con la finalidad de capturar la relación entre el conocimiento individual (K_j), la creencia individual (B_j), el conocimiento común (C) y la creencia común (D)[155]. A diferencia de este último, la teoría KB se ocupa únicamente de la relación entre conocimiento individual y creencia individual.

[155] Mientras que el *conocimiento individual* y la *creencia individual* hacen referencia, respectivamente, a los conocimientos y creencias particulares de un agente, el *conocimiento común* y la *creencia común* refieren, respectivamente, a un tipo especial de conocimiento y creencia al interior de un grupo, en donde no sólo todos los miembros del grupo conocen (creen), digamos, una proposición p, sino que, además, cualquiera de ellos sabe (cree) que todos los miembros del grupo conocen (creen) p, y así sucesivamente *ad infinitum.* Respecto al conocimiento común y su tratamiento formal, véase Gochet P. and Gribomont P., "Epistemic Logic", *Handbook of the History of Logic* (Volume 7: Logic and the Modalities in the Twentieth Century), Dov M. Gabbay and John Woods (Editors), Elsevier (Ed.), 2006, pp. 106-108. Para el

Definimos en $\mathcal{L}_{KB}$ el sistema axiomático KB.

Axiomas de KB

Sean α, β y γ *fbf* 's cualesquiera de $\mathcal{L}_{KB}$ y sea j=1, …, n.

A1: $\alpha\rightarrow(\beta\rightarrow\alpha)$,

A2: $(\alpha\rightarrow(\beta\rightarrow\gamma))\rightarrow((\alpha\rightarrow\beta)\rightarrow(\alpha\rightarrow\gamma))$,

A3: $(\neg\alpha\rightarrow\neg\beta)\rightarrow((\neg\alpha\rightarrow\beta)\rightarrow\alpha)$,

K: $K_j(\alpha\rightarrow\beta)\rightarrow(K_j\alpha\rightarrow K_j\beta)$ [Propiedad de distribución],

T: $K_j\alpha\rightarrow\alpha$ [Propiedad de conocimiento],

A5: $\neg K_j\alpha\rightarrow K_j\neg K_j\alpha$ [Propiedad de introspección negativa],

K^B: $B_j(\alpha\rightarrow\beta)\rightarrow(B_j\alpha\rightarrow B_j\beta)$ [Axioma K para creencia],

D^B: $B_j\alpha\rightarrow\neg B_j\neg\alpha$ [Axioma D para creencia],

KB1: $K_j\alpha\rightarrow B_j\alpha$,

KB2: $B_j\alpha\rightarrow K_jB_j\alpha$.

Reglas de inferencia de KB

Sean α y β *fbf* 's cualesquiera de $\mathcal{L}_{KB}$.

MP: Si $\vdash^{KB}\alpha$ y $\vdash^{KB}\alpha\rightarrow\beta$, entonces $\vdash\beta$ [*Modus ponens*] y

KG: Si $\vdash^{KB}\alpha$, entonces $\vdash^{KB} K_i\alpha$ [Generalización del conocimiento].

Definición de demostración:

Una demostración en KB es una lista finita de *fbf* 's de $\mathcal{L}_{KB}$, tal que en cada una de sus líneas aparece un axioma de KB o una *fbf* obtenida de fórmulas que aparecen en líneas anteriores por aplicación de alguna regla de inferencia de KB.

---

sistema propuesto por Kraus y Lehmann, véase Sarit Kraus and Daniel Lehmann. "Knowledge, belief and time". *Theoretical Computer Science*, 58, 1988, p. 155-174.

Definición de teorema:

Se considera un teorema de *KB* a la *fbf* que aparece en la última línea de una demostración.

Nótese que el sistema se obtiene a partir de la teoría S5 para la LE estándar al agregar cuatro axiomas[156]. K^B es la versión doxástica del axioma K para conocimiento y, por tanto, constituye una suerte de *modus ponens* al nivel de la creencia. D^B es la contraparte doxástica del axioma D para conocimiento, y en el contexto del sistema KB, tiene la función de expresar la idea de que, a diferencia del conocimiento proposicional, que precisa de la verdad de la proposición involucrada (Axioma T), la creencia únicamente requiere de consistencia: el axioma D^B se limita a decirnos que si un agente cree una proposición, entonces no cree la negación de la misma. Por su parte, a los axiomas KB1 y KB2 se les conoce como *axiomas puente*, pues manifiestan una relación entre conocimiento y creencia. El primero de ellos, expresa nuestra intuición de que el conocimiento implica creencia, intuición que, como vimos, se encuentra también recogida en la tradición filosófica. El segundo, declara una especie de introspección positiva en el ámbito de las creencias, ya que nos dice que si un agente cree una proposición, entonces él sabe que la cree. A diferencia de lo ocurre con KB1, no es claro que KB2 sea un axioma completamente intuitivo; sin embargo, una defensa de éste podría ser la siguiente: dado que la creencia es algo completamente subjetivo en carácter, *i. e.*, no requiere de ninguna nota objetiva para ser (como la verdad, por ejemplo), un agente es capaz de saber, por mera introspección, si cree o no alguna proposición. Una vez definido el sistema KB, veamos la semántica correspondiente a éste.

6.4. Semántica Formal para el sistema KB

Como en el caso de la LE estándar, definir una semántica para nuestro lenguaje bimodal $\mathcal{L}_{KB}$ consiste en dar una noción de interpretación, así como una noción de verdad para

[156] Ver la sección correspondiente a la teoría S5 en el capítulo III del presente trabajo.

éste. Para la primera, emplearemos una vez más estructuras de Kripke, de modo que una interpretación para $\mathcal{L}_{KB}$ consta de lo siguiente:

Definición de interpretación para $\mathcal{L}_{KB}$

Una estructura de Kripke, *i. e.*, una interpretación, para $\mathcal{L}_{KB}$ es un tuplo:

$\mathcal{E}_M = <S, \pi, R_1, \ldots, R_n, R^B{}_1, \ldots, R^B{}_n>$.

Donde:

1) $S \neq \emptyset$ S es un conjunto no vacío (llamado el conjunto de estados s de $\mathcal{E}_M$).

~~2) π~~: L P(S) π es una función que va de L al conjunto potencia de S, donde L es el conjunto de letras proposicionales de $\mathcal{L}_{KB}$ ($\pi(p_i) \subseteq S$).

3) Para cada j=1, ..., n, $R_j \subseteq S^2$ y $R^B{}_j \subseteq S^2$ (R y R^B reciben el nombre de relación de conocimiento y relación de creencia para j, respectivamente).

Ahora, para obtener una KB-interpretación ($\mathcal{E}_{KB}$), *i. e.*, una noción de interpretación que nos permita probar los metateoremas de corrección y completez para KB, debemos agregar las siguientes condiciones respecto a las relaciones de conocimiento y creencia:

4) Para cada j=1, ..., n, R_j es una relación de equivalencia (*i. e.*, reflexiva, simétrica y transitiva).

5) Para cada j=1, ..., n, $R^B{}_j$ es una relación serial (*i. e.*, que para todo $s \in S$ existe $s' \in S$ tal que $<s, s'> \in R^B{}_i$) y euclidiana (*i. e.*, que para cualesquiera s, t, $u \in S$ sea el caso que: si $<s, t> \in R^B{}_i$ y $<s, u> \in R^B{}_i$, entonces $<t, u> \in R^B{}_i$).

6) $R^B{}_j \subseteq R_j$.

7) Para cualesquiera s, t, $u \in S$, si $<s, t> \in R_j$ y $<t, u> \in R^B{}_j$, entonces $<s, u> \in R^B{}_j$[157].

[157] Mientras que los axiomas A1-K y K^B quedan garantizados por nuestra noción de interpretación para $\mathcal{L}_{KB}$, los axiomas restantes son rescatados por las condiciones 4) a 7) de una KB-interpretación. Así, 4) rescata los axiomas T y A5; 5), el axioma D^B; 6), el axioma KB1 y 7), el axioma KB2.

Nuestra definición de verdad para $\mathcal{L}_{KB}$ es esencialmente la misma que aquella que ofrecimos para nuestro lenguaje de LE estándar, con la salvedad de que agregaremos una cláusula adicional para los operadores B_j.

Definición de verdad para $\mathcal{L}_{KB}$

Definimos primero, recursivamente, para cada $s \in S$ y cada *fbf* σ de $\mathcal{L}_{KB}$, lo que significa:

$\mathcal{E}_M, s \vDash \sigma$ (se lee: la fórmula σ es verdadera (realizada o satisfecha) en el estado s bajo la interpretación (estructura) $\mathcal{E}_M$).

Definición de satisfacibilidad:

Cláusula 1: σ es una letra proposicional

Por lo tanto, $\sigma \in L$ (σ es una letra proposicional p_i)

$\mathcal{E}_M, s \vDash p_i$ sii $s \in \pi(p_i)$.

Cláusula 2: $\mathcal{E}_M, s \vDash \neg\sigma$ sii $\mathcal{E}_M, s \nvDash \sigma$.

Cláusula 3: $\mathcal{E}_M, s \vDash \alpha \rightarrow \beta$ sii $\mathcal{E}_M, s \nvDash \alpha$ o $\mathcal{E}_M, s \vDash \beta$.

Cláusula 4: Para j=1, …, n, $\mathcal{E}_M, s \vDash K_j\sigma$ sii $\mathcal{E}_M, s' \vDash \sigma$ para todo $s' \in S$ tal que $\langle s,s' \rangle \in R_j$.

Cláusula 5: Para j=1, …, n, $\mathcal{E}_M, s \vDash B_j\sigma$ sii $\mathcal{E}_M, s' \vDash \sigma$ para todo $s' \in S$ tal que $\langle s,s' \rangle \in R^B{}_j$.

A partir de lo anterior, podemos dar la siguiente definición de verdad para nuestro lenguaje:

Sea σ cualquier *fbf* de $\mathcal{L}_{KB}$.

Definición de verdad:

σ es verdadera en una estructura $\mathcal{E}_M$ sii para todo estado s de $\mathcal{E}_M$ se tiene que $\mathcal{E}_M, s \vDash \sigma$.

Notación: $\mathcal{E}_M \vDash \sigma$

Definamos ahora lo que significa que σ sea una fórmula válida de nuestro lenguaje (una verdad lógica epistémica de $\mathcal{L}_{KB}$):

Definición de validez lógca:

σ es válida sii para toda estructura $\mathcal{E}_M$ se tiene que $\mathcal{E}_M \vDash \sigma$

Notación: $\vDash\sigma$

Si consideramos la noción de una KB-interpretación ($\mathcal{E}_{KB}$), podemos reformular las definiciones anteriores como sigue

Sea σ cualquier *fbf* de $\mathcal{L}_{KB}$.

Definición de KB-verdad:

σ es verdadera en una estructura $\mathcal{E}_{KB}$ sii para todo estado s de $\mathcal{E}_{KB}$ se tiene que $\mathcal{E}_{KB}, s\vDash\sigma$.

Notación: $\mathcal{E}_{KB}\vDash\sigma$

Definamos ahora lo que significa que σ sea una fórmula KB-válida de nuestro lenguaje.

Definición de KB-validez:

σ es KB-válida sii para toda estructura $\mathcal{E}_{KB}$ se tiene que $\mathcal{E}_{KB}\vDash\sigma$

Notación: $\vDash^{KB}\sigma$

Dada esta semántica, *i. e.*, dada la noción de una KB-interpretación y la noción de KB-validez, se puede obtener una demostración de los metateoremas de corrección y completez para la teoría KB. La formulación de estos metateoremas quedaría de la siguiente manera, con σ cualquier *fbf* de $\mathcal{L}_{KB}$:

Corrección: Si $\vdash^{KB}\sigma$, entonces $\vDash^{KB}\sigma$.
Completez: Si $\vDash^{KB}\sigma$, entonces $\vdash^{KB}\sigma$[158].

6.5. La paradoja del creyente perfecto

De entre los teoremas de KB, destaca uno bastante controversial conocido como la *paradoja del creyente perfecto*. En el lenguaje $\mathcal{L}_{KB}$, este teorema se expresa como sigue:

---

[158] En este trabajo, no elaboraré la prueba de tales metateoremas. Para ésta, véase Sarit Kraus and Daniel Lehmann. Knowledge, belief and time. *Theoretical Computer Science*, 58, 1988, p. 155-174.

$B_jK_j\alpha\rightarrow K_j\alpha$, con α una *fbf* de $\mathcal{L}_{KB}$ y j= 1, …, n. Tal fórmula nos comunica la idea de que para un agente y una proposición arbitrarios, si el primero cree saber la segunda, entonces, *ipso facto*, la sabe. Esto es a todas luces contraintuitivo ya que nos llevaría a conclusiones absurdas como lo sería, *e. g.*, admitir que en cada ocasión en que el arqueólogo Heinrich Schliemann creyó saber que una ciudad hallada en la colina de Hisarlik era la antigua ciudad de Troya de los relatos homéricos, lo sabía en efecto[159]. Además de esta contradicción con el sentido común, la paradoja en cuestión también entra en conflicto con algunos tópicos de la historia de la filosofía. Por ejemplo, las refutaciones de los diálogos platónicos, en las que Sócrates exhibe la ignorancia de aquellos que tan sólo creen saber pero no saben en realidad, se tornarían imposibles, pues bastaría con que el interlocutor de Sócrates creyera saber una definición para que, en efecto, la supiera. Del mismo modo, la formulación de los casos Gettier sería en extremo difícil, cuando imposible, pues en todos o en la mayoría de estos casos figura un agente que cree saber una proposición que de hecho no sabe.

Vale la pena exhibir aquí la demostración en KB de la paradoja del creyente perfecto dado que esto nos permitirá identificar los axiomas que permiten su derivación:

$\vdash^{KB} B_jK_j\alpha\rightarrow K_j\alpha$

1. $K_j\neg K_j\alpha\rightarrow B_j\neg K_j\alpha$	***KB1*** *($\alpha/\neg K_j\alpha$)*
2. $B_jK_j\alpha\rightarrow\neg B_j\neg K_j\alpha$	***D^B*** *($\alpha/K_j\alpha$)*
3. $B_j\neg K_j\alpha\rightarrow\neg B_jK_j\alpha$	*2, Contraposición*
4. $K_j\neg K_j\alpha\rightarrow\neg B_jK_j\alpha$	*1, 3 Silogismo hipotético*
5. $\neg K_j\alpha\rightarrow K_j\neg K_j\alpha$	***A5***
6. $\neg K_j\alpha\rightarrow\neg B_jK_j\alpha$	*5, 4 Silogismo hipotético*
7. $B_jK_j\alpha\rightarrow K_j\alpha$	*6, Contraposición*

[159] H. Schliemann fue uno de los pioneros en la búsqueda de la ciudad perdida de Príamo. En la colina de Hisarlik, cerca de los Dardanelos, Schliemann halló nueve ciudades superpuestas, cada una levantada sobre las ruinas de la anterior. De entre éstas, es la ciudad conocida como Troya VII la mejor candidata para identificarse como la ciudad de Príamo del relato.

8. $K_j\alpha \rightarrow \alpha$ *T*

9. $B_jK_j\alpha \rightarrow \alpha$ *7, 8 Silogismo hipotético*

10. $\neg\alpha \rightarrow \neg B_jK_j\alpha$ *9, Contraposición*

Hagamos dos observaciones sobre esta demostración. Primero, repárese en que, a partir de la paradoja del creyente perfecto, la cual aparece en el séptimo paso de nuestra demostración, y del axioma T, se obtiene otro teorema igualmente contraintuitivo, con el que termina la demostración: $\neg\alpha \rightarrow \neg B_jK_j\alpha$. Tal teorema nos dice que los agentes sólo creen saber proposiciones que en efecto son verdaderas. Esto resulta absurdo cuando pensamos, por ejemplo, que por mucho tiempo, al menos hasta el siglo XVII, mucha gente creyó saber que nuestro planeta tenía apenas unos cuantos miles de años de existencia, y esto pese a su falsedad. Por otro lado, nuestra demostración hace patente que los axiomas cruciales para la obtención de la paradoja del creyente perfecto son los axiomas modales KB1, D^B y A5. Por tanto, si queremos hacer imposible su derivación, una alternativa consiste en remover del sistema alguno de estos tres axiomas[160]. De hecho, más allá de su interés intrínseco, la importancia de esta paradoja reside justamente en que ha motivado diversos desarrollos formales encaminados a evitar su demostración. En lo que sigue, expondré tres de estas propuestas formales, cada una de las cuales tiene como eje central la remoción de uno de los tres axiomas antes mencionados.

6.6. El sistema OK & RIB

En 1992, Frans Voorbraak presentó una teoría bimodal en el lenguaje $\mathscr{L}_{KB}$ para la LE: el sistema OK & RIB[161]. Para tornar imposible la demostración de la paradoja del creyente

---

[160] Otra vía sería prohibir, en algunos de los tres esquemas axiomáticos, que α se sustituya por *fbf's* de carácter modal.

[161] OK (objective knowledge) and RIB (rational introspective belief). La expresión 'conocimiento objetivo' hace referencia aquí a un tipo de conocimiento susceptible de aplicarse a cualquier agente capaz de procesar información, con independencia de si se le pueden adscribir o no estados de creencia consciente. Por consiguiente, la forma de conocimiento que es axiomatizada en OK & RIB es aplicable a objetos o dispositivos tales como un termostato o un receptor de televisión. Ver Voorbraak F.,

perfecto, Voorbraak elimina el axioma $K_j\alpha\rightarrow B_j\alpha$ (KB1), y en su lugar adopta el axioma $B_j\alpha\rightarrow B_jK_j\alpha$ (KB3), el cual rescata la relación entre conocimiento y creencia involucrada en la paradoja de Moore que analizamos en la primera sección de este capítulo, *i. e.* "S cree que α" implica "S cree que S sabe que α"[162]. Los axiomas del sistema de Voorbraak son, entonces:

Axiomas de OK & RIB

Sean α, β y γ *fbf 's* cualesquiera de $\mathcal{L}_{KB}$ y sea j=1, ..., n.

A1: $\alpha\rightarrow(\beta\rightarrow\alpha)$,

A2: $(\alpha\rightarrow(\beta\rightarrow\gamma))\rightarrow((\alpha\rightarrow\beta)\rightarrow(\alpha\rightarrow\gamma))$,

A3: $(\neg\alpha\rightarrow\neg\beta)\rightarrow((\neg\alpha\rightarrow\beta)\rightarrow\alpha)$,

K: $K_j(\alpha\rightarrow\beta)\rightarrow(K_j\alpha\rightarrow K_j\beta)$ [Propiedad de distribución],

T: $K_j\alpha\rightarrow\alpha$ [Propiedad de conocimiento],

A5: $\neg K_j\alpha\rightarrow K_j\neg K_j\alpha$ [Propiedad de introspección negativa],

K^B: $B_j(\alpha\rightarrow\beta)\rightarrow(B_j\alpha\rightarrow B_j\beta)$ [Axioma K para creencia],

D^B: $B_j\alpha\rightarrow\neg B_j\neg\alpha$ [Axioma D para creencia],

KB2: $B_j\alpha\rightarrow K_jB_j\alpha$ [Axioma puente],

KB3: $B_j\alpha\rightarrow B_jK_j\alpha$ [Axioma puente][163].

Respecto a este último axioma (KB3), cabe preguntar <<si Goldbach creía en la verdad de su conjetura, ¿de ello se sigue, como resultaría del axioma de Voorbraak, que Goldbach creía también que él sabía su verdad?>>[164]. La respuesta es no debido a que en la cuestión anterior 'creer' parece significar *conjeturar*, mientras que con el sistema

---

Generalized Kripke models for epistemic logic, Morgan Kaufmann Publishers, San Mateo, 1992, pp. 214-228.

[162] Para la prueba de que tanto en el sistema OK E RIB como en los dos sistemas bimodales que se expondrán en el resto de este capítulo no se demuestra cualquier instancia de la paradoja del creyente perfecto, véase el apéndice III.

[163] Tanto las reglas de inferencia como la definición de demostración y teorema para OK E RIB son esencialmente las mismas, *mutatis mutandis*, que para la teoría KB.

[164] Gochet P. and Gribomont P., "Epistemic Logic", *Handbook of the History of Logic* (Volume 7: Logic and the Modalities in the Twentieth Century)..., p. 115.

OK & RIB se intenta axiomatizar justo el sentido de creencia implicado en la paradoja de Moore, *viz., estar cierto de una proposición* o *estar convencido de su verdad*. Por otra parte, se debe plantear la interrogante acerca de en qué consiste exactamente el *conocimiento objetivo* (que es tipo de conocimiento que Voorbraak pretende formalizar), pues al no precisar del axioma KB1: $K_j\alpha \rightarrow B_j\alpha$, tal forma de conocimiento no sólo entra en conflicto con el sentido común, sino que rompe con una larga tradición de análisis conceptual del conocimiento que, como vimos, se remonta hasta la filosofía de Platón.

6.7. El sistema KB-$\{D^B\}$

La segunda alternativa para evitar la demostración de la paradoja del creyente perfecto consiste en eliminar del sistema KB el axioma D^B; el resultado de tal eliminación es la teoría bimodal KB-$\{D^B\}$. Tanto en esta teoría como en KB es posible demostrar la fórmula $(B_j\alpha \wedge B_j\beta) \leftrightarrow B_j(\alpha \wedge \beta)$. Ya que su demostración es idéntica en ambos sistemas, expongámosla en cualquiera de ellos.

$\vdash^{KB} (B_j\alpha \wedge B_j\beta) \leftrightarrow B_j(\alpha \wedge \beta)$

1. $(\alpha \wedge \beta) \rightarrow \alpha$	*Simplificación*
2. $B_j(\alpha \wedge \beta) \rightarrow B_j\alpha$	*1, RD1^B*
3. $(\alpha \wedge \beta) \rightarrow \beta$	*Simplificación*
4. $B_j(\alpha \wedge \beta) \rightarrow B_j\beta$	*3, RD1^B*
5. $B_j(\alpha \wedge \beta) \rightarrow (B_j\alpha \wedge B_j\beta)$	*2, 4, Implicación y MP(x2)*
6. $\alpha \rightarrow (\beta \rightarrow (\alpha \wedge \beta))$	*Adjunción*
7. $B_j\alpha \rightarrow B_j(\beta \rightarrow (\alpha \wedge \beta))$	*6, RD1^B*
8. $B_j(\beta \rightarrow (\alpha \wedge \beta)) \rightarrow (B_j\beta \rightarrow B_j(\alpha \wedge \beta))$	*K^B $(\alpha/\beta, \beta/(\alpha \wedge \beta))$*
9. $B_j\alpha \rightarrow (B_j\beta \rightarrow B_j(\alpha \wedge \beta))$	*7, 8, Transitividad y MP(x2)*
10. $(B_j\alpha \wedge B_j\beta) \rightarrow B_j(\alpha \wedge \beta)$	*9, Importación/Exportación*
11. $(B_j\alpha \wedge B_j\beta) \leftrightarrow B_j(\alpha \wedge \beta)$	*5, 10, Equivalencia y MP(x2)*[165]

[165] En la prueba anterior se emplea la regla derivada RD1^B, cuya prueba sería la siguiente:
Sean α y β *fbf´s* cualesquiera de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$ y sea j=1, …, n.

Ahora bien, a partir de una instancia de este teorema (al que se le conoce como *principio de factorización*) se demuestra en KB la fórmula $\neg B_j(\alpha\wedge\neg\alpha)$.

$\vdash^{KB}\neg B_j(\alpha\wedge\neg\alpha)$

1. $(B_j\alpha\wedge B_j\neg\alpha)\leftrightarrow B_j(\alpha\wedge\neg\alpha)$	*Principio de factorización* $(\alpha/\alpha, \beta/\neg\alpha)$
2. $\neg(B_j\alpha\wedge B_j\neg\alpha)\leftrightarrow\neg B_j(\alpha\wedge\neg\alpha)$	*1, Tautología*
3. $\neg(B_j\alpha\wedge B_j\neg\alpha)\rightarrow\neg B_j(\alpha\wedge\neg\alpha)$	*2, Tautología*
4. $B_j\alpha\rightarrow\neg B_j\neg\alpha$	D^B
5. $\neg(B_j\alpha\wedge B_j\neg\alpha)$	*4, Def.* $\wedge$
6. $\neg B_j(\alpha\wedge\neg\alpha)$	*3, 5 y MP*

Esto significa que la teoría KB prohíbe explícitamente, para cualquier agente, creer una contradicción. Por el contrario, en KB-$\{D^B\}$, al remover el axioma D^B: $B_j\alpha\rightarrow\neg B_j\neg\alpha$ (que es lógicamente equivalente a la *fbf* $\neg(B_j\alpha\wedge B_{j\neg}\alpha)$), no se tiene como teorema la fórmula $\neg B_j(\alpha\wedge\neg\alpha)$, y esto significa que con la remoción del axioma D^B la teoría KB-$\{D^B\}$ evita la derivación de la paradoja del creyente perfecto, pero a un alto precio, pues ésta no sólo contempla la posibilidad de que los agentes crean una proposición así como su negación, sino que con ello deja abierta la posibilidad de que éstos crean una contradicción[166]. Si bien ambas posibilidades repugnan a nuestro sentido común, existe una diferencia importante entre ellas. En sentido estricto, aceptar la posibilidad de la

---

Supongamos que $\vdash^{\mathcal{K}}\alpha\rightarrow\beta$. Entonces existe una demostración en $\mathcal{K}$ que termina en la fórmula '$\alpha\rightarrow\beta$'. Podemos entonces continuar la demostración como sigue y llegar a la fórmula '$B_j\alpha\rightarrow B_j\beta$':

- --------

m. $\alpha\rightarrow\beta$	
m+1. $K_j(\alpha\rightarrow\beta)$	*m, KG*
m+2. $K_j(\alpha\rightarrow\beta)\rightarrow B_j(\alpha\rightarrow\beta)$	*KB1* $(\alpha/(\alpha\rightarrow\beta))$
m+4. $B_j(\alpha\rightarrow\beta)$	*m+1, m+2 y MP*
m+5. $B_j(\alpha\rightarrow\beta)\rightarrow(B_j\alpha\rightarrow B_j\beta)$	K^B
M+6. $(B_j\alpha\rightarrow B_j\beta)$	*m+4, m+5 y MP*

[166] Respecto a la prueba de que no toda instancia de los esquemas $B_j\alpha\rightarrow\neg B_j\neg\alpha$ y $\neg B_j(\alpha\wedge\neg\alpha)$ es teorema de KB-$\{D^B\}$, ver el apéndice III.

fórmula $B_j\alpha\wedge B_j\neg\alpha$ no significa admitir la posibilidad de que el agente j crea una contradicción, ya que ésta sólo declara que el agente cree proposiciones contradictorias, *i. e.*, que cree la proposición α así como su negación. Más todavía, es posible aceptar $B_j\alpha\wedge B_j\neg\alpha$ y rechazar $B_j(\alpha\wedge\neg\alpha)$. Ésta es justo la posición, *e. g.*, del filósofo norteamericano Donal Davidson en su explicación del fenómeno del autoengaño, que es un caso de *irracionalidad* muy común en el que un individuo cree tanto una proposición como su negación[167]. Para dar cuenta de la posibilidad del autoengaño, Davidson recurre a la hipótesis de la división de la mente, *i. e.*, a la idea de que en la mente existen divisiones o conjuntos de creencias no conectados entre sí. De este modo, en la mente de un agente que se autoengaña existen dos conjuntos de creencias desvinculados entre sí: uno, al que pertenece la creencia en la proposición α y otro, al que pertenece la creencia en la proposición $\neg\alpha$. Dado que tales creencias coexisten de forma separada, el agente no está obligado a combinarlas y, por tanto, a creer la contradicción $\alpha\wedge\neg\alpha$. En este sentido, <<un agente es similar a una comunidad en la que diferentes personas pueden tener opiniones distintas, aunque ninguna defenderá contradicciones. En pocas palabras, creencias que derivan de diferentes marcos o estados mentales no tienen por qué ser combinadas por el agente>>[168]. Estas ideas de Davidson en Filosofía de la Mente tienen su equivalente en LE. Por ejemplo, R. Barcan Marcus propone una lógica para la creencia en la que se rechaza el axioma D^B: $B_j\alpha\rightarrow\neg B_j\neg\alpha$ así como la inferencia de $B_j(\alpha\wedge\neg\alpha)$ a partir de $(B_j\alpha\wedge B_{j\neg}\alpha)$[169]. Por otro lado, Ronald Fagin y Joseph Y. Halpern han desarrollado un grupo de semánticas mediante

---

[167] Para el tratamiento de D. Davidson del autoengaño, véase *Problems of Irrationality*; específicamente, los ensayos “Paradoxes of Irrationality” y “Deception and Division”.

[168] Thijsse, Elias, *Partial Logic and Knowledge Representation*, PhD thesis, 1992, p. 170.

[169] De hecho, R. Barcan rechaza una de las implicaciones del principio de factorización, *viz.*, $(B_j\alpha\wedge B_j\neg\alpha)\rightarrow B_j(\alpha\wedge\neg\alpha)$. Con ello, la lógica doxástica que elabora deja de ser un sistema regular de Lógica Modal. Ver Gochet P. and Gribomont P., “Epistemic Logic”, *Handbook of the History of Logic* (Volume 7: Logic and the Modalities in the Twentieth Century)…, p. 116.

las cuales es posible modelar la división o compartimentación de la mente de un agente que cree dos proposiciones sin creer la conjunción de ambas[170].

6.8. El sistema KB^{PG}

La tercera vía (y la última que se comentará en este capítulo) para eludir la paradoja del creyente perfecto es renunciar a la introspección negativa, *i. e.*, al axioma A5: $\neg K_j\alpha \rightarrow K_j\neg K_j\alpha$.

En *Epistemic logic*, Paul Gochet y Pascal Gribomot presentan un sistema bimodal para la LE en el que prescinden del axioma A5, y al que denominaremos el sistema KB^{PG}[171]. Su conjunto de axiomas es el siguiente:

Axiomas de KB^{PG}

Sean α, β y γ *fbf* 's cualesquiera de $\mathcal{L}_{KB}$ y sea j=1, …, n.

A1: $\alpha \rightarrow (\beta \rightarrow \alpha)$,

A2: $(\alpha \rightarrow (\beta \rightarrow \gamma)) \rightarrow ((\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrow \gamma))$,

A3: $(\neg\alpha \rightarrow \neg\beta) \rightarrow ((\neg\alpha \rightarrow \beta) \rightarrow \alpha)$,

K: $K_j(\alpha \rightarrow \beta) \rightarrow (K_j\alpha \rightarrow K_j\beta)$ [Propiedad de distribución],

T: $K_j\alpha \rightarrow \alpha$ [Propiedad de conocimiento],

K^B: $B_j(\alpha \rightarrow \beta) \rightarrow (B_j\alpha \rightarrow B_j\beta)$ [Axioma K para creencia],

D^B: $B_j\alpha \rightarrow \neg B_j\neg\alpha$ [Axioma D para creencia],

KB1: $K_j\alpha \rightarrow B_j\alpha$ [Axioma puente],

KB3: $B_j\alpha \rightarrow B_jK_j\alpha$ [Axioma puente][172].

Con la remoción del axioma de introspección negativa, los autores consiguen no sólo evadir la paradoja del creyente perfecto, sino que, al mismo tiempo, evitan la

---

[170] Ronald Fagin and Joseph Y. Halpern, *Belief, awareness, and limited reasoning. Artificial Intelligence,* 34:39–76, 1988.

[171] Véase Gochet P. and Gribomont P., "Epistemic Logic", *Handbook of the History of Logic* (Volume 7: Logic and the Modalities in the Twentieth Century)…, pp. 116-7.

[172] Tanto las reglas de inferencia como la definición de demostración y teorema para KB^{PG} son las mismas que para los sistemas bimodales anteriores.

demostración en el sistema de la fórmula $B_j\alpha\rightarrow\alpha$, o fórmula (Ω), la cual se conoce como *paradoja de la infalibilidad*, pues nos dice que los agentes nunca yerran en sus creencias, *i. e.*, que cualquier proposición que crean será verdadera. Para aclarar la relación entre el principio de introspección negativa y la paradoja de la infalibilidad, haré uso de una objeción que el filósofo alemán W. Lenzen formuló en contra del primero en su *Recent work in epistemic logic*: <<si un individuo *a* está completamente persuadido de la verdad de *p*, pero se equivoca al respecto, entonces él evidentemente no sabe que *p,* pese a que cree saberlo; y, por tanto, está lejos de saber que no sabe que *p*>>[173]. Podemos reconstruir esta crítica empleando los axiomas de los sistemas bimodales anteriores:

Sea j un individuo cualquiera y sea α una proposición arbitraria.

P1) $B_j\alpha\wedge\neg\alpha$	(hipótesis)
Pi1) $B_j\alpha$	(de P1, por simplificación)
Pi2) $\neg\alpha$	(de P1, por simplificación)
P2) $K_j\alpha\rightarrow\alpha$	(axioma T)
Pi3) $\neg K_j\alpha$	(de P2 y Pi2, por *Modus Tollens*)
P3) $B_j\alpha\rightarrow B_jK_j\alpha$	(axioma KB3)
Pi4) $B_jK_j\alpha$	(de P3 y Pi1, por *Modus Ponens*)
P4) $B_jK_j\alpha\rightarrow\neg B_j\neg K_j\alpha$	(axioma D^B)
Pi5) $\neg B_j\neg K_j\alpha$	(de P4 y Pi4, por *Modus Ponens*)
P5) $K_j\neg K_j\alpha\rightarrow B_j\neg K_j\alpha$	(axioma KB1)
Pi6) $\neg K_j\neg K_j\alpha$	(de P5 y Pi5, por *Modus Tollens*)
Pi7) $\neg K_j\alpha\wedge\neg K_j\neg K_j\alpha$	(de Pi3 y Pi6, por adjunción)

$\therefore$

C) $\neg(\neg K_j\alpha\rightarrow K_j\neg K_j\alpha)$ (de Pi7, por equivalencia lógica)

Tal reconstrucción del argumento de Lenzen muestra que, a partir de la hipótesis inicial ($B_j\alpha\wedge\neg\alpha$) y de los axiomas T, D^B, KB1 y KB3, se sigue lógicamente la negación del

[173] Wolfgang, Lenzen, *Recent work in epistemic logic*, Acta Philosophica Fennica, 30:5–219, 1978, p. 79.

axioma de introspección negativa (en este sentido, la objeción de Lenzen puede leerse como sigue: dado que los agentes creen proposiciones falsas y estos cuatro axiomas son verdaderos –o al menos sumamente razonables–, tenemos que rechazar la introspección negativa como un principio del conocimiento proposicional). Por consiguiente, si partimos de A5: $\neg K_j\alpha \rightarrow K_j\neg K_j\alpha$ y los cuatro axiomas recién mencionados, podemos concluir la fórmula $\neg(B_j\alpha \wedge \neg\alpha)$, que es lógicamente equivalente a $B_j\alpha \rightarrow \alpha$ (Ω). Puesto en otras palabras, cualquier sistema bimodal en $\mathcal{L}_{KB}$ con los axiomas A5, T, D^B, KB1 y KB3, además la regla MP, tendrá tanto a la paradoja del creyente perfecto como a la paradoja de la infalibilidad como teoremas. Veamos la demostración en cuestión:

1. $K_j\neg K_j\alpha \rightarrow B_j\neg K_j\alpha$	***KB1*** *(α/¬K_jα)*
2. $B_jK_j\alpha \rightarrow \neg B_j\neg K_j\alpha$	***D^B*** *(α/K_jα)*
3. $B_j\neg K_j\alpha \rightarrow \neg B_jK_j\alpha$	*2, Contraposición*
4. $K_j\neg K_j\alpha \rightarrow \neg B_jK_j\alpha$	*1, 3 Silogismo hipotético*
5. $\neg K_j\alpha \rightarrow K_j\neg K_j\alpha$	***A5***
6. $\neg K_j\alpha \rightarrow \neg B_jK_j\alpha$	*5, 4 Silogismo hipotético*
7. $B_jK_j\alpha \rightarrow K_j\alpha$	*6, Contraposición*
	[paradoja del creyente perfecto]
8. $K_j\alpha \rightarrow \alpha$	***T***
9. $B_jK_j\alpha \rightarrow \alpha$	*7, 8 Silogismo hipotético*
10. $B_j\alpha \rightarrow B_jK_j\alpha$	***KB3***
11. **$B_j\alpha \rightarrow \alpha$**	*10, 9 Silogismo hipotético* **[Ω]**

Por lo tanto, con el sistema KB^{PG}, en el que no se demuestra la fórmula de introspección negativa ($\neg K_j\alpha \rightarrow K_j\neg K_j\alpha$), se elude a un tiempo la paradoja del creyente perfecto, así como la paradoja de infalibilidad[174].

[174] Para la prueba de que no toda instancia del esquema $\neg K_j\alpha \rightarrow K_j\neg K_j\alpha$, así como no toda instancia de la paradoja de la infalibilidad, es teorema de este sistema, véase el apéndice III.

Conclusiones

Como epílogo a estas consideraciones, dejo al lector una mención, aunque somera, de uno de los problemas centrales de la Lógica Epistémica Modal, y al que he aludido de pasada en la sección última del capítulo II[175], *viz.*, la omnisciencia lógica. Por otra parte, también le entrego una consideración final en torno a esta lógica, en la que se presenta un balance general de todo lo hasta aquí expuesto, así como una consecuencia de ello para la teoría del conocimiento.

C.1. El problema de la omniscencia lógica

Entre las objeciones más importantes que se esgrimen en contra de la Lógica Epistémica Modal se halla el problema de la omnisciencia lógica u omnisciencia deductiva, *i. e.*, la idea de que los agentes epistémicos saben todas las consecuencias lógicas de las proposiciones por ellos conocidas. En otras palabras:

> Sea *S* un agente cualquiera y sean α y β proposiciones cualesquiera.
>
> *Si S sabe que α, y α implica lógicamente a β, entonces S también sabe que β.*

En los sistemas axiomáticos que hemos estudiado se encuentran elementos suficientes para reconstruir esta idea, *viz.*, la regla de inferencia KG y el axioma K, por lo que se puede afirmar que la omnisciencia lógica es un supuesto de todos estos sistemas. Sin embargo, quisiera analizar aquí el problema no desde una perspectiva sintáctica, sino a

[175] *Cf.*, la sección "Consideraciones en torno al sistema K".

partir de la semántica misma que hemos definido para nuestras teorías formales, *i. e.*, la semántica de mundos posibles.

Partiré del hecho de que toda instancia de las tautologías proposicionales en el lenguaje $\mathcal{L}$-{K_1, ..., K_n} es una *fbf* válida de dicho lenguaje. Lo anterior tiene como consecuencia, dada la cláusula modal de nuestra definición de verdad [Para j=1, ..., n, $\mathcal{E}$, s⊨$K_j\sigma$ sii $\mathcal{E}$, s'⊨σ para todo s'∈S tal que <s,s'>∈R_j], que cualquier instancia del esquema '$K_j\varphi$' es una fórmula válida de $\mathcal{L}$-{K_1, ..., K_n} siempre que 'φ' es instancia de tautología[176]. En otras palabras, si adoptamos la semántica de mundos posibles tal como ha sido definida, entonces nos vemos obligados a aceptar que cualquier agente epistémico conoce todas las tautologías proposicionales; esto último, claro está, resulta controversial, cuando no falso[177]. Más aún, esta consecuencia, ya de suyo problemática, nos conduce a la omnisciencia lógica. Veamos cómo. Hemos establecido en el capítulo II que cualquier instancia del esquema $K_j(\alpha\rightarrow\beta)\rightarrow(K_j\alpha\rightarrow K_j\beta)$ [axioma K] es una verdad lógica de $\mathcal{L}$-{K_1, ..., K_n}. Esto significa que nuestra semántica nos obliga a aceptar lo que antes he llamado el *supuesto de preservación de conocimiento*, que es una versión epistémica del *Modus Ponens*[178]. Por lo tanto, si los agentes conocen todas las tautologías (incluida cualquier implicación que exprese una relación de consecuencia lógica entre proposiciones), conocen también (dada la validez del supuesto anterior) cualquier proposición que se siga lógicamente de aquello que conocen. En otros términos, la semántica de mundos posibles nos fuerza a conceder que los agentes de conocimiento son lógicamente omniscientes, *i. e.*, que conocen, de hecho, todas las

[176] Nótese que las instancias de tautologías, así como de '$K_j\varphi$', donde 'φ' es instancia de tautología, también son fórmulas T-válidas, S4-válidas y S5-válidas.

[177] Se ha podido decir, sea lo que esto signifique, que los agentes poseen conocimiento implícito, entre el que se cuenta, por ejemplo, el conocimiento de las tautologías proposicionales. Al respecto, ver las secciones *2.9 Implicit Versus Explicit Beliefs* y *6.2 Belief, A Borderline Concept Between Logic And Psychology*, en Gochet P. and Gribomont P., "Epistemic Logic", *Handbook of the History of Logic* ..., en específico, pp. 119-120 y 157-158.

[178] El supuesto de preservación del conocimiento declara que si un agente sabe que una proposición implica a otra y conoce la primera de ellas, entonces (por *Modus Ponens*) también conoce la segunda proposición. Ver la sección final de capítulo II.

consecuencias lógicas de aquellas proposiciones por ellos sabidas[179]. El problema con la idea de la omnisciencia lógica es que resulta falsa si contemplamos a los seres humanos dentro del universo de los agentes epistémicos: la habilidad para extraer e identificar las consecuencias lógicas de nuestro conocimiento es variable entre individuos, pues depende de entrenamiento y sagacidad intelectual, entre otras cosas; lo que es evidente para una inteligencia excepcional podría permanecer oculto y, por tanto, desconocido para un número importante de personas. Además, es improbable no ya que los seres humanos seamos lógicamente omniscientes, sino siquiera que alguno de nosotros sea un razonador tal que conozca de hecho cualquier proposición que se halle implicada por su conocimiento. Por otro lado, la tesis de la omnisciencia lógica lleva a consecuencias absurdas. Así, en caso de que ésta fuera cierta, disciplinas como la Matemática, en la que se indaga si una determinada proposición es teorema o *consecuencia lógica* de un conjunto de axiomas, perdería gran parte de su relevancia, ya que al conocer los axiomas se conocería también el conjunto entero de sus consecuencias lógicas[180]. Por lo tanto, la tesis de la omnisciencia lógica resulta insostenible cuando se halla referida al saber humano.

Lo anterior nos proporciona un primer elemento para responder a la cuestión acerca de si la semántica de mundos posibles es adecuada o no para el análisis de los conceptos epistemológicos: el problema de la omnisciencia lógica muestra que la versión, digamos, estándar de esta semántica tiene que ser objeto de alguna modificación si es que ha de servir para capturar de modo satisfactorio el concepto de

---

[179] En sentido estricto, la omnisciencia lógica está restringida aquí al empleo de la Lógica Proposicional tanto estándar como epistémica; sin embargo, dicho problema se puede formular también en una semántica y un lenguaje para la Lógica Modal o Epistémica de primer orden.

[180] *Inter nos,* siempre me ha parecido que algo muy cercano a la omnisciencia lógica es sostenido cuando se afirma que la deducción no incrementa nuestro conocimiento debido a que su conclusión ya está contenida en las premisas, como si el saber no pudiera nutrirse en absoluto de las investigaciones concernientes al descubrimiento de las consecuencias deductivas de nuestras creencias.

conocimiento[181]. En otras palabras, a menos que se deje de lado el conocimiento proposicional humano y se restrinja la investigación a un conjunto de razonadores ideales, la semántica de mundos posibles tal como la hemos desarrollado aquí resulta inadecuada para analizar el saber como concepto modal.

C.2. Consideración final en torno a la Lógica Epistémica

Veamos ahora adónde hemos arribado a lo largo de nuestras consideraciones. En primer término, alcanzamos la idea de que tanto la cláusula modal epistémica como la tesis de preservación del conocimiento [principio K] abren la puerta al escepticismo en Epistemología: ésta, mediante su uso en el argumento escéptico de la transmisibilidad, y aquélla, por suscribir una noción de conocimiento infalible, la cual es prácticamente imposible de satisfacer. En estrecha relación con esto último, se halla nuestro resultado acerca del principio del conocimiento [principio T], según el cual dicho principio resulta incompatible con algunas concepciones falibilistas del saber; o, puesto de otra forma, la verdad como condición necesaria para saber nos expone al peligro de hacer del conocimiento, tal como sucede con la cláusula modal, algo infalible. En lo tocante a las tesis de introspección positiva y negativa [principios A4 y A5], hemos establecido, a través del argumento del regreso epistémico, que piden demasiado al saber y a la ignorancia, respectivamente, pues exige para ambos el conocimiento —y, por tanto, la justificación— de una cadena infinita de proposiciones, cada una de un orden mayor a las anteriores. Además, mostramos que justo este problema, el del regreso epistémico, parece inocuo para cualquier posición epistemológica la cual, además de la introspección positiva, suscriba el principio T. Por último, probamos que la introspección negativa es lógicamente incompatible con otras tesis epistémicas tomadas en conjunto, a saber: la tesis T, D^B [el supuesto de que, si se cree algo, entonces no se cree su negación], KB1 [el principio de que la creencia está implicada en el saber]

[181] Para tener un panorama general de las modificaciones en Lógica Epistémica encaminadas a resolver el problema de la omnisciencia lógica, ver la sección *6. LOGICAL OMNISCIENCE AND EPISTEMIC LOGIC,* en Gochet P. and Gribomont P., “Epistemic Logic”, *Handbook of the History of Logic* ..., pp. 157-168.

y KB3 [o la tesis de que, si cree algo, entonces se cree saberlo]. Preguntémonos entonces por el significado de todo lo anterior para la Lógica Epistémica y, en general, para la Epistemología; ¿qué lección sacaremos de todo ello?

Mi contestación dependerá, en lo fundamental, de dos supuestos. Primero, todo campo de investigación genuinamente racional debe comprometerse con el aprendizaje a partir de los errores y, por consiguiente, deberá contemplar suficientes mecanismos para la detección de errores o fallos generales[182]. Segundo, las pretensiones o atribuciones de conocimiento son, en cierto sentido, contextuales; de modo que una pretensión de conocimiento válida desde la perspectiva de la vida cotidiana podría no serlo, en principio, desde el punto de vista del epistemólogo[183]. Pasemos, pues, a la contestación.

Todos los resultados recién enumerados no parecen arrojar un saldo favorable para la Lógica Epistémica Modal; consideraciones como las anteriores, en conjunción con el problema de la omnisciencia deductiva, nos llevan a cuestionar la naturaleza genuinamente epistémica de semejante lógica: ¿no es ésta simplemente un intento curioso pero fallido de aplicar Lógica Modal al análisis de conceptos epistemológicos? Más aún, ¿no se reduce a un juego formal que nada tiene que ver con el conocimiento efectivo, y que, a la larga, termina deformando los conceptos que originalmente

[182] Este supuesto metodológico lo he tomado prestado de Charles S. Peirce; *cf.*, Laudan, Larry, "Detectar errores y aprender de ellos en un sistema de apelaciones asimétricas", en: *Pruebas y estándares de prueba en el derecho*, UNAM, México, 2010, pp. 107-122. Ahora, cabe distinguir aquí entre equivocaciones y errores. Las primeras tienen que ver con fallas resultantes de la mala aplicación de una regla, o con la utilización incorrecta de una herramienta. Por ejemplo, cuando nos equivocamos al sumar: disponemos de un algoritmo para la suma el cual no hemos seguido al pie de la letra; de haberlo hecho, hubiésemos llegado al resultado correcto. Por su parte, los errores son fallos inducidos por las reglas o herramientas mismas. En nuestro caso, las fallas detectadas (la infalibilidad, por ejemplo) son producto de los conceptos o reglas de discernimiento de la Epistemología, y no resultado de su mala aplicación.

[183] Para aclarar este contextualismo, téngase en cuenta, *v. g.*, que los estándares de la justificación para declarar legítimamente que se sabe algo son, con seguridad, distintos en el contexto de la vida cotidiana y en los contextos de la ciencia o la Filosofía. Ver la sección "Usos de "saber" para fines prácticos", en el capítulo VI de *Creer, saber, conocer*, de Luis Villoro.

pretendía formalizar? Más de una vez se me han dirigido estas preguntas, aun yo mismo llegué a sospechar algo similar a lo largo de esta investigación[184]. He aquí mi respuesta. La lógica epistémica no inauguró ninguna de las tesis acerca del conocimiento a las que se ha aludido, antes bien, las tomó prestadas de la Epistemología para su análisis. Ya he hecho mención a los paralelismos entre algunos elementos de esta lógica y el escepticismo; sin embargo, existe todavía otra analogía: como el escéptico —quien no afirma ni niega nada, sino que trabaja con la materia ajena y, luego de un examen, exhibe las consecuencias y los fallos—, la Lógica Epistémica es un recurso de la Epistemología para la detección de errores, algo así como su espejo. La lógica que hemos desarrollado conforma un ambiente más o menos controlado y, por qué no, artificial, en donde se pueden examinar ordenadamente las implicaciones inferenciales y consecuencias diversas de las tesis epistemológicas. En este sentido, la Lógica Epistémica es una extensión del *Órganon* que siempre ha sido la lógica para la Filosofía; y en la medida en que contribuye a la eliminación de errores mediante su detección, la Lógica Epistémica forma parte del sistema inmune de la Epistemología. Por tanto, las consecuencias adversas de nuestro análisis, si es que cabe tipificarlas de tal manera, lo son primordialmente para la teoría del conocimiento, son errores atribuibles a ésta, y nuestra lógica no ha hecho sino identificarlos[185]: el análisis lógico-epistémico nos sugiere fuertemente que la Epistemología, o por lo menos una parte de ella, se ha tornado una disciplina o actividad tal que termina por destruir su propio objeto de estudio, pues en sus entrañas se ha gestado un concepto de *saber* el cual no tiene una contraparte en las prácticas de conocimiento, de modo que semejante concepto resulta vacío. En suma, no sería la Lógica, sino la Epistemología la que terminaría deformando aquello que pretendía analizar[186].

---

[184] Hace poco más de un año, incluso alguien llegó a preguntarme: "¿qué de epistémico tiene la Lógica Epistémica?".

[185] Me refiero aquí a todas las consecuencias antes mencionadas, con excepción, quizá, al problema de la omnisciencia lógica.

[186] Podemos hallar una formulación de esta idea en *Elusive Knowledge*, de David Lewis (*cf.*, la primera nota al pie de ese trabajo).

Como nota final a este respecto, he de decir que todas las consideraciones aquí desplegadas no fueron en su mayoría producto del análisis *formal* de la Lógica Epistémica a través de sistemas axiomáticos y modelos de Kripke, sino el resultado del ambiente ordenado (o sistemático)[187] y de la oportunidad que brinda esta lógica para aislar algunos conceptos y considerarlos desde una perspectiva distinta a las tradicionales. Confieso que esto no favorece al formalismo en Lógica Epistémica; después de todo, uno podría preguntar si todo ese desarrollo formal arroja una claridad inalcanzable mediante el análisis conceptual y la reflexión filosófica en lenguaje natural. En cuanto a la oportunidad de aislar tesis o principios epistemológicos, seguramente tal manera de proceder presenta algunas desventajas, tal vez las misma que cualquier situación experimental —en la que se aíslan ciertos elementos para su análisis—, pero, a la vez, como todo experimento, reporta también algunas ventajas; espero, a lo largo de estas páginas, haber mostrado un puñado de ellas.

[187] Siempre he creído que una de las ventajas y atractivos principales de la Lógica consiste en el orden que impone a los pensamientos, orden que, como dice el clásico en sus Reglas para la dirección del espíritu, debe ser observado por todo el que emprende <<el conocimiento de las cosas no menos que el hilo de Theseo por quien ha de entrar en el laberinto>> [A.T. X, 380].

Apéndice I

En este apéndice, se incluye la demostración de algunos teoremas y reglas derivadas de inferencia de las teorías axiomáticas vistas a lo largo de los capítulos II, III y IV. Las demostraciones están agrupadas según el sistema axiomático al que pertenecen, *i. e.*, el sistema axiomático más elemental en el que se demuestra el teorema o regla de inferencia en cuestión. En cada caso, se ofrece primero una lista de todo aquello que se va a demostrar[188].

A1.1. Teoremas y reglas de inferencia en la teoría $\mathbb{K}$

Teoremas de $\mathbb{K}$

T1$\vdash^{K}$ $(K_j(\alpha\rightarrow\beta)\wedge K_j(\beta\rightarrow\alpha))\rightarrow(K_j\alpha\leftrightarrow K_j\beta)$

T2$\vdash^{K}$ $K_j(\alpha\wedge\beta)\leftrightarrow(K_j\alpha\wedge K_j\beta)$

T3$\vdash^{K}$ $(K_j(\alpha\rightarrow\beta)\wedge K_j(\beta\rightarrow\alpha))\leftrightarrow K_j(\alpha\leftrightarrow\beta)$

T4$\vdash^{K}$ $K_j\neg(\alpha\vee\beta)\leftrightarrow (K_j\neg\alpha\wedge K_j\neg\beta)$*

T5$\vdash^{K}$ $\neg K_j\neg(\alpha\vee\beta)\leftrightarrow(\neg K_j\neg\alpha\vee\neg K_j\neg\beta)$*

T6$\vdash^{K}$ $K_j(\alpha\rightarrow\beta)\rightarrow(\neg K_j\neg\alpha\rightarrow\neg K_j\neg\beta)$*

T7$\vdash^{K}$ $(K_j\alpha\vee K_j\beta)\rightarrow K_j(\alpha\vee\beta)$

T8$\vdash^{K}$ $\neg K_j\neg(\alpha\wedge\beta)\leftrightarrow(\neg K_j\neg\alpha\wedge\neg K_j\neg\beta)$*

T9$\vdash^{K}$ $K_j(\neg\alpha\rightarrow\alpha)\leftrightarrow K_j\alpha$

T10$\vdash^{K}$ $K_j(\alpha\rightarrow\neg\alpha)\leftrightarrow K_{j\neg}\alpha$

T11$\vdash^{K}$ $(K_j(\beta\rightarrow\alpha)\wedge K_j(\neg\beta\rightarrow\alpha))\leftrightarrow K_j\alpha$

T12$\vdash^{K}$ $(K_j(\alpha\rightarrow\beta)\wedge K_j(\alpha\rightarrow\neg\beta))\leftrightarrow K_j\neg\alpha$

T13$\vdash^{K}$ $K_j\alpha\rightarrow K_j(\beta\rightarrow\alpha)$

T14$\vdash^{K}$ $Kj\neg\alpha\rightarrow K_j(\alpha\rightarrow\beta)$

T15$\vdash^{K}$ $K_j\alpha\rightarrow(\neg K_j\neg\beta\rightarrow\neg K_j\neg(\alpha\wedge\beta))$

Reglas derivadas de inferencia para $\mathbb{K}$

Regla derivada 1 (RD1)

Regla derivada 3 (RD3)

[188] Se recomienda al lector tener presentes las explicaciones acerca de las demostraciones que se incluyen en la sección “El sistema $\mathbb{K}$” del capítulo II.

Si $\vdash^{K} \alpha\to\beta$, entonces $\vdash^{K} K_j\alpha\to K_j\beta$.	Si $\vdash^{K} \alpha\to\beta$, entonces $\vdash^{K} \neg K_j\neg\alpha\to\neg K_j\neg\beta$.
Regla derivada 2 (RD2)	Regla derivada 4 (RD4)
Si $\vdash^{K} \alpha\leftrightarrow\beta$, entonces $\vdash^{K} K_j\alpha\leftrightarrow K_j\beta$.	Si $\vdash^{K} \alpha$, entonces $\vdash^{K} \neg K_j\neg\beta\to\neg K_j\neg(\alpha\wedge\beta)$.

Veamos ahora la demostración de cada uno de estos teoremas y reglas derivadas de inferencia.

Teoremas de $\mathbb{K}$

T1 $\vdash^{K}$ $(K_j(\alpha\to\beta)\wedge K_j(\beta\to\alpha))\to(K_j\alpha\leftrightarrow K_j\beta)$

1. $K_j(\alpha\to\beta) \to(K_j\alpha\to K_j\beta)$	*K*
2. $K_j(\beta\to\alpha) \to(K_j\beta\to K_j\alpha)$	*K (α/β, β/α)*
3. $(K_j(\alpha\to\beta)\wedge K_j(\beta\to\alpha))\to((K_j\alpha\to K_j\beta)\wedge (K_j\beta\to K_j\alpha))$	*1, 2, Dilema constructivo y MP(x2)*
4. $(K_j(\alpha\to\beta)\wedge K_j(\beta\to\alpha))\to(K_j\alpha\leftrightarrow K_j\beta)$	*3, Def. ↔*

T2 $\vdash^{K}$ $K_j(\alpha\wedge\beta)\leftrightarrow(K_j\alpha\wedge K_j\beta)$

1. $(\alpha\wedge\beta)\to\alpha$	*Simplificación*
2. $K_j(\alpha\wedge\beta)\to K_j\alpha$	*1, RD1*
3. $(\alpha\wedge\beta)\to\beta$	*Simplificación*
4. $K_j(\alpha\wedge\beta)\to K_j\beta$	*3, RD1*
5. $K_j(\alpha\wedge\beta)\to(K_j\alpha\wedge K_j\beta)$	*2, 4, Implicación y MP(x2)*
6. $\alpha\to(\beta\to(\alpha\wedge\beta))$	*Adjunción*
7. $K_j\alpha\to K_j(\beta\to(\alpha\wedge\beta))$	*6, RD1*
8. $K_j(\beta\to(\alpha\wedge\beta))\to(K_j\beta\to K_j(\alpha\wedge\beta))$	*K (α/β, β/(α∧β))*
9. $K_j\alpha\to(K_j\beta\to K_j(\alpha\wedge\beta))$	*7, 8, Transitividad y MP(x2)*
10. $(K_j\alpha\wedge K_j\beta)\to K_j(\alpha\wedge\beta)$	*9, Importación/Exportación*
11. $K_j(\alpha\wedge\beta)\leftrightarrow(K_j\alpha\wedge K_j\beta)$	*5, 10, Equivalencia y MP(x2)*

T3 $\vdash^{K}$ $(K_j(\alpha\to\beta)\wedge K_j(\beta\to\alpha))\leftrightarrow K_j(\alpha\leftrightarrow\beta)$

1. $(K_j(\alpha\to\beta)\wedge K_j(\beta\to\alpha))\leftrightarrow K_j((\alpha\to\beta)\wedge(\beta\to\alpha))$	*T2 (α/(α→β), β/(β→α)), Tautología*
2. $(K_j(\alpha\to\beta)\wedge K_j(\beta\to\alpha))\leftrightarrow K_j(\alpha\leftrightarrow\beta)$	*1, Def. ↔*

T4 $\vdash^{K}$ $K_j\neg(\alpha\vee\beta)\leftrightarrow (K_j\neg\alpha\wedge K_j\neg\beta)$

1. $K_j(\neg\alpha \wedge \neg\beta) \leftrightarrow (K_j\neg\alpha \wedge K_j\neg\beta)$ — *T2 (α/¬α, β/¬β)*

2. $K_j\neg(\alpha \vee \beta) \leftrightarrow (K_j\neg\alpha \wedge K_j\neg\beta)$ — *1, De Morgan*

T5 $\vdash^K \neg K_j\neg(\alpha \vee \beta) \leftrightarrow (\neg K_j\neg\alpha \vee \neg K_j\neg\beta)$

1. $K_j\neg(\alpha \vee \beta) \leftrightarrow (K_j\neg\alpha \wedge K_j\neg\beta)$ — *T4*

2. $\neg K_j\neg(\alpha \vee \beta) \leftrightarrow \neg(K_j\neg\alpha \wedge K_j\neg\beta)$ — *1, Contraposición*

3. $\neg K_j\neg(\alpha \vee \beta) \leftrightarrow (\neg K_j\neg\alpha \vee \neg K_j\neg\beta)$ — *2, De Morgan*

T6 $\vdash^K K_j(\alpha \rightarrow \beta) \rightarrow (\neg K_j\neg\alpha \rightarrow \neg K_j\neg\beta)$

1. $K_j(\neg\beta \rightarrow \neg\alpha) \rightarrow (K_j\neg\beta \rightarrow K_j\neg\alpha)$ — *K (α/¬β, β/¬α)*

2. $K_j(\alpha \rightarrow \beta) \rightarrow (\neg K_j\neg\alpha \rightarrow \neg K_j\neg\beta)$ — *1, Contraposición*

T7 $\vdash^K (K_j\alpha \vee K_j\beta) \rightarrow K_j(\alpha \vee \beta)$

1. $\alpha \rightarrow (\alpha \vee \beta)$ — *Adición*

2. $K_j\alpha \rightarrow K_j(\alpha \vee \beta)$ — *1, RD1*

3. $\beta \rightarrow (\alpha \vee \beta)$ — *Adición*

4. $K_j\beta \rightarrow K_j(\alpha \vee \beta)$ — *3, RD1*

5. $(K_j\alpha \vee K_j\beta) \rightarrow K_j(\alpha \vee \beta)$ — *2, 4 Tautología*

T8 $\vdash^K \neg K_j\neg(\alpha \wedge \beta) \leftrightarrow (\neg K_j\neg\alpha \wedge \neg K_j\neg\beta)$

1. $(\alpha \wedge \beta) \rightarrow \alpha$ — *Simplificación*

2. $\neg K_j\neg(\alpha \wedge \beta) \rightarrow \neg K_j\neg\alpha$ — *1, RD3*

3. $(\alpha \wedge \beta) \rightarrow \beta$ — *Simplificación*

4. $\neg K_j\neg(\alpha \wedge \beta) \rightarrow \neg K_j\neg\beta$ — *3, RD3*

5. $\neg K_j\neg(\alpha \wedge \beta) \leftrightarrow (\neg K_j\neg\alpha \wedge \neg K_j\neg\beta)$ — *2, 4 Composición*

T9 $\vdash^K K_j(\neg\alpha \rightarrow \alpha) \leftrightarrow K_j\alpha$

1. $(\neg\alpha \rightarrow \alpha) \leftrightarrow \alpha$ — *Ley de Clavius*

2. $K_j(\neg\alpha \rightarrow \alpha) \leftrightarrow K_j\alpha$ — *1, RD1*

T10 $\vdash^K K_j(\alpha \rightarrow \neg\alpha) \leftrightarrow K_{j\neg}\alpha$

1. $(\alpha \rightarrow \neg\alpha) \leftrightarrow_{\neg}\alpha$ — *Ley de Clavius*

2. $K_j(\alpha \rightarrow \neg\alpha) \leftrightarrow K_{j\neg}\alpha$ — *1, RD1*

T11$\vdash^K$ $(K_j(\beta\rightarrow\alpha)\land K_j(\neg\beta\rightarrow\alpha))\leftrightarrow K_j\alpha$

1. $((\beta\rightarrow\alpha)\land(\neg\beta\rightarrow\alpha))\leftrightarrow\alpha$	*Reducción al absurdo*
2. $K_j((\beta\rightarrow\alpha)\land(\neg\beta\rightarrow\alpha))\leftrightarrow K_j\alpha$	*1, RD1*
3. $K_j((\beta\rightarrow\alpha)\land(\neg\beta\rightarrow\alpha))\leftrightarrow(K_j(\beta\rightarrow\alpha)\land K_j(\neg\beta\rightarrow\alpha))$	*T2 ($\alpha/\beta\rightarrow\alpha$, $\beta/\neg\beta\rightarrow\alpha$)*
4. $(K_j(\beta\rightarrow\alpha)\land K_j(\neg\beta\rightarrow\alpha))\leftrightarrow K_j\alpha$	*2, 3 Tautología*

T12$\vdash^K$ $(K_j(\alpha\rightarrow\beta)\land K_j(\alpha\rightarrow\neg\beta))\leftrightarrow K_j\neg\alpha$

1. $((\alpha\rightarrow\beta)\land(\alpha\rightarrow\neg\beta))\leftrightarrow\neg\alpha$	*Reducción al absurdo*
2. $K_j((\alpha\rightarrow\beta)\land(\alpha\rightarrow\neg\beta))\leftrightarrow K_j\neg\alpha$	*1, RD1*
3. $K_j((\alpha\rightarrow\beta)\land(\alpha\rightarrow\neg\beta))\leftrightarrow(K_j(\alpha\rightarrow\beta)\land K_j(\alpha\rightarrow\neg\beta))$	*T2 ($\alpha/\alpha\rightarrow\beta$, $\beta/\alpha\rightarrow\neg\beta$)*
4. $(K_j(\alpha\rightarrow\beta)\land K_j(\alpha\rightarrow\neg\beta))\leftrightarrow K_j\neg\alpha$	*2, 3 Tautología*

T13$\vdash^K$ $K_j\alpha\rightarrow K_j(\beta\rightarrow\alpha)$

1. $\alpha\rightarrow(\beta\rightarrow\alpha)$	*Condicionalización*
2. $K_j\alpha\rightarrow K_j(\beta\rightarrow\alpha)$	*1, RD1*

T14$\vdash^K$ $Kj\neg\alpha\rightarrow K_j(\alpha\rightarrow\beta)$ 164

1. $\neg\alpha\rightarrow(\alpha\rightarrow\beta)$	*Principio de explosión*
2. $K_j\neg\alpha\rightarrow K_j(\alpha\rightarrow\beta)$	*1, RD1*

T15$\vdash^K$ $K_j\alpha\rightarrow(\neg K_j\neg\beta\rightarrow\neg K_j\neg(\alpha\land\beta))$

1. $\alpha\rightarrow(\beta\rightarrow(\alpha\land\beta))$	*Conjunción*
2. $K_j\alpha\rightarrow K_j(\beta\rightarrow(\alpha\land\beta))$	*1, RD1*
3. $K_j(\beta\rightarrow(\alpha\land\beta))\rightarrow(\neg K_j\neg\beta\rightarrow\neg K_j\neg(\alpha\land\beta))$	*T6 (α/β, $\beta/\alpha\land\beta$)*
4. $K_j\alpha\rightarrow(\neg K_j\neg\beta\rightarrow\neg K_j\neg(\alpha\land\beta))$	*2, 3 Transitividad*

Reglas derivadas de inferencia para $\mathbb{K}$

Sean α y β *fbf´s* cualesquiera de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$ y sea $j=1, \ldots, n$.

Regla derivada 1 (RD1)

Si $\vdash^K \alpha\rightarrow\beta$, entonces $\vdash^K K_j\alpha\rightarrow K_j\beta$.

Supongamos que $\vdash^K \alpha\rightarrow\beta$. Entonces existe una demostración en $\mathbb{K}$ que termina en la fórmula '$\alpha\rightarrow\beta$'. Podemos entonces continuar la demostración como sigue y llegar a la fórmula '$K_j\alpha\rightarrow K_j\beta$':

----

m. $\alpha\rightarrow\beta$

m+1. $K_j(\alpha\rightarrow\beta)$ *m, KG*

m+2. $K_j(\alpha\rightarrow\beta)\rightarrow(K_j\alpha\rightarrow K_j\beta)$ *K*

m+3. $K_j\alpha\rightarrow K_j\beta$ *m+1, m+2 MP*

Regla derivada 2 (RD2)

Si $\vdash^K \alpha\leftrightarrow\beta$, entonces $\vdash^K K_j\alpha\leftrightarrow K_j\beta$.

Supongamos que $\vdash^K \alpha\leftrightarrow\beta$. Entonces existe una demostración en $\mathbb{K}$ que termina en la fórmula '$\alpha\leftrightarrow\beta$'. Podemos entonces continuar la demostración como sigue y llegar a la fórmula '$K_j\alpha\leftrightarrow K_j\beta$':

----

m. $\alpha\leftrightarrow\beta$

m+1. $\alpha\rightarrow\beta$ *m, Tautología*

m+2. $K_j\alpha\rightarrow K_j\beta$ *m+1, RD1*

m+3. $\beta\rightarrow\alpha$ *m, Tautología*

m+4. $K_j\beta\rightarrow K_j\alpha$ *m+3, RD1*

m+5. $K_j\alpha\rightarrow K_j\beta\wedge K_j\beta\rightarrow K_j\alpha$ *m+2, m+4 Conjunción*

m+6. $K_j\alpha\leftrightarrow K_j\beta$ *m+5, Def. $\leftrightarrow$*

Regla derivada 3 (RD3)

Si $\vdash^K \alpha\rightarrow\beta$, entonces $\vdash^K \neg K_j\neg\alpha\rightarrow\neg K_j\neg\beta$.

Supongamos que $\vdash^K \alpha\rightarrow\beta$. Entonces existe una demostración en $\mathbb{K}$ que termina en la fórmula '$\alpha\rightarrow\beta$'. Podemos entonces continuar la demostración como sigue y llegar a la fórmula '$\neg K_j\neg\alpha\leftrightarrow\neg K_j\neg\beta$':

----

m. $\alpha\rightarrow\beta$

m+1. $K_j(\alpha\rightarrow\beta)$ *m, KG*

m+2. $K_j(\alpha\rightarrow\beta)\rightarrow(\neg K_j\neg\alpha\rightarrow\neg K_j\neg\beta)$ *T6*

m+3. $\neg K_j\neg\alpha\rightarrow\neg K_j\neg\beta$ *m+1, m+2 MP*

Regla derivada 4 (RD4)

Si $\vdash^K \alpha$, entonces $\vdash^K \neg K_j\neg\beta\rightarrow\neg K_j\neg(\alpha\wedge\beta)$.

Supongamos que $\vdash^K \alpha$. Entonces existe una demostración en $\mathbb{K}$ que termina en la fórmula ‘α’. Podemos entonces continuar la demostración como sigue y llegar a la fórmula ‘$\neg K_j\neg\beta\rightarrow\neg K_j\neg(\alpha\wedge\beta)$’:

----

m. α

m+1. $K_j\alpha$ *m, KG*

m+2. $K_j\alpha\rightarrow(\neg K_j\neg\beta\rightarrow\neg K_j\neg(\alpha\wedge\beta))$ *T15*

m+3. $\neg K_j\neg\beta\rightarrow\neg K_j\neg(\alpha\wedge\beta)$ *m+1, m+2 MP*

A1.2. Teoremas en la teoría $\mathbb{T}$

T16$\vdash^T \alpha\rightarrow\neg K_j\neg\alpha$

T17$\vdash^T K_j\alpha\rightarrow\neg K_j\neg\alpha$

T18$\vdash^T (K_j\alpha\wedge K_j\beta)\rightarrow\neg(\alpha\rightarrow\neg\beta)$[189]

Demostraciones:

T16$\vdash^T \alpha\rightarrow\neg K_j\neg\alpha$

1. $K_j\neg\alpha\rightarrow\neg\alpha$ *T (α/¬α)*
2. $\alpha\rightarrow\neg K_j\neg\alpha$ *1, Transposición*

T17$\vdash^T K_j\alpha\rightarrow\neg K_j\neg\alpha$

1. $K_j\alpha\rightarrow\alpha$ *T*
2. $\alpha\rightarrow\neg K_j\neg\alpha$ *T16*
3. $K_j\alpha\rightarrow\neg K_j\neg\alpha$ *1, 2 Transitividad*

---

[189] Para la demostración de estos teoremas, consúltese el apéndice I.

T18$\vdash^{T}$ $(K_j\alpha \wedge K_j\beta) \rightarrow \neg(\alpha \rightarrow \neg\beta)$

1. $K_j\alpha \rightarrow \alpha$	*T*
2. $K_j\beta \rightarrow \beta$	*T (α/β)*
3. $(K_j\alpha \wedge K_j\beta) \rightarrow (\alpha \wedge \beta)$	*1, 2 Tautología*
4. $(\alpha \wedge \beta) \rightarrow \neg(\alpha \rightarrow \neg\beta)$	*Implicación material*
5. $(K_j\alpha \wedge K_j\beta) \rightarrow \neg(\alpha \rightarrow \neg\beta)$	*3, 4 Transitividad*

A1.3. Teoremas y reglas de inferencia en la teoría S4

Teorema de S4

T19: $K_j\alpha \leftrightarrow K_jK_j\alpha$

Demostración

T19: $K_j\alpha \leftrightarrow K_jK_j\alpha$

1. $K_j\alpha \rightarrow K_jK_j\alpha$	*A4*
2. $K_jK_j\alpha \rightarrow K_j\alpha$	*T (α/ $K_j\alpha$)*
3. $(K_j\alpha \rightarrow K_jK_j\alpha) \wedge (K_jK_j\alpha \rightarrow K_j\alpha)$	*1, 2 Conjunción*
4. $K_j\alpha \leftrightarrow K_jK_j\alpha$	*3, Def. ↔*

Regla de inferencia en S4

KK: Si $\vdash^{S4}$ $K_j\alpha \rightarrow \beta$, entonces $\vdash^{S4}$ $K_j\alpha \rightarrow K_j\beta$.

Demostración

Supongamos que $\vdash^{S4}$ $K_j\alpha \rightarrow \beta$. Entonces existe una demostración en S4 que termina en la fórmula '$K_j\alpha \rightarrow \beta$'. Podemos entonces continuar la demostración como sigue y llegar a la fórmula '$K_j\alpha \rightarrow K_j\beta$':

----

m. $K_j\alpha \rightarrow \beta$	
m+1. $K_jK_j\alpha \rightarrow K_j\beta$	*m, RD1*
m+2. $K_j\alpha \rightarrow K_jK_j\alpha$	*A4*
m+3. $K_j\alpha \rightarrow K_j\beta$	*m+2, m+1 Transitividad*

A1.4. Teoremas en la teoría S5

T20$\vdash^{S5}$ $\neg K_j\neg K_j\alpha\to K_j\alpha$

T21$\vdash^{S5}$ $K_j\alpha\leftrightarrow\neg K_j\neg K_j\alpha$

T22$\vdash^{S5}$ $\neg K_j\alpha\leftrightarrow K_j\neg K_j\alpha$

T23$\vdash^{S5}$ $\alpha\to K_j\neg K_j\neg\alpha$ [Propiedad brouweriana]

T24$\vdash^{S5}$ $K_j\alpha\to K_jK_j\alpha$ [Tesis KK]

Demostraciones

T20$\vdash^{S5}$ $\neg K_j\neg K_j\alpha\to K_j\alpha$

1. $\neg K_j\alpha\to K_j\neg K_j\alpha$ *A5*
2. $\neg K_j\neg K_j\alpha\to K_j\alpha$ *1, Transposición*

T21$\vdash^{S5}$ $K_j\alpha\leftrightarrow\neg K_j\neg K_j\alpha$

1. $\neg K_j\neg K_j\alpha\to K_j\alpha$ *T20*
2. $K_j\alpha\to\neg K_j\neg K_j\alpha$ *T16 (α/ $K_j\alpha$)*
3. $(K_j\alpha\to\neg K_j\neg K_j\alpha)\wedge(\neg K_j\neg K_j\alpha\to K_j\alpha)$ *1, 2 Conjunción*
4. $K_j\alpha\leftrightarrow\neg K_j\neg K_j\alpha$ *3, Def. $\leftrightarrow$*

T22$\vdash^{S5}$ $\neg K_j\alpha\leftrightarrow K_j\neg K_j\alpha$

1. $K_j\neg K_j\alpha\to\neg K_j\alpha$ *T (α/$\neg K_j\alpha$)*
2. $\neg K_j\alpha\to K_j\neg K_j\alpha$ *A5*
3. $(\neg K_j\alpha\to K_j\neg K_j\alpha)\wedge(K_j\neg K_j\alpha\to\neg K_j\alpha)$ *1, 2 Conjunción*
4. $\neg K_j\alpha\leftrightarrow K_j\neg K_j\alpha$ *3, Def. $\leftrightarrow$*

T23$\vdash^{S5}$ $\alpha\to K_j\neg K_j\neg\alpha$

1. $\neg K_j\neg\alpha\to K_j\neg K_j\neg\alpha$ *A5 (α/$\neg\alpha$)*
2. $\alpha\to\neg K_j\neg\alpha$ *T16*
3. $\alpha\to K_j\neg K_j\neg\alpha$ *1, 2 Transposición*

T24$\vdash^{S5}$ $K_j\alpha\to K_jK_j\alpha$

1. $\neg K_j\neg K_j\alpha\to K_j\alpha$ *Teorema 20*
2. $K_j\neg K_j\neg K_j\alpha\to K_jK_j\alpha$ *1, RD1*

3. $K_j\alpha \rightarrow K_j\neg K_j\neg K_j\alpha$ *Teorema 23 (α/ $K_j\alpha$)*

4. $K_j\alpha \rightarrow K_jK_j\alpha$ *2, 3, SH*

Apéndice II

Probaremos ahora el metateorema de completez para cuatro de las teorías axiomáticas que hemos estudiado, *viz.*, las teorías K, T, S4 y S5. Para ello, construiremos un *modelo canónico* para cada uno de estos sistemas, el cual consiste en una estructura o interpretación de la teoría axiomática en cuestión, ya que hace verdaderos a todos los teoremas de la teoría, y sólo a ellos. Sin embargo, introduciremos primero la noción de lo que es un *conjunto maximal S-consistente de fórmulas*, y demostraremos algunos lemas referentes a éste.

A2.1. Conjuntos maximales S-consistentes

Sea S un sistema epistémico normal (*i. e.*, que incluye al sistema K)[190] arbitrario, y sea Γ un conjunto de fbf's de L-$\{K_1, \ldots, K_n\}$ cualesquiera.

Definición de S-consistencia[191]:

Γ es S-consistente sii no existen $\gamma_1, \ldots, \gamma_n \in \Gamma$ tales que $\vdash^S \neg(\gamma_1 \wedge \ldots \wedge \gamma_n)$.

[Por consiguiente, Γ es S-inconsistente sii existen $\gamma_1, \ldots, \gamma_n \in \Gamma$ tales que $\vdash^S \neg(\gamma_1 \wedge \ldots \wedge \gamma_n)$].

Definición de conjunto maximal:

---

[190] Entonces, un sistema epistémico normal contiene, por lo menos, a los axiomas A1-A3, el axioma K, las reglas de inferencia MP y KG, así como todos los resultados que obtuvimos en la teoría K a partir de éstos (*i. e.*, las reglas derivadas de inferencia de RD1 a RD4, y los 15 teoremas del sistema K).

[191] Desde luego, esta noción está estrechamente ligada con la de consistencia, según la cual, una teoría axiomática S en $\mathcal{L}$-$\{K_1, \ldots, K_n\}$ es consistente sii existe una fórmula α de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$ tal que α no se demuestra en S ($\nvdash^S \alpha$).

Γ es maximal sii para toda fórmula α de L-$\{K_1, \ldots, K_n\}$, $\alpha \in \Gamma$ o $\neg\alpha \in \Gamma$.

Supongamos ahora que Γ es un conjunto maximal y S-consistente (*i. e.*, que es maximal S-consistente). Veamos las propiedades que posee.

Lema 1:

1a. Para toda fórmula α, exactamente una (nada más una) de entre α y $\neg\alpha$ pertenece a Γ;

Para cualesquiera α y β fbf's:

b. $\alpha \vee \beta \in \Gamma$ sii $\alpha \in \Gamma$ o $\beta \in \Gamma$;

1c. $\alpha \wedge \beta \in \Gamma$ sii $\alpha \in \Gamma$ y $\beta \in \Gamma$

1d. si $\alpha \in \Gamma$ y $\alpha \rightarrow \beta \in \Gamma$, entonces $\beta \in \Gamma$[192].

Demostremos cada una de estas afirmaciones.

1a.

Por definición de maximal, por lo menos una de entre α y $\neg\alpha$ pertenece a Γ. Para demostrar que a los más una de entre α y $\neg\alpha$ pertenece a Γ, supongamos que $\alpha, \neg\alpha \in \Gamma$. Pero, $\vdash^{S} \neg(\alpha \wedge \neg\alpha)$, porque es instancia de tautología[193].

$\therefore$

Γ sería S-inconsistente! (Por hipótesis Γ es S-consistente).

$\therefore$

Para toda fórmula α, exactamente una de entre α y $\neg\alpha$ pertenece a Γ.

Quod erat demonstrandum

1b.

Supongamos que $\alpha \vee \beta \in \Gamma$, pero $\alpha \notin \Gamma$ y $\beta \notin \Gamma$. Por consiguiente, como Γ es maximal, $\neg\alpha$, $\neg\beta \in \Gamma$. Pero $\vdash^{S} \neg((\alpha \vee \beta) \wedge (\neg\alpha \wedge \neg\beta))$!

---

[192] 1d nos dice que Γ está cerrado bajo consecuencia lógica, mientras que 1a-1d, tomados en conjunto, nos dejan saber que Γ se comporta como un mundo.

[193] Recuérdese que el sistema $\mathbb{K}$ y, por consiguiente, cualquier sistema epistémico normal, al contener los axiomas A1-A3, es suficientemente fuerte para rescatar cualquier instancia de tautología.

∴

I) Si $\alpha\vee\beta\in\Gamma$, entonces $\alpha\in\Gamma$ o $\beta\in\Gamma$.

Supongamos ahora que $\alpha\in\Gamma$ o $\beta\in\Gamma$. Supongamos (sin perder generalidad) que $\alpha\in\Gamma$, pero que $\alpha\vee\beta\notin\Gamma$. Entonces $\neg(\alpha\vee\beta)\in\Gamma$. ¡Pero $\vdash^{S}\neg(\alpha\wedge\neg(\alpha\vee\beta))$!

∴

i) Si $\alpha\in\Gamma$, entonces $\alpha\vee\beta\in\Gamma$.

[Seguiríamos un procedimiento exactamente análogo al anterior si, en lugar de suponer que $\alpha\in\Gamma$, supusiéramos $\beta\in\Gamma$.]

∴

ii) Si $\beta\in\Gamma$, entonces $\alpha\vee\beta\in\Gamma$.

∴

II) Si $\alpha\in\Gamma$ o $\beta\in\Gamma$, entonces $\alpha\vee\beta\in\Gamma$ [de i y ii)].

De I) y II), se sigue que **$\alpha\vee\beta\in\Gamma$ sii $\alpha\in\Gamma$ o $\beta\in\Gamma$.**

Quod erat demonstrandum

1c.

Supongamos que $\alpha\in\Gamma$, $\beta\in\Gamma$, pero $\alpha\wedge\beta\notin\Gamma$. Entonces, $\neg(\alpha\wedge\beta)\in\Gamma$. Sin embargo, $\vdash^{S}\neg((\alpha\wedge\beta)\wedge\neg(\alpha\wedge\beta))$!, pues es tautología.

∴

I) Si $\alpha\in\Gamma$ y $\beta\in\Gamma$, entonces $\alpha\wedge\beta\in\Gamma$.

Supongamos ahora que $\alpha\wedge\beta\in\Gamma$, pero $\alpha\notin\Gamma$ o $\beta\notin\Gamma$. Entonces $\neg\alpha\in$, o bien, $\neg\beta\in\Gamma$. Supongamos que $\neg\alpha\in\Gamma$. No obstante, $\vdash^{S}\neg(\neg\alpha\wedge(\alpha\wedge\beta))$!

∴

i) Si $\alpha\wedge\beta\in\Gamma$, entonces $\alpha\in\Gamma$.

El caso para $\neg\beta$ es análogo.

∴

ii) Si $\alpha\wedge\beta\in\Gamma$, entonces $\beta\in\Gamma$.

∴

II) Si $\alpha\wedge\beta\in\Gamma$, entonces $\alpha\in\Gamma$ y $\beta\in\Gamma$ [de i) y ii)].

De I) y II), se sigue que **$\alpha\wedge\beta\in\Gamma$ sii $\alpha\in\Gamma$ y $\beta\in\Gamma$.**

Quod erat demonstrandum

1d.

Supongamos que $\alpha\in\Gamma$ y $\alpha\rightarrow\beta\in\Gamma$, pero $\beta\notin\Gamma$. Entonces $\neg\beta\in\Gamma$. Sin embargo, $\vdash^{S}\neg(\alpha\wedge\neg(\alpha\rightarrow\beta)\wedge\neg\beta)$!

∴

Si $\alpha\in\Gamma$ y $\alpha\rightarrow\beta\in\Gamma$, entonces $\beta\in\Gamma$.

Quod erat demonstrandum

Lema 2:

2a. Si $\vdash^{S}\alpha$, entonces $\alpha\in\Gamma$;

2b. Si $\alpha\in\Gamma$ y $\vdash^{S}\alpha\rightarrow\beta$, entonces $\beta\in\Gamma$.

Demostración:

2a.

Supongamos que $\vdash^{S}\alpha$ y $\alpha\notin\Gamma$. De 1a, se sigue que $\neg\alpha\in\Gamma$; pero como $\vdash^{S}\alpha$, $\vdash^{S}\neg\neg\alpha$ también [pues $\alpha\leftrightarrow\neg\neg\alpha$ es una instancia de tautología]! Pero esto último se contradice con la hipótesis de que Γ es S-consistente.

∴

Si $\vdash^{S}\alpha$, entonces $\alpha\in\Gamma$.

Quod erat demonstrandum

2b.

Supongamos que $\alpha\in\Gamma$ y $\vdash^{S}\alpha\rightarrow\beta$. De esto último y 2a, se tiene que $\alpha\rightarrow\beta\in\Gamma$; y de esto, de que $\alpha\in\Gamma$ y 1d, se sigue que $\beta\in\Gamma$.

∴

Si $\alpha\in\Gamma$ y $\vdash^{S}\alpha\rightarrow\beta$, entonces $\beta\in\Gamma$.

Quod erat demonstrandum

Lema 3[194]**.**

Sea Σ un conjunto S-consistente de fórmulas de L-$\{K_1, \ldots, K_n\}$. Entonces existe un conjunto maximal S-consistente Γ tal que $\Sigma \subseteq \Gamma$.

Demostración

Supongamos que tenemos a todas las fórmulas de L-$\{K_1, \ldots, K_n\}$ ordenadas en una lista numerable: $\alpha_1, \alpha_2, \ldots, \alpha_n, \ldots$ [tenemos un orden ya fijo][195].

Definimos por recursión un conjunto Γ_i para cada número natural i.

$$\Gamma_0=\Sigma$$

$$\Gamma_{n+1}=\begin{cases} \Gamma_n \cup \{\alpha_{n+1}\} & \text{si este conjunto es S-consistente.} \\ \Gamma_n \cup \{\neg\alpha_{n+1}\} & \text{en caso contrario.} \end{cases}$$

Así, por ejemplo:

$\Gamma_0=\Sigma$; $\Gamma_1=\Sigma\cup\{\alpha_1\}$; $\Gamma_2=\Gamma_1\cup\{\alpha_2\}$; etcétera[196].

Afirmación:

Para toda n, Γ_n es S-consistente.

Demostración por inducción matemática

Base de la inducción: n=0

Γ_0 es S-consistente porque $\Gamma_0=\Sigma$ y Σ es S-consistente por hipótesis.

Paso inductivo

[194] Este lema es conocido como *lema de Lindenbaum.*

[195] Podemos ordenar a todos los símbolos de $\mathcal{L}$-$\{K_1, \ldots, K_n\}$ y luego ordenar a todas sus fórmulas como en un diccionario

[196] Podría suceder, por mencionar un ejemplo, que α_1 ya estuviera en Σ, o bien, que en vez de agregar α_1, debamos agregar su negación ($\neg\alpha_1$)

Hipótesis de inducción (HI): $n=m$, con m un número natural arbitrario.

Supongamos que Γ_m es S-consistente.

P.D. Γ_{m+1} también es S-consistente.

Supongamos que Γ_{m+1} es S-inconsistente. Entonces tanto $\Gamma_m \cup \{\alpha_{m+1}\}$ como $\Gamma_m \cup \{\neg\alpha_{m+1}\}$ son S-inconsistente. Por definición, esto quiere decir que existen fórmulas $\beta_1, \ldots, \beta_i$ y $\gamma_1, \ldots, \gamma_j$ en Γ_m tales que:

$\vdash^S \neg (\beta_1 \wedge \ldots \wedge \beta_i \wedge \alpha_{m+1})$ y $\vdash^S \neg (\gamma_1 \wedge \ldots \wedge \gamma_j \wedge \neg\alpha_{m+1})$

$\therefore$

$\vdash^S (\beta_1 \wedge \ldots \wedge \beta_i) \rightarrow \alpha_{m+1}$ y $\vdash^S (\gamma_1 \wedge \ldots \wedge \gamma_j) \rightarrow \neg\alpha_{m+1}$

$\therefore$

$\vdash^S (\beta_1 \wedge \ldots \wedge \beta_i \wedge \gamma_1 \wedge \ldots \wedge \gamma_j) \rightarrow (\alpha_{m+1} \wedge \neg\alpha_{m+1})$

$\therefore$

$\vdash^S \neg (\beta_1 \wedge \ldots \wedge \beta_i \wedge \gamma_1 \wedge \ldots \wedge \gamma_j)$

$\therefore$

Γ_m sería S-inconsistente! Pero esto es contradice con HI.

Por lo tanto, **Para toda n, Γ_n es S-consistente.**

Quod erat demonstrandum

Ahora bien, sea $\Gamma = \bigcup_{n \in N} \Gamma_n$[197].

Afirmación:

1) $\Sigma \subseteq \Gamma$;
2) Γ es S-consistente;
3) Γ es maximal.

Demostración

[197] Es decir, $\Gamma = \Gamma_0 \cup \Gamma_1 \cup \Gamma_2 \cup \ldots$ Nótese que Γ posee infinitas fórmulas, y que para cada fórmula γ, $\gamma \in \Gamma$ sii $\gamma \in \Gamma_n$ para alguna n.

1) pasa porque $\Sigma=\Gamma_0 \therefore \boldsymbol{\Sigma\subseteq\Gamma}$

Por otro lado, si Γ fuera S-inconsistente, existirían $\gamma_1, \ldots \gamma_m \in \Gamma$ tales que $\vdash^S \neg(\gamma_1 \wedge \ldots \wedge \gamma_m)$. Como este conjunto de fórmulas es **finito**, existe $n \in \mathbf{N}$ tal que $\gamma_1, \ldots \gamma_m \in \Gamma_n$! Pero esto se contradice con lo que recién establecimos, a saber: para toda n, Γ_n es S-consistente.

$\therefore$

Γ es S-consistente

Por último, probemos que Γ es maximal.

Sea α cualquier fórmula.

P.D. $\alpha \in \Gamma$ o $\neg\alpha \in \Gamma$.

α es α_{n+1} para alguna n, puesto que en la lista $\alpha_1, \alpha_2, \ldots, \alpha_n, \ldots$ las ordenamos a todas: por definición, $\alpha_{n+1} \in \Gamma_{n+1}$ o $\neg\alpha_{n+1} \in \Gamma_{n+1}$.

$\therefore$

$\alpha \in \Gamma$ o $\neg\alpha \in \Gamma$ (pues $\Gamma_{n+1} \subseteq \Gamma$).

$\therefore$

Γ es maximal.

Quod erat demonstrandum

Lema 4:

Definición: $K_j\text{-}(\Gamma) \stackrel{\text{def}}{=} \{\beta: K_j\beta \in \Gamma\}$, con j=1, ..., n.

Sea Γ un conjunto maximal S-consistente y supongamos que α es una fórmula tal que $\neg K_j\alpha \in \Gamma$ (*i. e.*, dado 1a, $K_j\alpha \notin \Gamma$). Entonces el conjunto $K_j\text{-}(\Gamma) \cup \{\neg\alpha\}$ es S-consistente.

Demostración

Supongamos que $K_j\text{-}(\Gamma) \cup \{\neg\alpha\}$ es S-inconsistente. Entonces existen $\beta_1, \ldots, \beta_n \in K_j\text{-}(\Gamma)$ tales que:

$\vdash^S \neg(\beta_1 \wedge \ldots \wedge \beta_n \wedge \neg\alpha)$

$\therefore$

$\vdash^S (\beta_1 \wedge \ldots \wedge \beta_n) \rightarrow \alpha$ *Tautología*

∴

$\vdash^S K_j(\beta_1\wedge\ldots\wedge\beta_n)\rightarrow K_j\alpha$ *RD1*

∴

$\vdash^S (K_j\beta_1\wedge\ldots\wedge K_j\beta_n)\rightarrow K_j\alpha$ *Teorema 2 (de K)*

∴

$\vdash^S\neg(K_j\beta_1\wedge\ldots\wedge K_j\beta_n\wedge\neg K_j\alpha)$ *Tautología*

Pero $\beta_1, \ldots, \beta_n\in\Gamma$ y $\neg K_j\alpha\in\Gamma$.

∴

Γ sería S-inconsistente! (Sin embargo, esto se contradice con nuestra hipótesis inicial)

∴

K_j-(Γ)∪{¬α} es S-consistente[198].

Quod erat demonstrandum

A2.2. Construcción del modelo canónico de S

El modelo canónico de S es $E_S= <W_S, \pi_S, R_{1S}, \ldots, R_{nS}>$, donde:

1) W_S={w: w es conjunto maximal S-consistente} [W_S es el conjunto de todos los conjuntos maximales S-consistentes de fórmulas].

2) $<w, w'>\in R_{jS}$ sii K_j-(w)⊆w', con j=1, …, n [$<w, w'>\in R_{jS}$ sii para toda fórmula α, si $K_j\alpha\in$w, entonces α∈w'].

3) Sea P una letra proposicional y sea $w\in W_S$.

$\pi_K(P)=\{w\in W_S: P\in w\}$ [P es verdadera sii P∈w].

---

[198] Nótese en que podría no haber fórmulas de la forma $K_j\beta$ en Γ (*i. e.*, K_j-(Γ)=∅). Entonces tendríamos que probar que {¬α} es S-consistente. Veamos, si no lo fuera, $\vdash^S\neg\neg\alpha \therefore \vdash^S\alpha \therefore \vdash^S K_j\alpha$ (por regla KG). Pero por lema 2a, tendríamos que $K_j\alpha\in\Gamma$**!** (No obstante, esto se contradice con nuestra hipótesis inicial) De modo que {¬α} sería S-consistente.

TEOREMA FUNDAMENTAL DEL MODELO CANÓNICO:

Para toda fórmula α y todo $w \in W_S$, E_S, $w \vDash \alpha$ sii $\alpha \in w$ (*i. e.*, verdad y pertenencia coinciden).

Demostraremos la afirmación por inducción sobre la complejidad de α.

Base de la inducción: α es una letra proposicional *p*.

La afirmación vale porque así definimos a E_S ∴ $E_S, w \vDash p$ sii $p \in w$.

Hipótesis de inducción (HI):

Supongamos la afirmación válida para toda fórmula con menos de m conectivos y operadores (con m un número natural fijo).

Sea α con m conectivos y operadores en conjunto.

P.D. $E_S, w \vDash \alpha$ sii $\alpha \in w$.

Caso 1: $\alpha = \neg\beta$

Dada nuestra definición de satisfacibilidad, $E_S, w \vDash \neg\beta$ sii $E_S, w \nvDash \beta$. Pero por HI, $E_S, w \nvDash \beta$ sii $\beta \notin w$. Como w es maximal, tenemos —por lema 1a— $\beta \notin w$ sii $\neg\beta \in w$ ∴ $\mathbf{E_S, w \vDash \neg\beta}$ **sii** $\mathbf{\neg\beta \in w}$.

Caso 2: $\alpha = \beta \rightarrow \gamma$

Por nuestra definición de satisfacibilidad, tenemos que $E_S, w \vDash \beta \rightarrow \gamma$ sii $E_S, w \nvDash \beta$ o $E_S, w \vDash \gamma$. Por HI, se tiene que $E_S, w \nvDash \beta$ o $E_S, w \vDash \gamma$ sii $\beta \notin w$ o $\gamma \in w$. De lema 1a, se sigue que $\beta \notin w$ o $\gamma \in w$ sii $\neg\beta \in w$ o $\gamma \in w$. Ahora, de lema 1b, tenemos que $\neg\beta \in w$ o $\gamma \in w$ sii $\neg\beta \vee \gamma \in w$. Pero, por lema 2b, se tiene que $\neg\beta \vee \gamma \in w$ sii $\beta \rightarrow \gamma \in w$ ∴ $\mathbf{E_S, w \vDash \beta \rightarrow \gamma}$ **sii** $\mathbf{\beta \rightarrow \gamma \in w}$.

Caso 3: $\alpha = K_j\beta$

P. D. 1) Si $E_S, w \vDash K_j\beta$, entonces $K_j\beta \in w$, y 2) Si $K_j\beta \in w$, entonces $E_S, w \vDash K_j\beta$.

Probemos el primero de estos condicionales (por transposición).

Supongamos que $K_j\beta \notin w$. De esto y lema 1a, se sigue que $\neg K_j\beta \in w$. Por lema 4, K_j-$(w) \cup \{\neg\beta\}$ es S-consistente. Por lema 3, existe un conjunto maximal S-consistente, w', tal que K_j-$(w) \cup \{\neg\beta\} \subseteq w'$ (este w' está en W_S de E_S) $\therefore$ K_j-$(w) \subseteq w'$ y $\neg\beta \in w'$ $\therefore$ Por definición de R_{jS} y lema 1a, respectivamente, $\langle w, w' \rangle \in R_{jS}$ y $\beta \notin w'$ $\therefore$ Por HI, $E_S, w' \nvDash \beta$ $\therefore$ Por nuestra definición de satisfacibilidad, $E_S, w \nvDash K_j\beta$.

$\therefore$

1) Si $E_S, w \vDash K_j\beta$, entonces $K_j\beta \in w$.

Probemos ahora el segundo condicional (también por transposición).

Supongamos que $E_S, w \nvDash K_j\beta$. Por definición de satisfacibilidad, existe $w' \in W_S$ tal que $\langle w, w' \rangle \in R_{jS}$ y $E_S, w' \nvDash \beta$ $\therefore$ Por HI, $\beta \notin w'$. Como $\langle w, w' \rangle \in R_{jS}$, tenemos, por definición de R_{jS}, K_j-$(w) \subseteq w'$ $\therefore$ $K_j\beta \notin w$.

$\therefore$

2) Si $K_j\beta \in w$, entonces $E_S, w \vDash K_j\beta$.

$\therefore$

De 1) y 2), **$E_S, w \vDash K_j\beta$ sii $K_j\beta \in w$.**

Con esto concluye la demostración del Teorema Fundamental.

Quod erat demonstrandum

Veamos ahora un corolario de este último lema.

Corolario:

Sea S un sistema epistémico normal en L-$\{K_1, \ldots, K_n\}$ y sea E_S su modelo canónico. Entonces, para toda fórmula φ, $\vdash^S \varphi$ sii $E_S \vDash \varphi$.

Demostración

Supongamos primero que $\vdash^S \varphi$. Por lema 2a, $\varphi \in w$ para todo $w \in W_S$. Por teorema fundamental, para todo $w \in W_S$, $E_S, w \vDash \varphi$ $\therefore E_S \vDash \varphi$.

Supongamos ahora que $\nvdash^{S}\varphi$. Entonces $\{\neg\varphi\}$ es S-consistente. Por tanto, existe un conjunto maximal S-consistente que contiene a $\{\neg\varphi\}$. Como los elementos de W_S son los conjuntos maximales S-consistentes, concluimos que $\neg\varphi\in w$, para algún $w\in W_S$. Por el teorema fundamental, $E_S, w\vDash\neg\varphi$. Por tanto, $E_S, w\nvDash\varphi \therefore E_S\nvDash\varphi$.

$\therefore$

Para toda fórmula φ, $\vdash^{S}\varphi$ sii $E_S\vDash\varphi$.

Quod erat demonstrandum

Si empleamos el corolario anterior, es relativamente sencillo demostrar el metateorema de completez para cada uno de nuestros sistemas axiomáticos. Veamos.

A2.3. Metateorema de completez para K, T, S4 y S5

Sea S un sistema epistémico normal arbitrario, E_S su modelo canónico, y sea **C** cualquier clase de interpretaciones o estructuras de Kripke que contiene a E_S.[199].

Afirmamos que si φ es cualquier fórmula verdadera en cualquier interpretación de C, $E_S\vDash\varphi$. Por el teorema fundamental, $\vdash^{S}\varphi$.

Ahora sí estamos listos para enunciar y demostrar el metateorema de completez para los sistemas K, T, S4 y S5.

Metateorema de completez para K:

Toda fórmula válida es teorema de K. Es decir, si $\vDash\varphi$, entonces $\vdash^{K}\varphi$.

Esto se sigue por el evidente hecho de que el modelo canónico de K pertenece a la clase de todas las interpretaciones, y todo teorema de K es verdadero en todas las interpretaciones.

Quod erat demonstrandum

Para probar los correspondientes metateoremas de completez para T, S4 y S5, basta probar, respectivamente:

---

[199] **C** puede ser, por ejemplo, la clase de todas las interpretaciones (*i. e.*, la clase de las estructuras así sin más, en las que no existe restricción alguna para las relaciones de accesibilidad R_j).

1) En E_T (el modelo canónico del sistema T), para cada $j=1, \ldots, n$, R_{jT} es reflexiva

2) En E_{S4} (el modelo canónico del sistema S4), para cada $j=1, \ldots, n$, R_{jS4} es reflexiva y transitiva.

3) En E_{S5} (el modelo canónico del sistema S5), para cada $j=1, \ldots, n$, R_{jS5} es reflexiva y euclidiana.

Teorema 1:

Sea $E_T= <W_T, \pi_T, R_{1T}, \ldots, R_{nT}>$ el modelo canónico de T. Entonces, para cada $j=1, \ldots, n$, R_{jT} es reflexiva.

Demostración

Sea $w \in W_T$. P.D. $<w,w> \in R_{jT}$ [esto significa, por construcción del modelo canónico, K_j-$(w) \subseteq w$]. Supongamos $K_j\alpha \in w$. P.D. $\alpha \in w$. Si $K_j\alpha \in w$, como $\vdash^T K_j\alpha \rightarrow \alpha$, tenemos —por lema 2b— $\alpha \in w$.

$\therefore$

$<w,w> \in R_{jT}$ [R_{jT} es reflexiva]

Mmetateorema de completez para T:

Toda fórmula T-válida es teorema de T. Esto es, si $\vDash^T \varphi$, entonces $\vdash^T \varphi$.

Quod erat demonstrandum

Teorema 2:

Sea $E_{S4}= <W_{S4}, \pi_{S4}, R_{1S4}, \ldots, R_{nS4}>$ el modelo canónico de S4. Entonces, para cada $j=1, \ldots, n$, R_{jS4} es reflexiva y transitiva.

Demostración

R_{jS4} es reflexiva porque $T \subseteq S4$ (*i. e.*, todo teorema de T es teorema de S4, en particular, $\vdash^{S4} K_j\alpha \rightarrow \alpha$).

P. D. Si $<w, w'> \in R_{jS4}$ y $<w', w''> \in R_{jS4}$, entonces $<w, w''> \in R_{jS4}$. Supongamos que $<w, w'> \in R_{jS4}$ y $<w', w''> \in R_{jS4}$. P. D. $<w, w''> \in R_{jS4}$. Es decir, K_j-$(w) \subseteq w''$. Tomemos α tal que

$K_j\alpha\in w$. P. D. $\alpha\in w''$. Como $\vdash^{S4}K_j\alpha\rightarrow K_jK_j\alpha$, por lema 2b, $K_jK_j\alpha\in w$. De esto, como $\langle w, w'\rangle\in R_{jS4}$, se tiene, por construcción del modelo canónico de S4, $K_j\alpha\in w'$. Como $\langle w', w''\rangle\in R_{jS4}$, $\alpha\in w''$.

$\therefore$

$\langle w,w\rangle\in R_{jS4}$, y si $\langle w, w'\rangle\in R_{jS4}$ y $\langle w', w''\rangle\in R_{jS4}$, entonces $\langle w, w''\rangle\in R_{jS4}$ [R_{jS4} es reflexiva y transitiva].

Metateorema de completez para S4:

Toda fórmula S4-válida es teorema de S4. Es decir, si $\vDash^{S4}\varphi$, entonces $\vdash^{S4}\varphi$.

Quod erat demonstrandum

Teorema 3:

Sea $E_{S5}= \langle W_{S5}, \pi_{S5}, R_{1S5}, \ldots, R_{nS5}\rangle$ el modelo canónico de S5. Entonces, para cada j=1, …, n, R_{Js5} es reflexiva y euclidiana.

Demostración

R_{jS5} es reflexiva porque $T\subseteq S5$ (*i. e.*, todo teorema de T es teorema de S5, en particular, $\vdash^{S5} K_j\alpha\rightarrow\alpha$).

P. D. Si $\langle w, w'\rangle\in R_{jS5}$ y $\langle w, w''\rangle\in R_{jS5}$, entonces $\langle w', w''\rangle\in R_{jS5}$. Supongamos que $\langle w, w'\rangle\in R_{jS5}$ y $\langle w, w''\rangle\in R_{jS5}$. P. D. $\langle w', w''\rangle\in R_{jS5}$. Es decir, K_j-$(w')\subseteq w''$. P. D. Si $K_j\alpha\in w'$, entonces $\alpha\in w''$. Por contraposición, supongamos que $\alpha\notin w''$. Como $\langle w, w''\rangle\in R_{jS5}$, $\alpha\notin K_j$-(w). Por consiguiente, $K_j\alpha\notin w$. Entonces tenemos, por lema 1a, $\neg K_j\alpha\in w$. Como $\vdash^{S5}\neg K_j\alpha\rightarrow K_j\neg K_j\alpha$, por lema 2b, $K_j\neg K_j\alpha\in w$. De esto, como $\langle w, w'\rangle\in R_{jS5}$, se tiene, por construcción del modelo canónico de S5, $\neg K_j\alpha\in w'$. Pero, por lema 1a, $K_j\alpha\notin w'$.

$\therefore$

$\langle w,w\rangle\in R_{jS5}$, y si $\langle w, w'\rangle\in R_{jS5}$ y $\langle w, w''\rangle\in R_{jS5}$, entonces $\langle w', w''\rangle\in R_{jS5}$ [R_{jS5} es reflexiva y euclidiana].

Metateorema de completez para S5:

Toda fórmula S5-válida es teorema de S5. Esto es, si $\vDash^{S5}\varphi$, entonces $\vdash^{S5}\varphi$.

Quod erat demonstrandum

Apéndice III

Lo que sigue a continuación es la prueba del metateorema de corrección para tres de los cuatro sistemas bimodales que se examinaron en el capítulo IV del presente trabajo, a saber, los sistemas OK E RIB, KB-$\{D^B\}$ y KB^{PP}. Además, con el apoyo de este metateorema, se prueba, entre otras cosas, que no toda instancia de la paradoja del creyente perfecto es teorema de las tres teorías mencionadas.

A3.1. Metateorema de correccción para OK & RIB

Definición de una OK & RIB-interpretación para $\mathcal{L}_{KB}$

Una OK E RIB–interpretación ($\mathcal{E}_{\text{OK \& RIB}}$) para $\mathcal{L}_{KB}$ es un tuplo:

$$\mathcal{E}_{\text{OK \& RIB}} = <S, \pi, R_1, \ldots, R_n, R^B_1, \ldots, R^B_n>.$$

Donde:

1) $S \neq \emptyset$ S es un conjunto no vacío (llamado el conjunto de estados s de $\mathcal{E}_{\text{OK \& RIB}}$).

2) π: $L \longrightarrow P(S)$ π es una función que va de L al conjunto potencia de S, donde L es el conjunto de letras proposicionales de $\mathcal{L}_{KB}$ ($\pi(p_i) \subseteq S$).

3) Para cada j=1, …, n, $R_j \subseteq S^2$ y $R^B{}_j \subseteq S^2$.

4) Para cada j=1, …, n, R_j es una relación de equivalencia.

5) Para cada j=1, …, n, $R^B{}_j$ es una relación serial y euclidiana.

7) Para cualesquiera s, t, u$\in$S, si $<s, t> \in R_j$ y $<t, u> \in R^B{}_j$, entonces $<s, u> \in R^B{}_j$.

8) Para cualesquiera s, t, u$\in$S, si $<s, t> \in R^B{}_j$ y $<t, u> \in R_j$, entonces $<s, u> \in R^B{}_j$[200].

---

[200] Nótese que una OK & RIB-interpretación se obtiene a partir de una KB-interpretación (de una interpretación para el sistema KB) al remover de ésta la condición 6): $R^B{}_j \subseteq R_j$, y agregar la condición 8): *Para cualesquiera s, t, u$\in$S, si $<s, t> \in R^B{}_j$ y $<t, u> \in R_j$, entonces $<s, u> \in R^B{}_j$.* Por ello, he decidido conservar la numeración que se tenía para las condiciones de una $\mathcal{KB}$-interpretación.

Definición de OK & RIB-verdad para $\mathcal{L}_{KB}$

Sea $\mathcal{E}$ una OK & RIB–interpretación cualquiera.

Definimos para cada $s \in S$ y cada *fbf* σ de $\mathcal{L}_{KB}$ lo que significa:

$\mathcal{E}, s \vDash \sigma$ (se lee: la fórmula σ es verdadera (realizada o satisfecha) en el estado s bajo la interpretación (estructura) $\mathcal{E}$).

Definición de OK E RIB-satisfacibilidad:

Cláusula 1: σ es una letra proposicional

Por lo tanto, $\sigma \in L$ (σ es una letra proposicional p_i)

$\mathcal{E}, s \vDash p_i$ sii $s \in \pi(p_i)$.

Cláusula 2: $\mathcal{E}, s \vDash \neg\sigma$ sii $\mathcal{E}, s \nvDash \sigma$.

Cláusula 3: $\mathcal{E}, s \vDash \alpha \rightarrow \beta$ sii $\mathcal{E}, s \nvDash \alpha$ o $\mathcal{E}, s \vDash \beta$.

Cláusula 4: Para $j=1, \ldots, n$, $\mathcal{E}, s \vDash K_j\sigma$ sii $\mathcal{E}, s' \vDash \sigma$ para todo $s' \in S$ tal que $\langle s,s' \rangle \in R_j$.

Cláusula 5: Para $j=1, \ldots, n$, $\mathcal{E}, s \vDash B_j\sigma$ sii $\mathcal{E}, s' \vDash \sigma$ para todo $s' \in S$ tal que $\langle s,s' \rangle \in R^B_j$.

Sea σ cualquier *fbf* de $\mathcal{L}_{KB}$.

Definición de OK & RIB-verdad:

σ es verdadera en una OK & RIB–interpretación $\mathcal{E}$ sii para todo estado s de $\mathcal{E}$ se tiene que $\mathcal{E}, s \vDash \sigma$.

Notación: $\mathcal{E} \vDash \sigma$

Definamos ahora lo que significa que σ sea una fórmula OK & RIB-válida de nuestro lenguaje.

Definición de OK & RIB-validez:

σ es OK & RIB-válida sii para toda OK & RIB–interpretación $\mathcal{E}$ se tiene que $\mathcal{E} \vDash \sigma$.

Notación: $\vDash^{OK} \sigma$

Con esta esta semántica, obtendremos el metateorema de corrección para la teoría OK E RIB. Formulemos el metateorema de corrección para OK E RIB de la manera siguiente:

Metateorema (corrección para OK & RIB)

Todo teorema de OK & RIB es OK & RIB-válido.

Probemos esta afirmación estableciendo que todos los axiomas de OK & RIB son OK & RIB-válidos y que sus reglas de inferencia preservan OK & RIB-validez. Téngase en cuenta que los axiomas y reglas de inferencia de OK & RIB son:

Axiomas de OK & RIB

Sean α, β y γ *fbf*´s cualesquiera de $\mathcal{L}_{KB}$ y sea j=1, ..., n.

A1: $\alpha\to(\beta\to\alpha)$,

A2: $(\alpha\to(\beta\to\gamma))\to((\alpha\to\beta)\to(\alpha\to\gamma))$,

A3: $(\neg\alpha\to\neg\beta)\to((\neg\alpha\to\beta)\to\alpha)$,

K: $K_j(\alpha\to\beta)\to(K_j\alpha\to K_j\beta)$ [Propiedad de distribución],

T: $K_j\alpha\to\alpha$ [Propiedad de conocimiento],

A5: $\neg K_j\alpha\to K_j\neg K_j\alpha$ [Propiedad de introspección negativa],

K^B: $B_j(\alpha\to\beta)\to(B_j\alpha\to B_j\beta)$ [Axioma K para creencia],

D^B: $B_j\alpha\to\neg B_j\neg\alpha$ [Axioma D para creencia],

KB2: $B_j\alpha\to K_jB_j\alpha$,

KB3: $B_j\alpha\to B_jK_j\alpha$.

Reglas de inferencia de OK & RIB

Sean α y β *fbf*´s cualesquiera de $\mathcal{L}_{KB}$.

MP: Si $\vdash^{OK}\alpha$ y $\vdash^{OK}\alpha\to\beta$, entonces $\vdash\beta$ [*Modus ponens*] y

KG: Si $\vdash^{OK}\alpha$, entonces $\vdash^{OK}K_i\alpha$ [Generalización del conocimiento].

Dado que los axiomas A1 a K son fórmulas válidas, *i. e.*, verdaderas en cualquier E estructura, también son fórmulas OK & RIB-válidas. Además, en el capítulo II se ha probado que el axioma T es T-válido y que el axioma A5 es S5-válido. En consecuencia, 184 ambos axiomas son también fórmulas OK & RIB-válidas. Sólo resta, pues, mostrar la OK & RIB -validez de los axiomas K^B, D^B, KB2 y KB3.

$\vDash^{OK} B_j(\alpha\to\beta)\to((B_j\alpha\to B_j\beta)$.

Prueba por reducción al absurdo.

Supongamos que existe una estructura OK & RIB–interpretación $\mathcal{E}$ y existe s de $\mathcal{E}$ tal que:

1) $\mathcal{E}$, s$\nvDash B_j(\alpha\to\beta)\to(B_j\alpha\to B_j\beta)$. De esta suposición se sigue que:

2) $\mathcal{E}$, s$\vDash B_j(\alpha\to\beta)$ y 3) $\mathcal{E}$, s$\nvDash B_j\alpha\to B_j\beta$.

De 3), tenemos que 4) $\mathcal{E}$, s$\vDash B_j\alpha$ y 5) $\mathcal{E}$, s$\nvDash B_j\beta$.

De 5), se sigue que existe s' de $\mathcal{E}$ tal que $<$ s, s'$>\in R^{B}{}_j$ y 6) $\mathcal{E}$, s'$\nvDash\beta$.

De 4), como $<$ s, s'$>\in R^{B}{}_j$, se sigue que 7) $\mathcal{E}$, s'$\vDash\alpha$.

De 6) y 7), tenemos que 8) $\mathcal{E}$, s'$\nvDash \alpha\to\beta$.

De 8), como $<$ s, s'$>\in R^{B}{}_j$, se sigue que 9) $\mathcal{E}$, s $\nvDash B_j(\alpha\to\beta)$**!**

Pero 9) se contradice con 2). Por lo tanto,

$\vDash^{OK} B_j(\alpha\to\beta)\to((B_j\alpha\to B_j\beta)$.

$\vDash^{OK} B_j\alpha\to\neg B_j\neg\alpha$.

Prueba por reducción al absurdo.

Supongamos que existe una OK & RIB–interpretación $\mathcal{E}$ y existe s de $\mathcal{E}$ tal que:

1) $\mathcal{E}$, s$\nvDash B_j\alpha\to\neg B_j\neg\alpha$. De esta suposición se sigue que:

2) $\mathcal{E}$, s$\vDash B_j\alpha$ y 3) $\mathcal{E}$, s$\nvDash\neg B_j\neg\alpha$.

De 3), tenemos que 4) $\mathcal{E}$, s$\vDash B_j\neg\alpha$.

Como $R^{B}{}_j$ es serial se tiene que 5) existe s' de $\mathcal{E}$ tal que $<$ s, s'$>\in R^{B}{}_j$.

De 5) y 2), se sigue que 6) $\mathcal{E}$, s'$\vDash\alpha$.

Ahora, de 5) y 4), tenemos que 7) $\mathcal{E}$, s'$\vDash\neg\alpha$. De esto se sigue que:

8) $\mathcal{E}$, s'$\nvDash\alpha$**!** Pero 8) se contradice con 6). Por lo tanto,

$\vDash^{OK} B_j\alpha\to\neg B_j\neg\alpha$.

$\models^{OK} B_j\alpha \rightarrow K_jB_j\alpha$

Prueba por reducción al absurdo.

Supongamos que existe una estructura OK & RIB–interpretación $\mathcal{E}$ y existe s de $\mathcal{E}$ tal que:

1) $\mathcal{E}$, s$\not\models B_j\alpha \rightarrow K_jB_j\alpha$. De esta suposición se sigue que:

2) $\mathcal{E}$, s$\models B_j\alpha$ y 3) $\mathcal{E}$, s$\not\models K_jB_j\alpha$.

De 3), se sigue que existe s’ de $\mathcal{E}$ tal que $<$ s, s’$>\in R_j$ y 4) $\mathcal{E}$, s’$\not\models B_j\alpha$.

De 4), se sigue que existe s’’ de $\mathcal{E}$ tal que $<$ s’, s’’$>\in R^B{}_j$ y 5) $\mathcal{E}$, s’’$\not\models \alpha$.

Como $<$ s, s’$>\in R_j$ y $<$ s’, s’’$>\in R^B{}_j$, tenemos, por condición (7) de nuestra definición de una OK E RIB-interpretación, que $<$ s, s’’$>\in R^B{}_j$. De esto y 5), se sigue que:

6) $\mathcal{E}$, s$\not\models B_j\alpha$**!** Pero 6) se contradice con 2). Por lo tanto,

$\models^{OK} B_j\alpha \rightarrow K_jB_j\alpha$.

$\models^{OK} B_j\alpha \rightarrow B_jK_j\alpha$

Prueba por reducción al absurdo.

Supongamos que existe una OK & RIB–interpretación y existe s de $\mathcal{E}$ tal que:

1) $\mathcal{E}$, s$\not\models B_j\alpha \rightarrow B_jK_j\alpha$. De esta suposición se sigue que:

2) $\mathcal{E}$, s$\models B_j\alpha$ y 3) $\mathcal{E}$, s$\not\models B_jK_j\alpha$.

De 3), se sigue que existe s’ de $\mathcal{E}$ tal que $<$ s, s’$>\in R^B{}_j$ y 4) $\mathcal{E}$, s’$\not\models K_j\alpha$.

De 4), se sigue que existe s’’ de $\mathcal{E}$ tal que $<$ s’, s’’$>\in R_j$ y 5) $\mathcal{E}$, s’’$\not\models \alpha$.

Como $<$ s, s’$>\in R^B{}_j$ y $<$ s’, s’’$>\in R_j$, tenemos, por condición (8) de nuestra definición de una OK E RIB-interpretación, que $<$ s, s’’$>\in R^B{}_j$. De esto y 5), se sigue que:

6) $\mathcal{E}$, s$\not\models B_j\alpha$**!** Pero 6) se contradice con 2). Por lo tanto,

$\models^{OK} B_j\alpha \rightarrow B_jK_j\alpha$

Ya que se ha probado que los axiomas de OK & RIB son fórmulas OK & RIB-válidas, probemos que sus reglas de inferencias preservan OK & RIB-validez[201]. Las reglas de inferencia de OK & RIB son MP (*Modus ponens*) y KG (*Generalización del conocimiento*). Sabemos que el *Modus ponens* es una forma de razonamiento correcto, *i. e.*, preserva verdad y, por tanto, preserva también OK & RIB-validez. Respecto a si la regla KG preserva OK & RIB-validez, considérese la siguiente prueba.

> Sea α una *fbf* de $\mathcal{L}_{KB}$ t. q. $\vDash^{OK} \alpha$ y sean $\mathcal{E}$ una OK & RIB-estructura cualquiera y s un estado cualquiera de $\mathcal{E}$.

Si aplicamos a α la regla KG, obtendremos la fórmula $K_j\alpha$, con j= 1, …, n. Como $\vDash^{OK} \alpha$, tendremos que, $\mathcal{E}$, s$\vDash K_j\alpha$, pues $\mathcal{E}$, s'$\vDash\alpha$ para cualquier s' de $\mathcal{E}$ tal que $<s. s'>\in R_j$. Como s es un estado cualquiera de $\mathcal{E}$, tenemos, por definición de OK E RIB-verdad, que $\mathcal{E}\vDash K_j\alpha$. Pero como $\mathcal{E}$ es una OK & RIB-estructura cualquiera, tenemos, por definición de OK & RIB-validez, que $\vDash^{OK} K_j\alpha$. Por lo tanto, la regla KG también preserva OK & RIB -validez.

Ahora que se ha establecido que todos los axiomas del sistema OK & RIB son fórmulas OK & RIB-válidas y que sus reglas de inferencia preservan OK & RIB-validez, probemos el metateorema de corrección para este sistema.

<u>Metateorema de corrección para OK & RIB</u>

> Sea φ una *fbf* cualquiera $\mathcal{L}_{KB}$.
>
> Si $\vdash^{OK} \varphi$, entonces $\vDash^{OK} \varphi$.

Demostración

Demostración por inducción matemática sobre el número m de pasos de la demostración de φ.

---

[201] Lo que viene a continuación en este apartado es esencialmente una transcripción, con las modificaciones pertinentes, de la prueba del metateorema de corrección para el sistema K que presenté en el capítulo II. Pese a ello, resolví incluir aquí dicha transcripción para evitar que el lector se remita al capítulo en cuestión.

Base: m=1 (la demostración de σ tiene un renglón)

Entonces φ se demuestra en un paso. Por lo tanto, φ es un axioma de OK & RIB (A1-KB3). Pero sabemos que todo axioma de OK & RIB es una fórmula OK & RIB-válida. Por lo tanto, $\vDash^{OK} \varphi$.

m= n, para n>1 (la demostración de φ consta de n renglones, con n>1)

HI: El metateorema vale para todas las fórmulas que aparecen en los renglones anteriores a n.

P.D. El metateorema vale para φ (que aparece en el renglón n).

Hay tres casos para φ: φ es axioma de OK & RIB, es consecuencia de anteriores por MP, o bien, es consecuencia de anteriores por KG.

Caso 1: es axioma de OK & RIB. Se procede igual que en la base.

Caso 2: σ es consecuencia de anteriores por MP. Entonces φ se obtuvo, por MP, de las fórmulas $\alpha \rightarrow \varphi$ y α, que aparecen en renglones anteriores de la demostración. Por consiguiente, a estas dos fórmulas les podemos aplicar HI, de modo que $\vDash^{OK} \alpha \rightarrow \varphi$ y $\vDash^{OK} \alpha$. Como la regla de inferencia MP preserva OK & RIB-validez, tenemos que $\vDash^{OK} \varphi$.

Caso 3: φ es consecuencia de anteriores por KG. Entonces φ es de la forma $K_j\alpha$ y se obtuvo, por KG, de la fórmula α, que aparece en un renglón anterior de la demostración. En consecuencia, a esta fórmula se le puede aplicar HI, de forma que tenemos que $\vDash^{OK} \alpha$. Como la regla de inferencia KG preserva OK & RIB-validez, tenemos que $\vDash^{OK} \varphi$.

De la base y los casos 1-3, se sigue, por *principio de inducción matemática completa* (PIMC), que:

Si $\vdash^{OK} \varphi$, entonces $\vDash^{OK} \varphi$.

Ahora que hemos probado el metateorema de corrección para OK & RIB, *i. e.*, que todos sus teoremas son OK & RIB-válidos, estamos en condiciones de probar que no toda instancia del esquema $K_j\alpha \rightarrow B_j\alpha$, así como no toda instancia la paradoja del creyente perfecto ($B_jK_j\alpha \rightarrow K_j\alpha$), es teoremas de este sistema. Para ello bastará con mostrar que alguna de sus instancias no es OK & RIB-válida. Probemos que las fórmulas $K_1p_1 \rightarrow B_1p_1$ (instancia de $K_j\alpha \rightarrow B_j\alpha$) y $B_1K_1p_1 \rightarrow K_1p_1$ (instancia de $B_jK_j\alpha \rightarrow K_j\alpha$) no son OK & RIB-válidas.

$\not\models^{OK} K_1p_1 \rightarrow B_1p_1$

Considérese la siguiente OK & RIB-interpretación para $\mathscr{L}_{KB}$.

$\mathcal{E} = \langle S, \pi, R_1, \ldots, R_n, R^B_1, \ldots, R^B_n \rangle$.

$S = \{r, s\}$

$\pi: L \longrightarrow P(S)$

$\pi(\psi) = \emptyset$, para toda letra proposicional ψ, excepto la letra proposicional p_1.

$\pi(p_1) = \{r\}$

$R_j = \{\langle r, r\rangle, \langle s, s\rangle\}$, con $j = 1, \ldots, n$

$R^B_j = \{\langle r, s\rangle, \langle s, s\rangle\}$, con $j = 1, \ldots, n$

Prueba:

Como $r \in \pi(p_1)$ y $s \notin \pi(p_1)$, tenemos, respectivamente, por cláusula 1 de nuestra definición de satisfacibilidad:

1) $\mathcal{E}, r \models p_1$ y

2) $\mathcal{E}, s \not\models p_1$.

Dadas R_1 y R^B_1, de 1) y 2) se sigue, por cláusula 4 y 5 de nuestra definición de satisfacibilidad, respectivamente, que:

3) $\mathcal{E}, r \models K_1p_1$ y

4) $\mathcal{E}, r \not\models B_1p_1$.

De 3) y 4) se tiene, por cláusula 3 de nuestra definición de satisfacibilidad:

6) $\mathcal{E}, r \not\models K_1p_1 \rightarrow B_1p_1$.

Por lo tanto, dada nuestra definición de OK & RIB-validez:

$\not\models^{OK} K_1p_1 \rightarrow B_1p_1$.

Tal prueba muestra que la fórmula $B_1K_1p_1 \rightarrow K_1p_1$, que es una instancia del esquema $K_j\alpha \rightarrow B_j\alpha$, no es OK & RIB-válida. En consecuencia, por el metateorema de corrección, $\nvdash^{OK} B_1K_1p_1 \rightarrow K_1p_1$. Por tanto, no toda instancia del esquema en cuestión es teorema del sistema OK & RIB.

$\nvDash^{OK} B_1K_1p_1 \rightarrow K_1p_1$

Considérese la siguiente OK & RIB-interpretación para $\mathcal{L}_{KB}$.

$\mathcal{E} = \langle S, \pi, R_1, \ldots, R_n, R^B{}_1, \ldots, R^B{}_n \rangle$.

$S = \{r, s\}$

$\pi: L \longrightarrow P(S)$

$\pi(\psi) = \emptyset$, para toda letra proposicional ψ, excepto la letra proposicional p_1.

$\pi(p_1) = \{s\}$

$R_j = \{\langle r, r\rangle, \langle s, s\rangle\}$, con $j = 1, \ldots, n$

$R^B{}_j = \{\langle r, s\rangle, \langle s, s\rangle\}$, con $j = 1, \ldots, n$

Prueba:

Como $r \notin \pi(p_1)$ y $s \in \pi(p_1)$, tenemos, respectivamente, por cláusula 1 de nuestra definición de satisfacibilidad:

1) $\mathcal{E}, r \nvDash p_1$ y

2) $\mathcal{E}, s \vDash p_1$.

Dada R_1, de 1) y 2) se sigue, respectivamente, por cláusula 4 de nuestra definición de satisfacibilidad:

3) $\mathcal{E}, r \nvDash K_1p_1$ y

4) $\mathcal{E}, s \vDash K_1p_1$.

Dada $R^B{}_1$, de 4) se sigue, por cláusula 5 de nuestra definición de satisfacibilidad:

5) $\mathcal{E}, r \vDash B_1K_1p_1$.

De 5) y 3) se tiene, por cláusula 3 de nuestra definición de satisfacibilidad:

6) $\mathcal{E}, r \nvDash B_1K_1p_1 \rightarrow K_1p_1$.

Por lo tanto, dada nuestra definición de OK & RIB-validez:

$\nvDash^{OK} B_1K_1p_1 \rightarrow K_1p_1$.

Esta prueba muestra que la fórmula $B_1K_1p_1 \rightarrow K_1p_1$, que es una instancia de la paradoja del creyente perfecto, no es OK & RIB-válida. En consecuencia, por el metateorema de corrección, $\nvDash^{OK}$ $B_1K_1p_1 \rightarrow K_1p_1$. Por tanto, no toda instancia de la paradoja del creyente perfecto es teorema del sistema OK & RIB.

A3.2. Metateorema de corrección para KB-{D^B}

Definición de una KB-{D^B}-interpretación para $\mathcal{L}_{KB}$

Una KB-{D^B}-interpretación ($\mathcal{E}_{KB\text{-}\{D^B\}}$) para $\mathcal{L}_{KB}$ es un tuplo:

$\mathcal{E}_{KB\text{-}\{D^B\}} = \langle S, \pi, R_1, \ldots, R_n, R^B_1, \ldots, R^B_n \rangle$.

Donde:

1) $S \neq \emptyset$ S es un conjunto no vacío (llamado el conjunto de estados s de $\mathcal{E}_{KB\text{-}\{D^B\}}$).

2) $\pi: L \longrightarrow P(S)$ π es una función que va de L al conjunto potencia de S, donde L es el conjunto de letras proposicionales de $\mathcal{L}_{KB}$ ($\pi(p_i) \subseteq S$).

3) Para cada j=1, …, n, $R_j \subseteq S^2$ y $R^B_j \subseteq S^2$.

4) Para cada j=1, …, n, R_j es una relación de equivalencia.

6) $R^B_j \subseteq R_j$.

7) Para cualesquiera s, t, u∈S, si $\langle s, t \rangle \in R_j$ y $\langle t, u \rangle \in R^B_j$, entonces $\langle s, u \rangle \in R^B_j$[202].

Definición de KB-{D^B}-verdad para $\mathcal{L}_{KB}$

Las cláusulas para KB-{D^B}-satisfacibilidad son, *mutatis mutandis*, las mismas que para la semántica del sistema OK & RIB.

Sea σ cualquier *fbf* de $\mathcal{L}_{KB}$.

Definición de KB-{D^B}-verdad:

[202] Una KB-{D^B}-interpretación se obtiene a partir de una KB-interpretación (de una interpretación para el sistema KB) al remover de ésta la condición 5): *Para cada j=1, …, n, R^B_j es una relación serial y euclidiana.*

σ es verdadera en una KB-{D^B}-interpretación $\mathcal{E}$ sii para todo estado s de $\mathcal{E}$ se tiene que $\mathcal{E}$, s⊨σ.

Notación: $\mathcal{E}$⊨σ

Definamos ahora lo que significa que σ sea una fórmula KB-{D^B}-válida de nuestro lenguaje.

Definición de KB-{D^B}-validez:

σ es KB-{D^B}-válida sii para toda KB-{D^B}-interpretación $\mathcal{E}$ se tiene que $\mathcal{E}$⊨σ

Notación: $\vDash^{\{Db\}}\sigma$

Con esta esta semántica, obtendremos el metateorema de corrección para la teoría KB-{D^B}.

Metateorema (corrección para KB-{D^B})

Todo teorema de KB-{D^B} es KB-{D^B}-válido.

Una vez más, probaremos esta afirmación estableciendo que todos los axiomas de KB-{D^B} son KB-{D^B}-válidos y que sus reglas de inferencia preservan KB-{D^B}-validez. Téngase en cuenta que los axiomas y reglas de inferencia de KB-{D^B} son:

Axiomas de KB-{D^B}

Sean α, β y γ *fbf´s* cualesquiera de $\mathcal{L}_{KB}$ y sea j=1, …, n.

A1: $\alpha\rightarrow(\beta\rightarrow\alpha)$,

A2: $(\alpha\rightarrow(\beta\rightarrow\gamma))\rightarrow((\alpha\rightarrow\beta)\rightarrow(\alpha\rightarrow\gamma))$,

A3: $(\neg\alpha\rightarrow\neg\beta)\rightarrow((\neg\alpha\rightarrow\beta)\rightarrow\alpha)$,

K: $K_j(\alpha\rightarrow\beta)\rightarrow(K_j\alpha\rightarrow K_j\beta)$ [Propiedad de distribución],

T: $K_j\alpha\rightarrow\alpha$ [Propiedad de conocimiento],

A5: $\neg K_j\alpha\rightarrow K_j\neg K_j\alpha$ [Propiedad de introspección negativa],

K^B: $B_j(\alpha\rightarrow\beta)\rightarrow(B_j\alpha\rightarrow B_j\beta)$ [Axioma K para creencia],

KB1: $K_j\alpha\rightarrow B_j\alpha$,

KB2: $B_j\alpha\rightarrow K_jB_j\alpha$.

Reglas de inferencia de KB-$\{D^B\}$

Sean α y β *fbf*'s cualesquiera de $\mathcal{L}_{KB}$.

MP: Si $\vdash^{\{Db\}} \alpha$ y $\vdash^{\{Db\}} \alpha\rightarrow\beta$, entonces $\vdash\beta$ [*Modus ponens*] y

KG: Si $\vdash^{\{Db\}} \alpha$, entonces $\vdash^{\{Db\}} K_i\alpha$ [Generalización del conocimiento].

A excepción del axioma KB1, para la prueba de que los axiomas de la teoría KB-$\{D^B\}$ son KB-$\{D^B\}$-válidos se procede de la misma manera que en el caso de los axiomas del sistema OK & RIB. Por consiguiente, únicamente resta probar que el axioma KB1: $K_j\alpha\rightarrow B_j\alpha$ es KB-$\{D^B\}$-válido también.

$\vDash^{\{Db\}} K_j\alpha\rightarrow B_j\alpha$

Prueba por reducción al absurdo.
Supongamos que existe una KB-$\{D^B\}$-interpretación $\mathcal{E}$ y existe s de $\mathcal{E}$ tal que:

1) $\mathcal{E}$, s$\nvDash K_j\alpha\rightarrow B_j\alpha$. De esta suposición se sigue que:

2) $\mathcal{E}$, s$\vDash K_j\alpha$ y 3) $\mathcal{E}$, s$\nvDash B_j\alpha$.

De 3), se sigue que 4) existe s' de $\mathcal{E}$ tal que $<$ s, s'$>\in R^B{}_j$ y 5) $\mathcal{E}$, s'$\nvDash\alpha$.

Como $R^B{}_j\subseteq R_j$, de 4) se tiene que:

6) $<$ s, s'$>\in R_j$. De 6) y 5) se sigue que 7) $\mathcal{E}$, s$\nvDash K_j\alpha$**!**

Pero 7) se contradice con 2). Por lo tanto,

$\vDash^{\{Db\}} K_j\alpha\rightarrow B_j\alpha$.

Para establecer que las reglas de inferencia de KB-$\{D^B\}$ (MP y KG) preservan KB-$\{D^B\}$-validez, así como para probar el metateorema de corrección para este sistema, se procede de la misma forma que para la teoría OK & RIB.

De lo anterior se sigue que:

Si $\vdash^{\{Db\}} \varphi$, entonces $\vDash^{\{Db\}} \varphi$.

Una vez más, emplearemos el metateorema de corrección para probar que ciertas fórmulas no son teoremas del sistema KB-$\{D^B\}$; probaremos que no toda instancia de los esquemas de fórmulas $B_j\alpha\rightarrow\neg B_j\neg\alpha$ y $\neg B_j(\alpha\wedge\neg\alpha)$, así como no toda instancia de la paradoja del perfecto creyente, es teorema de esta teoría. Para ello, probemos que las fórmulas $B_1p_1\rightarrow\neg B_1\neg p_1$ (instancia de $B_j\alpha\rightarrow\neg B_j\neg\alpha$), $\neg B_1(p_1\wedge\neg p_1)$ (instancia de $\neg B_j(\alpha\wedge\neg\alpha)$) y $B_1K_1p_1\rightarrow K_1p_1$ (instancia de la paradoja del creyente perfecto) no son KB-{D^B}-válidas.

$\nvDash^{\{Db\}} B_1p_1\rightarrow\neg B_1\neg p_1$

Considérese la siguiente KB-{D^B}-interpretación para $\mathcal{L}_{KB.}$

$\mathcal{E}$= <S, π, R_1, ..., R_n, $R^B{}_1$, ..., $R^B{}_n$ >.

S={r}

π: L⟶P(S)

π(ψ)=∅, para toda letra proposicional ψ.

R_j= {<r, r>}, con j= 1, ..., n

$R^B{}_j$= ∅, con j= 1, ,..., n

Prueba:

Dada $R^B{}_1$, se tiene, por cláusula 5 de nuestra definición de satisfacibilidad:

1) $\mathcal{E}$, r$\vDash B_1p_1$ y

2) $\mathcal{E}$, r$\vDash B_1\neg p_1$

De 2) se tiene, por cláusula 2 de nuestra definición de satisfacibilidad:

3) $\mathcal{E}$, r$\nvDash\neg$B1$\neg$p1

De 1) y 2) se sigue, por cláusula 3 de nuestra definición de satisfacibilidad:

5) $\mathcal{E}$, r$\vDash B_1p_1\rightarrow\neg B_1\neg p_1$

Por lo tanto, dada nuestra definición de KB-{D^B}-validez:

$\nvDash^{\{Db\}} B_1p_1\rightarrow\neg B_1\neg p_1$.

Esta prueba muestra que la fórmula $B_1p_1\rightarrow\neg B_1\neg p_1$, que es una instancia de la fórmula $B_j\alpha\rightarrow\neg B_j\neg\alpha$, no es KB-{$D^B$}-válida. En consecuencia, por el metateorema de corrección, $\nvdash^{\{Db\}} B_1p_1\rightarrow\neg B_1\neg p_1$. Por tanto, no toda instancia de la fórmula en cuestión es teorema del sistema KB-{D^B}.

$\nvDash^{KB\text{-}\{Db\}} \neg B_1(p_1\wedge\neg p_1)$

Empleemos la interpretación anterior para probar esta afirmación.

Prueba:

Dada R^{B}_{1}, se tiene, por cláusula 5 de nuestra definición de satisfacibilidad:

1) $\mathcal{E}$, r$\vDash B_1(p_1\wedge\neg p_1)$.

De 1) se tiene, por cláusula 2 de nuestra definición de satisfacibilidad:

2) $\mathcal{E}$, r$\nvDash \neg B_1(p_1\wedge\neg p_1)$.

Por lo tanto, dada nuestra definición de KB-{D^B}-validez:

$\nvDash^{\{Db\}} \neg B_1(p_1\wedge\neg p_1)$.

Tal prueba establece que la fórmula $\neg B_1(p_1\wedge\neg p_1)$, que es una instancia de la fórmula $\neg B_j(\alpha\wedge\neg\alpha)$, no es KB-{$D^B$}-válida. En consecuencia, por el metateorema de corrección, $\nvdash^{\{Db\}} \neg B_1(p_1\wedge\neg p_1)$. Por tanto, no toda instancia de la fórmula en cuestión es teorema del sistema KB-{D^B}.

$\nvDash^{\{Db\}} B_1K_1p_1\rightarrow K_1p_1$

Probemos esta afirmación empleando una vez más la interpretación inicial.

Prueba:

Como $r\notin\pi(p_1)$, tenemos, por cláusula 1 de nuestra definición de satisfacibilidad:

1) $\mathcal{E}$, r$\nvDash p_1$.

Dada R_1, de 1) se sigue, por cláusula 4 de nuestra definición de satisfacibilidad:

3) $\mathcal{E}$, r$\nvDash K_1p_1$.

Dada R^{B}_{1}, de 3) se tiene, por cláusula 5 de nuestra definición de satisfacibilidad:

4) $\mathcal{E}$, r$\vDash B_1K_1p_1$

De 4) y 3) se sigue, por cláusula 3 de nuestra definición de satisfacibilidad:

5) $\mathcal{E}$, r$\vDash B_1K_1p_1\rightarrow K_1p_1$

Por lo tanto, dada nuestra definición de KB-{D^B}-validez:

$\nvDash^{\{Db\}} B_1K_1p_1\rightarrow K_1p_1$

La prueba anterior establece que la fórmula $B_1K_1p_1\rightarrow K_1p_1$, que es una instancia de la paradoja del creyente perfecto, no es KB-{D^B}-válida. En consecuencia, por el metateorema de corrección, $\nvdash^{\{Db\}}$ $B_1K_1p_1 \rightarrow K_1p_1$. Por tanto, no toda instancia de la paradoja del creyente perfecto es teorema del sistema KB-$\{D^B\}$.

A3.3. Metateorema de corrección para KB^{PG}

Definición de una KB^{PG}-interpretación para $\mathcal{L}_{KB}$

Una KB^{PG}-interpretación ($\mathcal{E}_{KB}{}^{PG}$) para $\mathcal{L}_{KB}$ es un tuplo:

$\mathcal{E}_{KB}{}^{PG} = \langle S, \pi, R_1, \ldots, R_n, R^B{}_1, \ldots, R^B{}_n \rangle$.

Donde:

1) $S \neq \emptyset$ S es un conjunto no vacío (llamado el conjunto de estados s de $\mathcal{E}_{KB}{}^{PG}$).

2) $\pi: L \longrightarrow P(S)$ π es una función que va de L al conjunto potencia de S, donde L es el conjunto de letras proposicionales de $\mathcal{L}_{KB}$ ($\pi(p_i) \subseteq S$).

3) Para cada j=1, …, n, $R_j \subseteq S^2$ y $R^B{}_j \subseteq S^2$.

4) Para cada j=1, …, n, R_j es una relación reflexiva.

5) Para cada j=1, …, n, $R^B{}_j$ es una relación serial y euclidiana.

6) $R^B{}_j \subseteq R_j$.

7) Para cualesquiera s, t, $u \in S$, si $\langle s, t \rangle \in R^B{}_j$ y $\langle t, u \rangle \in R_j$, entonces $\langle s, u \rangle \in R^B{}_j$.

Definición de KB^{PG}-verdad para $\mathcal{L}_{KB}$

Las cláusulas para KB^{PG}-satisfacibilidad son, *mutatis mutandis*, las mismas que para la semántica del sistema OK & RIB.

Sea σ cualquier *fbf* de $\mathcal{L}_{KB}$.

Definición de KB^{PG}-verdad:

σ es verdadera en una KB^{PG}-interpretación $\mathcal{E}$ sii para todo estado s de $\mathcal{E}$ se tiene que $\mathcal{E}$, $s \vDash \sigma$.

Notación: $\mathcal{E} \vDash \sigma$

Definamos ahora lo que significa que σ sea una fórmula KB^{PG}-válida de nuestro lenguaje.

Definición de KB^{PG}-validez:

σ es KB^{PG}-válida sii para toda KB^{PG}-interpretación $\mathcal{E}$ se tiene que $\mathcal{E}\vDash\sigma$

Notación: $\vDash^{PG}\sigma$

Con esta semántica, obtendremos el metateorema de corrección para la teoría KB^{PG}.

Metateorema (corrección para KB^{PG})

Todo teorema de KB^{PG} es KB^{PG}-válido.

Al igual que en los sistemas anteriores, probaremos esta afirmación estableciendo que todos los axiomas de KB^{PG} son KB^{PG}-válidos y que sus reglas de inferencia preservan KB^{PG}-validez. Recuérdese que los axiomas y reglas de inferencia de KB^{PG} son:

Axiomas de KB^{PG}

Sean α, β y γ *fbf's* cualesquiera de $\mathcal{L}_{KB}$ y sea $j=1, \ldots, n$.

A1: $\alpha\rightarrow(\beta\rightarrow\alpha)$,

A2: $(\alpha\rightarrow(\beta\rightarrow\gamma))\rightarrow((\alpha\rightarrow\beta)\rightarrow(\alpha\rightarrow\gamma))$,

A3: $(\neg\alpha\rightarrow\neg\beta)\rightarrow((\neg\alpha\rightarrow\beta)\rightarrow\alpha)$,

K: $K_j(\alpha\rightarrow\beta)\rightarrow(K_j\alpha\rightarrow K_j\beta)$ [Propiedad de distribución],

T: $K_j\alpha\rightarrow\alpha$ [Propiedad de conocimiento],

K^B: $B_j(\alpha\rightarrow\beta)\rightarrow(B_j\alpha\rightarrow B_j\beta)$ [Axioma K para creencia],

D^B: $B_j\alpha\rightarrow\neg B_j\neg\alpha$ [Axioma D para creencia],

KB1: $K_j\alpha\rightarrow B_j\alpha$,

KB3: $B_j\alpha\rightarrow B_jK_j\alpha$.

Reglas de inferencia de KB^{PG}

Sean α y β *fbf's* cualesquiera de $\mathcal{L}_{KB}$.

MP: Si $\vdash^{PG} \alpha$ y $\vdash^{PG} \alpha \rightarrow \beta$, entonces $\vdash^{PG} \beta$ [*Modus ponens*] y

KG: Si $\vdash^{PG} \alpha$, entonces $\vdash^{PG} K_i\alpha$ [Generalización del conocimiento].

La justificación de que los axiomas A1 a T son KB^{PG}-válidos es básicamente la misma que para el sistema OK & RIB. Ahora bien, para la prueba de que el axioma KB1 es KB^{PG}-válido, se procede de manera idéntica a como se hizo en el sistema KB-$\{D^B\}$, mientras que para probar que los axiomas D^B y KB3 son KB^{PG}-válidos, que las reglas de inferencia de KB^{PG} (MP y KG) preservan KB^{PG}-validez, así como para probar el metateorema de corrección para este sistema, se procede exactamente igual que para la teoría OK & RIB.

De lo anterior se sigue que:

Si $\vdash^{PG}\varphi$, entonces $\vDash^{PG}\varphi$.

Al igual que con los sistemas OK & RIB y KB-$\{D^B\}$, emplearemos el metateorema de corrección para KB^{PG} con la finalidad de establecer que determinadas fórmulas no son teoremas de este sistema. Probaremos que no toda instancia de los esquemas de fórmulas $\neg K_j\alpha \rightarrow K_j\neg K_j\alpha$ (introspección negativa), $B_jK_j\alpha \rightarrow K_j\alpha$ (paradoja del creyente perfecto) y $B_j\alpha \rightarrow \alpha$ (paradoja de la infalibilidad) es teorema de KB^{PG}. Con este propósito, probaremos primero que sus instancias respectivas $\neg K_1p_1 \rightarrow K_1\neg K_1p_1$, $B_1K_1p_1 \rightarrow K_1p_1$ y $B_1p_1 \rightarrow p_1$ no son KB^{PG}-válidas.

$\nvDash^{PG}\neg K_1p_1 \rightarrow K_1\neg K_1p_1$

Considérese la siguiente KB^{PG}-interpretación para $\mathcal{L}_{KB}$.

$\mathcal{E} = <S, \pi, R_1, \ldots, R_n, R^B_1, \ldots, R^B_n>$.

$S=\{r, s\}$

$\pi: L \longrightarrow P(S)$

$\pi(\psi)=\emptyset$, para toda letra proposicional ψ, excepto la letra proposicional p_1.

$\pi(p_1)=\{s\}$

$R_j= \{<r, r>, <s, s>, <r, s>\}$, con $j= 1, \ldots, n$

$R^B_j= \{<r, s>, <s, s>\}$, con $j= 1, \ldots, n$

Prueba:

Como $r \notin \pi(p_1)$ y $s \in \pi(p_1)$, tenemos, respectivamente, por cláusula 1 de nuestra definición de satisfacibilidad:

1) $\mathcal{E}, r \nvDash p_1$ y

2) $\mathcal{E}, s \vDash p_1$.

Dada R_1, de 1) y 2) se sigue, respectivamente, por cláusula 4 de nuestra definición de satisfacibilidad:

3) $\mathcal{E}, r \nvDash K_1p_1$ y

4) $\mathcal{E}, s \vDash K_1p_1$.

De 3) y 4) tenemos, respectivamente, por cláusula 4 de nuestra definición de satisfacibilidad:

4) $\mathcal{E}, r \vDash \neg K_1p_1$ y

5) $\mathcal{E}, s \nvDash \neg K_1p_1$.

Dada R_1, de 5) se sigue, por cláusula 4 de nuestra definición de satisfacibilidad:

6) $\mathcal{E}, r \nvDash K_1\neg K_1p_1$

De 6) y 4) se tiene, por cláusula 3 de nuestra definición de satisfacibilidad:

6) $\mathcal{E}, r \nvDash \neg K_1p_1 \rightarrow K_1\neg K_1p_1$.

Por lo tanto, dada nuestra definición de KB^{PG}-validez:

$\nvDash^{PG} \neg K_1p_1 \rightarrow K_1\neg K_1p_1$.

Esta prueba muestra que la fórmula $\neg K_1p_1 \rightarrow K_1\neg K_1p_1$, que es una instancia del esquema $\neg K_j\alpha \rightarrow K_j\neg K_j\alpha$, no es KB^{PG}-válida. En consecuencia, por el metateorema de corrección, $\nvdash^{PG} \neg K_1p_1 \rightarrow K_1\neg K_1p_1$. Por tanto, no toda instancia del esquema en cuestión es teorema del sistema KB^{PG}.

$\nvDash^{PG} B_1K_1p_1 \rightarrow K_1p_1$

La interpretación anterior nos permite probar también esta afirmación.

Prueba:

Como $r \notin \pi(p_1)$ y $s \in \pi(p_1)$, tenemos, respectivamente, por cláusula 1 de nuestra definición de satisfacibilidad:

1) $\mathcal{E}, r \nvDash p_1$ y

2) $\mathcal{E}, s \vDash p_1$.

Dada R_1, de 1) y 2) se sigue, respectivamente, por cláusula 4 de nuestra definición de satisfacibilidad:

3) $\mathcal{E}, r \nvDash K_1p_1$ y

4) $\mathcal{E}, s \vDash K_1p_1$.

Dada $R^B{}_1$, de 4) se sigue, por cláusula 5 de nuestra definición de satisfacibilidad:

5) $\mathcal{E}, r \nvDash B_1K_1p_1$

De 5) y 3) se tiene, por cláusula 3 de nuestra definición de satisfacibilidad:

6) $\mathcal{E}, r \nvDash B_1K_1p_1 \rightarrow K_1p_1$.

Por lo tanto, dada nuestra definición de KB^{PG}-validez:

$\nvDash^{PG} B_1K_1p_1 \rightarrow K_1p_1$.

Esta prueba muestra que la fórmula $B_1K_1p_1 \rightarrow K_1p_1$, que es una instancia de la paradoja del creyente perfecto, no es KB^{PG}-válida. En consecuencia, por el metateorema de corrección, $\nvdash^{PG} B_1K_1p_1 \rightarrow K_1p_1$. Por tanto, no toda instancia de esta paradoja es teorema del sistema KB^{PG}.

$\nvDash^{PG} B_1p_1 \rightarrow p_1$

Emplearemos la misma interpretación que para las dos afirmaciones anteriores.

Prueba:

Como $r \notin \pi(p_1)$ y $s \in \pi(p_1)$, tenemos, respectivamente, por cláusula 1 de nuestra definición de satisfacibilidad:

1) $\mathcal{E}, r \nvDash p_1$ y

2) $\mathcal{E}, s \vDash p_1$.

Dada $R^B{}_1$, de 2) se sigue, por cláusula 5 de nuestra definición de satisfacibilidad:

3) $\mathcal{E}, r \vDash B_1p_1$

De 3) y 1) se tiene, por cláusula 3 de nuestra definición de satisfacibilidad:

6) $\mathcal{E}, r \nvDash B_1p_1 \rightarrow p_1$.

Por lo tanto, dada nuestra definición de KB^{PG}-validez:

$\not\models^{PG} B_1p_1 \rightarrow p_1$.

La prueba anterior muestra que la fórmula $B_1p_1 \rightarrow p_1$, que es una instancia de la paradoja de la infalibilidad, no es KB^{PG}-válida. En consecuencia, por el metateorema de corrección, $\nvdash^{PG} B_1p_1 \rightarrow p_1$. Por tanto, no toda instancia de esta paradoja es teorema del sistema KB^{PG}.

Referencias bibliográficas

Alfred Jules Ayer, *The Problem of Knowledge*, Macmillan, London, 1956.

Aristóteles, *Metafísica,* Trad. Valentín García Yebra, Gredos (Ed.), Madrid, 1970.

---------------, "Tópicos y Sobre las refutaciones sofísticas", en *Órganon,* Vol. 1, Ed. y Trad. M. Candel Sanmartín, Gredos (Ed.), Madrid, 1988.

Feldman, Richard, "Skepticism (I)", *Epistemology*, New Jersey, Prentice Hall, 2003.

Freund, Max, "Lógica epistémica", en *Enciclopedia iberoamericana de filosofía* (Volumen 7), Trotta (Ed.), 1995.

Kant, Immanuel, *Lógica. Acompañada de Una Selección de Reflexiones del Legado de Kant,* Akal (Ed.), Madrid, 2001.

M. Gabbay y Wood J. (ed.), *Handbook of the history of logic* (vol 7), M. Gabbay y Wood J. (Editores), 2006.

Platón, *Menón*, Trad. J. Calonge, Gredos (Ed.), Madrid, 1983.

Radu J. Bogdan (ed.), '*Jaakko Hintikka*', D. Reidel Publishing Company. 1987.

Rosenhouse, Jason, *The Monty Hall Problem: The Remarkable Story of Math's Most Contentious Brainteaser*, Oxford University Press, New York, 2009.

Villoro, Luis, *Creer, saber, conocer,* Siglo XXI (Ed.), México, 1989.

Wolfgang, Lenzen, *Recent work in epistemic logic*, Acta Philosophica Fennica, 30:5–219, 1978.

Casa abierta al tiempo
UNIVERSIDAD AUTÓNOMA METROPOLITANA

ACTA DE EXAMEN DE GRADO

No. 00408
Matrícula: 2163800531

Consideraciones en torno a la Lógica Epistémica.

Con base en la Legislación de la Universidad Autónoma Metropolitana, en la Ciudad de México se presentaron a las 13:00 horas del día 2 del mes de marzo del año 2021 POR VÍA REMOTA ELECTRÓNICA, los suscritos miembros del jurado designado por la Comisión del Posgrado:

DR. JOSE MARCOS NICOLAS JAVIER DE TERESA OCHOA
DR. JOSE JORGE MAX FERNANDEZ DE CASTRO TAPIA
DR. VICTOR CANTERO FLORES
DRA. YOLANDA MAGDA TORRES FALCON

Bajo la Presidencia del primero y con carácter de Secretaria la última, se reunieron para proceder al Examen de Grado cuya denominación aparece al margen, para la obtención del grado de:

MAESTRO EN HUMANIDADES (FILOSOFIA)

DE: ADAN CONTRERAS QUINTERO

ADAN CONTRERAS QUINTERO
ALUMNO

y de acuerdo con el artículo 78 fracción III del Reglamento de Estudios Superiores de la Universidad Autónoma Metropolitana, los miembros del jurado resolvieron:

APROBAR

Acto continuo, el presidente del jurado comunicó al interesado el resultado de la evaluación y, en caso aprobatorio, le fue tomada la protesta.

REVISÓ

MTRA. ROSALIA SERRANO DE LA PAZ
DIRECTORA DE SISTEMAS ESCOLARES

DIRECTOR DE LA DIVISIÓN DE CSH

DR. JUAN MANUEL HERRERA CABALLERO

PRESIDENTE

DR. JOSE MARCOS NICOLAS JAVIER DE TERESA OCHOA

VOCAL

DR. JOSE JORGE MAX FERNANDEZ DE CASTRO TAPIA

VOCAL

DR. VICTOR CANTERO FLORES

SECRETARIA

DRA. YOLANDA MAGDA TORRES FALCON