UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

DOCTORADO EN FILOSOFÍA DE LA CIENCIA,
FILOSOFÍA DE LAS MATEMÁTICAS Y LÓGICA DE LAS CIENCIAS

LÓGICA, CREENCIAS Y NORMATIVIDAD: EL PUENTE ENTRE LA LÓGICA EPISTÉMICA Y LA FILOSOFÍA

TESIS
QUE PARA OPTAR POR EL GRADO DE
DOCTOR EN FILOSOFÍA DE LA CIENCIA
PRESENTA

ROSA MARÍA ESPINOZA CORONEL

Tutor principal:
DR. MAX FERNÁNDEZ DE CASTRO (Universidad Autónoma Metropolitana-Iztapalapa)

Miembros del Comité Tutor:
Dra. Atocha Aliseda Llera (IIF-UNAM)
Dr. Francisco Hernández Quiroz (Facultad de Ciencias UNAM)
Dr. Fernando Velázquez Quesada (University of Bergen)
Dra. Karen González Fernández (Universidad Panamericana)

CIUDAD DE MÉXICO, MAYO, 2024

VNIVERSIDAD NACIONAL AVTONOMA DE MEXICO

UNAM POSGRADO
Filosofía de la Ciencia

PROPUESTA UNIVERSITARIA DE INTEGRIDAD Y HONESTIDAD ACADÉMICA Y PROFESIONAL
(Graduación con trabajo escrito)

De conformidad con lo dispuesto en los artículos 87, fracción V, del Estatuto General, 68, primer párrafo, del Reglamento General de Estudios Universitarios y 26, fracción 1, y 35 del Reglamento General de Exámenes, me comprometo en todo tiempo a honrar a la Institución y a cumplir con los principios establecidos en el Código de Ética de la Universidad Nacional Autónoma de México, especialmente con los de integridad y honestidad académica.

De acuerdo con lo anterior, manifiesto que el trabajo escrito titulado "Lógica, creencias y normatividad: el puente entre la Lógica Epistémica y la Filosofía" que presenté para obtener el grado de **Doctorado** es original, de mi autoría y lo realicé con el rigor metodológico exigido por mi programa de posgrado, citando las fuentes de ideas, textos, imágenes, gráficos u otro tipo de obras empleadas para su desarrollo.

En consecuencia, acepto que la falta de cumplimiento de las disposiciones reglamentarias y normativas de la Universidad, en particular las ya referidas en el Código de Ética, llevará a la nulidad de los actos de carácter académico administrativo del proceso de graduación.

Atentamente

RMEC

Rosa María Espinoza Coronel
No. Cuenta: 514006422

Índice general

Capítulo 1

Introducción: Lógica, Omnisciencia y Normatividad.

1.1. Preambulo

En este trabajo abordaré el problema de la omnisciencia lógica desde un punto de vista filosófico en sentido amplio. Con esto me refiero a una labor de esclarecimiento conceptual y de búsqueda de conexiones con temas que pueden aportar información relevante a la elucidación de una cuestión, pero que no han sido tomados en cuenta en el problema central de esta tesis, -el de la omnisciencia lógica. Parte de la originalidad a la que aspira esta investigación es a establecer un puente entre ciertos temas que corresponden a la filosofía de la lógica y un problema que se ha tratado exclusivamente de manera formal. Para ello se requiere de ciertos presupuestos que a continuación expongo.

La relación entre la lógica y el razonamiento humano es un supuesto que vio la luz desde tiempos aristotélicos y como tal ha prevalecido como un elemento histórico distintivo de la lógica como disciplina filosófica. Se afirmaba que las leyes de la lógica eran las leyes del pensamiento. Una forma de plasmar esta idea consiste en señalar que el ser humano en condiciones adecuadas razona de acuerdo con los principios y las normas lógicas (Stein 2011). Es decir, la lógica describiría la competencia razonante del ser humano, su habilidad para hacer inferencias cuando no se encuentra perturbado por elementos circunstanciales que le impiden ejercerla. Sin embargo, un número importante de filósofos han puesto en duda dicha tesis inicial básicamente por dos razones. Primero, porque las leyes de la lógica no son explícitas del razonamiento, es decir, no aluden en su contenido a creencias ni a razonamientos (Harman 1986). Segundo, por el auge que han tomado los experimentos mentales en psicología cognitiva en los que se muestra que el ser humano no siempre razona de acuerdo con principios lógicos, sino que presenta de manera recurrente errores de razonamiento (Stein 2011). Quienes se posan en el primer cuadro, es decir, aquellos que defienden la influencia de la lógica en el razonamiento apelan a la tesis de la normatividad lógica (MacFarlane, 2004. Steinberger, 2015). Dicha tesis afirma que la lógica no representa propiamente cómo razona el ser humano sino cómo debería razonar y cómo debería formar sus creencias de tal manera que la verdad se preserve. Esta afirmación disuelve en parte la discordancia que hay con los mencionados experimentos de la psicología cognitiva. Pero queda pendiente la objeción antes mencionada de que las verdades lógicas no son principios de razonamiento. Al respecto han surgido propuestas muy importantes que intentan conectar formalmente a los principios lógicos con normas de razonamiento.

Una de las aportaciones de esta investigación será sostener que los desacuerdos mencionados en el párrafo anterior impactan en el campo de la lógica epistémica. Una manera de enmarcar esta cuestión es mediante el problema que me propongo explorar, a saber el de la omnisciencia lógica. Este problema consiste en que en la representación formal del conocimiento se le atribuye al agente el conocimiento de todas las consecuencias lógicas de aquello que sabe, lo que es imposible para un ser humano real. Esta atribución se da dentro de los primeros sistemas de lógica epistémica creados por Hintikka en los años sesentas del siglo pasado. Otra manera de formular el problema es señalando que el sistema estándar que formaliza el conocimiento modela un agente humano capaz de conocer todas las consecuencias lógicas de aquello que sabe o cree. Quienes consideraron que esta propiedad era grave en un sistema que pretendía representar la manera en que los seres humanos formamos (o debemos formar) nuestras creencias apelaron a que la lógica epistémica debía respetar las capacidades cognitivas del agente humano y tomar en cuenta sus limitaciones de memoria y capacidad inferencial.

En un principio esta investigación se concentraba únicamente en destacar algunos elementos que podrían desarrollarse filosóficamente a partir del problema de la omnisciencia lógica. Por ejemplo, aceptar o no que la omnisciencia es un tipo de idealización necesaria para el desarrollo de la lógica epistémica, así como analizar las distintas nociones de creencia que tal vez no estén correctamente formalizadas en los sistemas de lógica doxástica. Sin embargo, algo que llamó mi atención es un diagnóstico en el que coinciden los autores que intentan formular sistemas lógicos no omniscientes, a saber, que los sistemas normales no toman en cuenta que las habilidades cognitivas del agente son límitadas y que, por ende, este no puede saber todas las consecuencias lógicas de sus creencias. Pero, ¿por qué Hintikka tendría que contemplar las limitaciones cognitivas del agente humano? ¿Es tarea de la lógica formalizar el razonamiento humano? Recientemente el estudio de la lógica se ha extendido a campos en los que no necesariamente está involucrado un retrato fiel del razonamiento humano, por ejemplo, en la inteligencia artificial. No obstante hay un aspecto histórico y argumentativo que respalda las objeciones en contra de Hintikka. Mi interés es mostrar un poco la parte argumentativa a favor y en contra de la normatividad de la lógica, sus resultados y su vínculo con la omnisciencia lógica.

Ahora bien, he formulado la dificultad en términos de conocimiento, pero se plantea de manera similar para la creencia y es así como yo la consideraré en este trabajo (con muy pocas excepciones)[1]. La razón es que tradicionalmente el conocimiento es definido como una creencia con ciertas características: verdad y justificación, por lo pronto. Estas complican el problema de la omnisciencia y podrían apartarme del foco de mi investigación. De esta manera, al limitarme a la creencia simplifico el tratamiento. Ahora bien, eso no significa que una solución al problema de la omnisciencia en el caso de la creencia bastaría para solventarlo en el caso del conocimiento. Tal vez, para este concepto más complejo otras dificultades se presenten. La salvedad mencionada se refiere a sistemas formales diseñados precisamente para modelar el conocimiento.

1.2. Preguntas de Investigación

Las tres preguntas centrales que guían esta investigación son:

a) ¿Qué relación guarda la lógica con el razonamiento humano?

b) ¿Cuál es el vínculo entre un problema como la omnisciencia lógica y el debate sobre la normatividad de la lógica?

c) ¿Qué debe de entenderse como una solución correcta al problema de la omnisciencia lógica?

Más precisamente, el problema principal al que me enfrento es el planteado en c), pero mi propuesta personal será abordarlo a través de las cuestiones a) y b).

[1]Por ello estrictamente yo no debería hablar de “omnisciencia”, pero conservo el mote por comodidad.

Empezaré formulando el problema de la omnisciencia lógica y esbozaré algunas distinciones conceptuales con las que pienso abordarlo, lo cual me permitirá a continuación establecer con mayor precisión mis objetivos.

1.3. El Contexto de la Omnisciencia Lógica.

El problema de la omnisciencia lógica se originó con los primeros sistemas formales de lógica epistémica, los cuales fueron formulados por Jaakko Hintikka en 1962. Los sistemas se proponen modelar los conocimientos o las creencias de un agente. Para ello, Hintikka introduce en el lenguaje de la lógica clásica un operador monádico $\mathcal{K}$ (o $\mathcal{B}$ si se trata de la creencia) cuyo comportamiento está regido por una serie de reglas sintácticas (en la versión axiomática) o semánticas (en la presentación de teoría de modelos). Si P es una proposición, $\mathcal{KP}$ es otra y puede interpretarse como que el agente sabe que P (y $\mathcal{B}P$ como que el agente cree que P). ¿Qué propiedades debe tener este operador? ¿Cómo debe comportarse? Una manera de ilustrar mejor la función de dichos operadores es revisando algunos casos en que pueden ser útiles, veamos:

1.3.1. Un caso Gettier

Revisemos uno de los conocidos ejemplos de Gettier (1963). Antes recordemos que algo que aportan dichos casos es un contrajemplo a la noción platónica de conocimiento. Más específicamente muestran que una creencia verdadera y justificada no necesariamente es conocimiento. Supongamos que dos individuos, Smith y Jones, están solicitando el mismo trabajo y que Smith tiene evidencia de que la siguiente proposición es verdadera:

(P) Jones es el hombre que obtendrá el trabajo y tiene 10 monedas en su bolsa.

La evidencia que tiene Smith para creer P se basa en que el presidente de la compañía le aseguró que Jones sería el elegido y en que, hace unos minutos, notó que Jones tenía 10 monedas en su bolsa. P implica a la siguiente proposición:

(Q) El hombre que obtendrá el trabajo tiene 10 monedas en su bolsa.

Smith hace la inferencia de P a Q, y acepta Q por la evidencia que obtuvo de P. Note que la inferencia lógica jugó dos papeles. Por un lado, condujo a

Smith a creer Q. Por otro aportó una justificación a la conclusión a partir de la premisa. Por lo tanto, Smith está justificado en creer que Q es verdadera. Ahora imaginemos que Smith es quien finalmente obtiene el trabajo. Además, sin haberlo notado antes, Smith también tiene 10 monedas en su bolsa. Por consiguiente, la proposición Q es verdadera y la proposición P es falsa, a pesar de que inferimos Q de P. Notemos que el siguiente conjunto de afirmaciones es verdadero:

1) Q es verdadera.

2) Smith cree que Q es verdadera.

3) Smith está justificado en creer que Q es verdadera.

Pero también es claro que Smith no sabe que Q es verdadera, porque esta proposición es verdadera en virtud del número de monedas que tiene Smith en su bolsa. Pero él basa su creencia en Q en el número de monedas que tiene Jones en su bolsa y en que cree erróneamente que es su rival el que obtendrá el trabajo.

Ahora bien, deseamos que el operador $\mathcal{B}$ pueda modelar esta situación. Como vimos, Smith pasó de sus creencia de que Jones era la persona que obtendría el empleo y de que traía diez monedas en el bolsillo (P), a la creencia de que la persona que obtendría el empleo tenía diez monedas en el bolsillo (2). ¿Cómo llegó a esta creencia? Porque la dedujo de la anterior. Si de nuevo utilizamos el símbolo $\mathcal{B}$ para simbolizar “Smith cree que...” podríamos formalizar esta situación de la siguiente manera:

1). $\mathcal{B}[OE(j) \wedge (\forall x)(OE(x) \rightarrow x = j) \wedge 10M(j)]$

2). $\mathcal{B}(\exists y)[(OE(y) \wedge (\forall x)(OE(x) \rightarrow x = y) \wedge 10M(y)]$

Es decir, suponemos que Smith hizo la inferencia correcta que lleva de 3) a 4):

3). $OE(j) \wedge (\forall x)(OE(x) \rightarrow x = j) \wedge 10M(j)$

4). $(\exists y)[(OE(y) \wedge (\forall x)(OE(x) \rightarrow x = y) \wedge 10M(y)]$

Parece natural pedir que si una proposición, Q, es consecuencia de otra, P, y el individuo cree la segunda, también podamos atribuirle la creencia en la primera, aún si el sujeto no “sabe” (al menos conscientemente) de este hecho lógico. Sin embargo, este primer ejemplo, nos advierte de una simplificación que, como luego veremos, Hintikka hace. Que Smith creyera P no lo llevó inmediatamente a su creencia en Q. Supone Gettier que Smith hizo esta inferencia y que solo entonces creyó Q. Pudo ocurrir que esta inferencia fuese inconsciente y casi inmediata. En otros casos, sin embargo, el agente solo llegará a la conclusión si hace una deducción consciente lo que, por supuesto, toma algún tiempo. Este factor temporal no aparece en los sistemas de Hintikka y puede pasar como una suerte de idealización. El siguiente ejemplo que tomo de Raymond Smullyan (1988) me servirá para destacar este aspecto y otro que será muy importante a lo largo de este trabajo.

1.3.2. Caballeros y Bribones

Vayamos a la isla de los caballero y los bribones de Smullyan donde todo habitante es caballero o bribón, pero no ambas cosas. Los caballeros siempre dicen la verdad y los bribones siempre mienten. Supongamos que un visitante llega a la isla y un habitante le dice: usted nunca creerá que soy un caballero. El visitante es del tipo presumido es decir, cree que si cree una proposición, entonces esta es verdadera. Ahora el visitante razona así: si él es un bribón, entonces mintió. Por lo tanto, alguna vez creeré que es un caballero. Si creeré que es un caballero, entonces efectivamente es un caballero (el visitante es presumido). Nosotros podemos ahora seguir razonando: el visitante cree que su interlocutor es un caballero y, por ende, este mintió. En realidad, es un bribón. El visitante cree algo falso.

Por supuesto que esto ocurrirá solo si el visitante sabe las reglas de la isla, escucha lo que se le dice, está dispuesto a inferir y tiene cierta capacidad lógica. Específicamente, supongamos que si Q es consecuencia lógica de P, entonces si el individuo cree P, creerá Q. Representaré por $\mathcal{B}\alpha$ el hecho de que el individuo cree una proposición α. Es decir, lo que aparece a la derecha de $\mathcal{B}$ representa los pensamientos del visitante. Smullyan propone representar el hecho de que un habitante de la isla dice P como $(K \leftrightarrow P)$ donde K significa que el habitante es un caballero (y $\neg K$ que es un bribón). Entonces pueden formalizarse así los razonamientos anteriores:

1. $\mathcal{B}(K \leftrightarrow \neg\mathcal{B}K)$
2. $\mathcal{B}(\neg K \rightarrow \mathcal{B}K)$
3. $\mathcal{B}(\mathcal{B}K \rightarrow K)$
4. $\mathcal{B}(\neg K \rightarrow K)$

5. $\mathcal{B}K$

6. $K \leftrightarrow \neg\mathcal{B}K$

7. $\neg K$

8. $(\neg K \wedge \mathcal{B}K)$

1) simboliza que el visitante cree en las reglas de la isla y escucha a su interlocutor. Los pasos de 2) a 5) representan las inferencias del agente fundadas en la capacidad lógica que le otorgamos (excepto (3) que se debe a su caracter presumido). En cambio, la continuación de la prueba formaliza el razonamiento que hacemos. El visitante no puede concluir que cree K con las capacidades que le hemos concedido. El agente puede razonar como lo hacemos nosotros, excepto que no puede pasar de "...por lo tanto es un caballero" a "ahora yo creo que es un caballero". Aunque razona de acuerdo con el cálculo proposicional clásico, esta última inferencia le está vedada. Con ello quiero enfatizar que la lógica que emplea o conoce el agente no tiene por qué ser aquella que empleamos nosotros para razonar sobre sus creencias. Incluso podría ocurrir que su lenguaje no coincidiese con el nuestro (como sucede en algunos sistemas formales que más tarde estudiaremos). También veremos que en el sistema $S5$ de Hintikka, que es muy frecuentemente usado para la representación del conocimiento, el agente y el observador tienen exactamente la misma lógica.

1.3.3. Aquiles y la tortuga.

Recordemos el ejemplo de Lewis Carroll (1895) de Aquiles y la tortuga: Aquiles intenta convencer a la tortuga de un teorema de Euclides, exhibiendo la demostración. Al final se llega a que tenemos dos líneas iguales a un tercera, y una de las nociones comunes de Euclides es que dos cosas iguales a una tercera son iguales entre sí. De allí la tortuga debe concluir que las dos líneas iniciales son iguales, pero se resiste a ello. Aquiles insiste en que debe hacerlo así porque es una ley de la lógica, a lo que la tortuga responde que entonces debe ser agregada como premisa. Es decir, si $\mathcal{B}P$ significa que la tortuga cree (o acepta) la proposición P, podemos esquematizar la situación de la siguiente manera:

- $(P1) \quad \mathcal{B}(a = b)$
- $(P2) \quad \mathcal{B}(b = c)$

Lo que me importa subrayar es que parece de lo más natural concluir de allí que:

- $(C) \quad \mathcal{B}(a = c)$

Sin necesidad de invocar la noción común de Euclides pues parece que, como en el ejemplo anterior, podemos atribuir a la tortuga la creencia muy inmediata de dos premisas que cree (sobre todo cuando considera explícitamente esa conclusión). Sin embargo, también tenemos como premisa que dos cosas iguales a una tercera son iguales entre sí y la tortuga la acepta.

- $(P3) \quad \mathcal{B}[(\forall x)(\forall y)(\forall z)(x = z) \wedge (y = z) \to (x = y)]$.

Intuitivamente cualquier persona que crea P1), P2) y P3) debería creer C), pero la tortuga insiste en que esto solo se da por una ley de la lógica que debe admitirse antes, es decir, agregarse como premisa. Pero una vez que Aquiles accede a este movimiento, cae en la trampa porque la tortuga repite el procedimiento y pide, como condición para aceptar la conclusión, que se agrege como nueva premisa:

$$[(a = c) \wedge (b = c) \wedge [(\forall x)(\forall y)(\forall z)((x = z) \wedge (y = z)) \to (x = y)] \to (a = c)$$

que asimismo es una verdad de la lógica. Así el argumento se alarga indefinidamente. Note que si le atribuimos a la tortuga la creencia en las consecuencias lógicas de lo que cree, y nos referimos a la lógica clásica, entonces P3) puede siempre suponerse. Parece natural postular que el operador $\mathcal{B}$ pueda concatenarse a una verdad de la lógica clásica para obtener una verdad de la lógica epistémica. Así podemos suponer que los agentes conocen las leyes de la lógica, al menos implícitamente. Sin embargo, este ejemplo me sirve para ilustar un aspecto que será relevante en lo que sigue: la conducta de la tortuga parece reprobable. Parece que estaría obligada a aceptar la conclusión desde que admite las dos primeras premisas. Allí se esboza una forma de normatividad asociada a la lógica.

1.3.4. El Teeteto

En el *Teeteto* de Platón, Sócrates pregunta a Teeteto sobre la naturaleza del conocimiento. Este responde, en una primera etapa, que el conocimiento es percepción. Si esto es así, razona Sócrates, entonces nadie necesita de un maestro como Protágoras, pues todos tenemos sensaciones igualmente válidas, lo mismo una persona reputada como sabia que un demente que alucina. Sócrates da otros argumentos de mayor peso, pero me basta con este para hacer un par de observaciones. El primer punto que deseo enfatizar es que la lógica no puede recomendarle a alguien que crea un absurdo aún si este es consecuencia de las creencias que defiende. Por ende, si la lógica es prescriptiva habrá que ver en qué sentido lo es. La segunda es que, al parecer, Teeteto creyó en algún momento que el conocimiento es percepción y también que se requieren maestros como Protágoras. Pero estas dos creencias tienen consecuencias contradictorias (dados ciertos supuestos de sentido común). ¿Podemos decir que todas sus creencias formaban un conjunto inconsistente? No es que haya creído directamente una proposición y su negación, pero sí dos proposiciones cuyas consecuencias lógicas conducen fácilmente[2] a una contradicción explícita.

Para plantear esta situación con mayor claridad, simbolizaré enseguida el primer argumento referido, para lo cual me basaré en la siguiente correspondencia:

$C =$ el conocimiento es percepción.

$A =$ las percepciones de un individuo son tan auténticas como las de otro.

$V =$ todos tenemos conocimientos igualmente válidos.

$T =$ Teeteto debe consultar a Protágoras.

$P =$ Protágoras tiene conocimientos más válidos que los de Teeteto.

Entonces si empleamos el operador $\mathcal{B}$ para representar las creencias de Teeteto su razonamiento puede esquematizarse así:

1). $\mathcal{B}C$

2). $\mathcal{B}A$

3). $\mathcal{B}(C \to (A \to V))$

4). $(\mathcal{B}C \to (\mathcal{B}A \to \mathcal{B}V))$

5). $\mathcal{B}V$

6). $\mathcal{B}(T \to P)$

7). $(\mathcal{B}T \to \mathcal{B}P)$

8). $\mathcal{B}(V \to \neg P)$

[2]Por supuesto que este adverbio es vago. Se refiere a la intuición según la cual ciertas consecuencias lógicas de algunas premisas son casi inmediatas, mientras a otra se llega después de un tortuoso camino inferencial, pero no pretendo tener una definición precisa.

9). $(\mathcal{B}V \rightarrow \mathcal{B}\neg P)$

10). $\mathcal{B}T$ (Teeteto cree que debe consultar a Protágoras)

11). $\mathcal{B}P$

12). $\mathcal{B}\neg P$.

No podría objetarse que Teeteto ha formulado una hipótesis (C), misma que debe retirar porque no ha resistido a la crítica. La sostuvo como una creencia suya. Otro ejemplo en que la inconsistencia salta más a la vista, aparece en un libro de Peter Singer donde menciona que cuando un comité de Harvard decidió que cualquier paciente cuyo cerebro haya dejado de funcionar está muerto, justificó su decisión diciendo que mantener a un individuo que está en coma irreversible conectado a un respirador era una carga para el mismo paciente. Comenta Singer "las personas muertas no están en coma, están muertas y nada puede ser una carga para ellas" (p. 45). Por supuesto, podría dar muchos otros ejemplos para mostrar que parece sensato decir que un sujeto puede mantener creencias contradictorias. Si consideramos, como yo lo haré, que la lógica es prescriptiva y que recomendará evitar en lo posible mantener un conjunto inconsistente de creencias, debe aceptar la posibilidad de que alguien crea una contradicción o tenga dos creencias que al cabo de breves inferencias generen una contradicción explícita. La lógica epistémica clásica no es útil, sea porque implica que el creyente de una contradicción cree (o debería creer) cualquier cosa, o bien (en sistemas que tienen el esquema D como teorema) porque no admite que el sujeto tenga una creencia contradictoria. En la lógica de Hintikka 11) y 12) permitirían inferir $\mathcal{B}P \wedge \neg P$ y $\mathcal{B}R$ para cualquier proposición R y, por lo tanto, también $\mathcal{B}(T \wedge \neg T)$.

Esta cuestión de si podemos realmente afirmar que un sujeto tiene creencias contradictorias y cómo debe tratarse es arduo y no lo analizaré con la profundidad debida. Sin embargo, sí observaré cómo diversos sistemas de lógica doxástica lo tratan. En general, como ya dije, soy partidaria de sistemas con una cierta flexibilidad que contemplen la posibilidad de que un sujeto caiga en una contradicción específica, pero que le recomienden salir de esa situación.

Los ejemplos anteriores pretenden exhibir la práctica de los operadores epistémicos, aunque igualmente nos dan un indicio de cómo se motiva el problema de la omnisciencia lógica pues en los tres primeros supusimos que si el agente cree una proposición P y de ella se sigue Q, entonces puede atribuírsele, o incluso debe exigírsele, la creencia en Q. En el cuarto, en cambio, el individuo aparentemente está obligado a revisar sus creencias anteriores. En cualquier caso parece que la lógica genera ciertos compromisos a un creyente o lo autoriza a tener ciertas actitudes. Pero si esto vale para toda consecuencia lógica, el agente debe ser capaz de reconocer como tal cualquier consecuencia lógica de lo que cree (hecha la salvedad del tiempo) y, en particular, todas las verdades de la lógica, lo que, sin duda, es excesivo para un ser humano estándar. Veremos que los sistemas de Hintikka efectivamente hacen al agente modelado lógicamente omnisciente (u omnicreyente), lo que es muy poco realista. Tal es el problema de la omnisciencia lógica. Sin embargo, cómo diré enseguida, esta formulación es insuficiente.

Otra función de los ejemplos anteriores es que me permiten hacer dos observaciones que desarrollaré a lo largo de este trabajo de investigación:

(1) En un sistema que pretende modelar o normar las creencias de un agente, es importante distinguir entre la lógica atribuida al agente y la que se emplea en el sistema mismo. En los casos anteriores concedimos al agente la misma capacidad inferencial que la que se emplea en la descripción de sus creencias, con excepción del ejemplo de Smullyan, en que el visitante de la isla, aunque usa la lógica clásica, no puede hacer ciertas deducciones concernientes a sus propias creencias. Ello me permite adoptar una actitud neutra y tolerante en el debate entre monismo y pluralismo lógico, discusión que rebasa los límites de mi investigación. Incluso adoptando la primera postura puede ser deseable evaluar el desempeño de un sujeto en términos de lo que este considera como una lógica correcta.

(2) El segundo elemento sobre el que quiero llamar la atención es que en la aplicación de las reglas de la lógica a estos casos parecen estar involucrados implíctamente ciertos operadores deónticos que modulan a los operadores epistémicos. Para ilustar este punto, considere que, en el primer ejemplo, Smith *tiene razones para* creer Q, mismas que provienen de P, vía la inferencia lógica. En el segundo, el visitante *puede* creer (en el sentido de estar autorizado) que el habitante es un caballero, aunque no está obligado a ello. En el tercero, parece irracional la conducta de la tortuga. En principio, *debería* aceptar la conclusión aún sin P3) (de hecho, aún sin P2)). En el último, obviamente Teeteto está obligado a creer las conclusiones a que su interlocutor lo lleva y que parten de premisas que acepta o bien a renunciar a estas últimas.

Trataré en detalle estas cuestiones en el primer capítulo.

1.3.5. La importancia de replantear la Omnisciencia Lógica

Para resolver el problema de la omnisciencia lógica muchos sistemas formales de lógica epistémica (sobre todo proposicional) han sido propuestos en las últimas décadas. Sus autores han pretendido solucionarlo o, por lo menos, reducir su dificultad. Ahora bien, si la solución de un autor es adecuada ¿por qué es seguida por las de otros autores? ¿Persiguen los mismos fines? ¿Hay algún criterio para decir que una es mejor que la otra? Examinando sus artículos puede verse que, si bien es relativamente fácil plantear el problema de la omnisciencia lógica (como yo misma acabo de hacerlo), hay un sentido en que esa formulación es insuficiente. Una de mis hipótesis de trabajo supone que el problema no está acotado en su totalidad, de allí la necesidad de usar un planteamiento distinto al de los sistemas lógicos. No sabemos con precisión qué debe contar como una solución del mismo. Para hacerlo patente voy a citar a algunos de ellos para mostrar cuán poco han delimitado el problema y cómo, por eso, es difícil saber si alcanzan el objetivo que se proponen. Además hay divergencia entre los fines que se persiguen en los diferentes sistemas. Comencemos por Levesque, quien formuló uno de los primeros sistemas para solucionar el problema de la omnisciencia lógica.

Levesque

En 1984 Levesque presenta el artículo "A Logic of Implicit and Explicit Belief". Una vez que el problema es expuesto de manera escueta, dice que algunas manifestaciones del problema de la omnisciencia lógica "podrían causar que uno rechazara una formalización en términos de mundos posibles como poco intuitiva, en el mejor de los casos, y como completamente irrealista, en el peor." A su juicio, Hintikka no modela lo que un agente cree sino cómo sería el mundo si lo que cree fuera verdadero. Agrega:

> "¿Cuál es entonces una semántica apropiada para tratar con las creencias reales de un agente? (...) Un modo de evitar todas esas tautologías es hacer la noción de cómo un agente piensa que es el mundo más relevante a lo que realmente cree". (Levesque, 1984, pag.198-199)

Hacia el final apunta:

> "¿Qué logramos con esta nueva lógica? una cosa es un lenguaje que puede ser usado para razonar formalmente acerca de las creencias de otros agentes sin asumir omnisciencia lógica". (Levesque 1984, pag. 201)

Evidentemente esperamos que el sistema no atribuya al agente la creencia en todas las verdades de la lógica o en todas las consecuencias lógicas de sus propias creencias. Eso es esencial, pero Levesque espera que su sistema represente más fielmente las creencias reales de un agente. Pero, ¿qué significa esto? ¿De un agente específico, o del sujeto normal? Parece entonces que la lógica sería un capítulo de la psicología o debería recurrir a esta disciplina en busca de un criterio de adecuación. No obstante estas citas ponen de relieve dos puntos comunes en la discusión que señalaré brevemente. Primero, la importancia de que los sistemas de lógica epistémica tomen en cuenta las limitaciones cognitivas de los seres humanos, específicamente porque dichas limitaciones impiden que los agentes crean todas las consecuencias lógicas de sus creencias. En segundo lugar, su solución al problema apunta hacia un modelo de creencia más adecuado al agente humano.

Fagin y Halpern

Tomemos otro ejemplo. Ronald Fagin y Joseph Halpern (1985) dicen respecto a la omnisciencia lógica:

> "Desafortunadamente en la vida real las personas ciertamente no son omniscientes. En verdad, los defensores de los mundos posibles han siempre subrayado que este estilo de semántica supone un razonador "ideal" con poderes computacionales infinitos. Sin embargo, para muchas aplicaciones, a uno le gustaría una lógica que proporcionara una representación más realista del ser humano". (Fagin y Halpern 1985. p. 40)

Más adelante consideraré la noción de ideal aquí mencionada y mostraré que el adverbio "desafortunadamente" no puede ser tomado en serio. Me interesa ahora observar que Fagin y Halpern tampoco son precisos en sus objetivos.

Ho Ngoc Duc

Revisemos brevemente la propuesta de Ho Ngoc Duc quien en 1997 en el artículo: *Reasoning about Rational, but not Logically Omniscient Agents.* propone algunos componentes novedosos para resolver el problema. Duc comienza criticando algunos tratamientos previos:

> "Casi todos los intentos en la literatura para resolver el problema consisten en debilitar el sistema epistémico estándar: sistemas más débiles son considerados, en donde el agente no posee las capacidades completas de razonamiento de un razonador ideal... Nosotros argumentaremos que esta solución no es satisfactoria: de esa manera la

> omnisciencia lógica puede ser evitada, pero muchas intuiciones acerca del conocimiento y la creencia se pierden". (Duc, 1997, pag. 1)

Tampoco está de acuerdo con otros enfoques que más adelante repasaremos:

> "Aunque el modelo de deducción, el enfoque de mundos imposibles y el enfoque del operador de conciencia solucionan LOP [el problema de la omnisciencia lógica] técnicamente, estos no pueden considerarse satisfactorios". (Duc 1997, p. 5)

Es decir, puede considerarse que un sistema soluciona "técnicamente" el problema si las creencias del agente que modelan no son cerradas bajo consecuencia lógica, pero no basta eso para dar una solución satisfactoria. ¿Cómo debemos entender este requirimiento? Para Duc la lógica debe modelar cómo el individuo adquiere nuevas creencias por medio de inferencias concretas. En su solución entrará el tiempo como una variable importante. Como vemos sus objetivos son muy diferentes de los de sus predecesores. Sin embargo, aún continúan siendo vagos.

Parikh

Mi posición al respecto concuerda en gran medida con la de Parikh quien en *Sentences, Belief and Logical Omniscience, or What Does Deduction tell us?*, artículo de 2008, señala un defecto en los planteamientos anteriores del problema:

> "Ha habido sugerencias de que estos "problemas" pueden tratarse usando herramientas como la conciencia, mundos posibles imposibles, o haciendo referencia a la complejidad computacional. Desde nuestro punto de vista estos métodos no señalan el problema fundamental, a saber, ¿qué estamos tratando de modelar? Para tener una teoría del conocimiento, debemos tener algún criterio para medir lo que alguien sabe. Pues comenzar con una lógica que produce omnisciencia lógica y después complicarlo con varios métodos, hace a un lado la pregunta de cuál es nuestro objetivo. A menos que tengamos una visión clara de nuestro objetivo no es probable que lo alcancemos." (Parikh, 2008, pag. 461)

Concuerdo completamente con Parikh. La pregunta es ¿qué estamos tratando de modelar? Solo, que a diferencia suya, yo trataré el problema, no para el conocimiento, sino solo para la creencia, siguiendo la estrategia cartesiana de dividir un problema en sus partes más simples. Aunque concedo que hay una sencilla explicación de qué está mal en un sistema que concede al agente omnisciencia lógica, sostengo que muchos términos implícitamente involucrados en la formulación de esta cuestión deben ser esclarecidos. Para dar un ejemplo de una empresa similar en que los términos son aclarados con antelación a que una solución sea ofrecida, considere el artículo clásico en que Tarski define la "verdad" para cierto tipo de lenguajes. Señala primeramente cuál es la noción vaga de "verdad" del lenguaje ordinario que intenta explicar. En segundo lugar muestra cómo no es posible definir un concepto preciso que corresponda totalmente a esa noción, pues esta es contradictoria. Luego ofrece un análisis de qué rasgos del lenguaje coloquial generan esas antinomias para sacar de allí la hipótesis de que, para cierto tipo de lenguajes formales, será posible dar dicha definición. Pero antes de ofrecer esta, Tarski determina con toda precisión cuáles serán los indicios de que su labor ha alcanzado la meta (a saber, cuando de la definición en cuestión y de ciertas verdades elementales sea posible deducir las convenciones T). No pretendo hacer una tarea semejante, sino solo señalo, con este ejemplo, el tipo de labor conceptual que no se ha hecho en la vasta literatura sobre omnisciencia lógica. Solo algunos autores dan algunas indicaciones de qué consideran que debe ser una solución a este problema. Yo me propongo avanzar en esta dirección. Hay otro punto en que la analogía anterior me es útil. Tarski dice que intenta recuperar la noción de verdad que queda indicada en la llamada teoría de la verdad como correspondencia, para ejemplificar la cual se remite a Aristóteles. Sin embargo, no hay duda de que alguien podría ofrecer otro concepto que intentara recuperar un aspecto distinto del uso ordinario de "verdadero". De la misma forma, yo defenderé una forma de dar mayor precisión al problema de la omnisciencia lógica. No niego que pudiera haber otras formas de entenderlo.

1.3.6. Objetivos de la investigación.

La motivación más importante de esta investigación es realizar un examen minucioso del problema de la omnisciencia lógica para mostrar que, subyacientes a esta problemática, se ocultan importantes cuestiones filosóficas. El objetivo general es mostrar la pertinencia de estas cuestiones que, aunque debatidas recientemente, no han sido consideradas, al menos de manera explícita, en la construcción de sistemas de lógica epistémica. Una vez revisados algunas de estas controversias pretendo mostrar cómo pueden incidir en la evaluación de sistemas que pretenden resolver el problema de la omnisciencia lógica. Para ello tomaré algunos ejemplos paradigmáticos de estos sistemas y los examinaré a la luz de las consideraciones previas.

He dicho que algunos conceptos y cuestiones filosóficas subyacen a un tratamiento de la omnisciencia lógica. Enumero enseguida algunos de ellos que me propongo analizar y mostrar su relevancia para el tema de la omnisciencia lógica.

1)Idealización: Diversos autores sostienen que la omnisciencia lógica es una idealización excesiva de las capacidades humanas cognitivas. No obstante, también hay autores que justifican esta idealización como una herramienta que se emplea en otras disciplinas científicas. Podemos emplearla para modelar lo esencial de un fenómeno y simplificar una teoría eliminando factores que más tarde podrán tomarse en cuenta. Como tal, la idealización también se usa con un objetivo pragmático para realizar predicciones. En estos casos podemos idealizar para describir, ya sea una parte o el todo de un fenómeno, y para luego comprobar si coincide con la descripción. Sin embargo, lo que es excesivo como una descripción de una capacidad inferencial puede ser justo como una norma de razonamiento. ¿Se trata de modelar la competencia inferencial de un sujeto o de normarla? Mi interés en la idealización consiste en indagar si realmente los sistemas de lógica epistémica clásicos están justificados en emplearla al momento de modelar o bien normar la creencia y el conocimiento. Sostendré que, aún en el segundo caso, la respuesta es negativa, es decir, consideraré que la idealización presente en los sistemas de lógica epistémica propuestos por Hintikka no se justifica aun tratándose de describir o de normar el razonamiento. Sin embargo, algún grado de idealización es inevitable en cualquier empresa científica y la lógica epistémica no es la excepción.

2) Creencia. El desarrollo de la sección dedicada a este tema tiene dos partes. En la primera revisaré las concepciones actuales de la creencia y cómo es definida en cada caso. Entre estas concepciones están el interpretacionismo y el representacionismo, entre otros. La aportación central de este punto es presentar dos grandes categorías en que se concibe la creencia, a saber, como representación mental o como atribución de actitudes. Se aclararán estos dos sentidos. En la segunda parte expongo las concepciones de cuatro importantes pensadores que ilustran las dos posiciones señaladas.

3) Normatividad Lógica: El tema de la normatividad lógica tiene que ver con la siguiente pregunta: ¿son los principios lógicos normas de razonamiento? Dicho de otra manera, ¿los seres humanos razonamos y formamos nuestras creencias de acuerdo con los principios de la lógica? Ambas preguntas tienen una historia en la filosofía, pues desde Kant e incluso con Aristóteles se concebía a la lógica como la ciencia del pensamiento. No obstante, en tiempos recientes ha surgido una nueva controversia que intenta desacreditar la tesis de que la lógica prescribe cómo debemos razonar. Se ha argumentado que los principios lógicos no contienen normas que nos permitan formar creencias, además de que dichos principios no son principios explícitos de razonamiento. La razón por la cual incluyo el debate sobre la normatividad de la lógica en la investigación tiene que ver con una intuición que se deriva de las críticas al sistema de Hintikka. Pues, en todas estas se señala que Hintikka no respeta las capacidades cognitivas del agente, las cuales son limitadas y no puede concluir todas las consecuencias lógicas de sus creencias. Desde mi perspectiva esta crítica conduce a una pregunta fundamental acerca de cuál es el papel de la lógica, sea epistémica o de otro tipo, en el razonamiento humano. Es decir, en el caso de los sistemas de lógica epistémica o doxástica ¿debemos interpretar fórmulas con el operador K o B como normas o como descripciones de las creencias? En la sección que corresponde a esta discusión se revisarán los argumentos a favor y en contra de la normatividad lógica, así como una propuesta para transformar enunciados lógicos en principios de razonamiento.

A partir de las aclaraciones hechas en el capítulo 2, los siguientes capítulos estarán dedicados a revisar los sistemas lógicos que se han propuesto como solución al problema de la omnisciencia lógica. Dividí esto en dos partes. En el capítulo 3 revisaré sistemas que son atemporales y que no modelan el cambio de creencia. En el capítulo 4 revisaré sistemas que, a diferencia de los anteriores, sí modelan el cambio de creencia y capturan la temporalidad de las deducciones.

1.4. Agradecimientos

Este trabajo no hubiera sido posible sin el acompañamiento de mis sinodales. Agradezco a la Doctora Atocha Aliseda por aceptar ser mi sinodal a mitad del camino, por su solidaridad y comprensión para continuar con este proyecto. Al Doctor Francisco Hernández por las observaciones que desde el principio hizo a esta investigación, en especial por sugerirme revisar los trabajos de sus alumnos Mauricio Andrade y Rodrigo Medina, los cuales me ayudaron a entender ciertas partes de la lógica de la justificación y su respuesta a la omnisciencia lógica. Al Doctor Fernando Velázquez le agradezco que haya compartido conmigo sus notas sobre lógica epistémica dinámica las cuales están presentes en este trabajo y que me fueron de gran utilidad para complementar las hipótesis de este trabajo. A la doctora Karen González Fernández por sus atenta lectura del trabajo y sus amables observaciones. Agradezco al Doctor Max Fernández de Castro por su guía y dirección a largo de varios años y por ayudarme a entender las partes más técnicas de este trabajo.

Asimismo agradezco el acompañamiento y solidaridad de algunas personas que fueron parte importante para concluir este proyecto. A mis hermanos Javier y Juan Pablo, por su ayuda. A José, porque su presencia ha sido mi mejor arma en muchos momentos difíciles. A mi hijo Emiliano, por ser un motivo para no abandonar esta tesis. A mis padres, cuya ausencia me ha enseñado la tristeza permanente.

Esta tesis fue elaborada gracias a una beca CONAHCYT otorgada del año 2016 al 2020.

Capítulo 2

Lógica Epistémica y Normatividad

2.1. Objetivos del capítulo.

En este segundo capítulo me propongo las siguientes tareas:

1) Describiré brevemente los sistemas normales para ver cómo aparece allí la omnisciencia lógica.

2) Analizaré la cuestion de si la omnisciencia lógica puede considerarse una idealización que es adecuada para los sistemas de lógica epistémica. Si así fuera, entonces es un aspecto que debemos tolerar, tal vez en aras de la simplicidad o de una mayor facilidad de cálculo. Argumentaré que no es así en ninguna de las concepciones de la creencia esbozadas.

3) Esbozaré diferentes concepciones de la creencia y, como dije, las dividiré en dos grandes categorías; una que corresponde a la creencia como evento mental privado, y otro según la cual la creencia es un instrumento de una teoría psicológica intuitiva con la cual predecimos la conducta de los otros.

4) Repasaré cuál es la relación entre el razonamiento humano y la lógica. Algunos autores han sostenido o que es obscura esta relación o que no existe. Sostendré el punto de vista más intuitivo y tradicional de acuerdo al cual la lógica no es descriptiva de la competencia razonante, pero sí normativa del cambio de creencias a través del razonamiento. Desde la perspectiva de la creencia como representación, la lógica guía al sujeto solitario en su captación de nueva información, mientras que desde la óptica de la creencia como atribución, la lógica

determina los deberes que contraemos o las autorizaciones que obtenemos a causa de nuestras creencias.

2.2. Lógica de la Creencia y del Conocimiento.

A principios de la década de los sesentas surgieron los primeros sistemas de lógica que pretendieron formalizar nociones epistémicas. Esos sistemas de lógica proposicional utilizaban el aparato semántico creado tanto por Kripke como por Hintikka de los mundos o situaciones posibles. Mientras que Kripke (1963) utilizó la idea de escenarios o mundos posibles para modelar la necesidad metafísica, Hintikka (1962), casi simultáneamente, la diseñó con el fin de representar lo que un individuo concibe en una situación dada. En este caso, que es el que nos concierne, el individuo se encuentra en el mundo actual, pero no lo distingue de una serie de mundos alternativos. Por ejemplo, el individuo no sabe si el dólar bajó ayer, por lo que concibe mundos alternativos en que el dólar descendió y otro en que tal evento no ocurrió, por lo demás siendo idénticos a lo que él piensa que realmente ocurre. No puede saber cuál de esos dos mundos es el real. En cambio, él cree (y, de hecho, sabe) que el PAN ganó la elección presidencial en México en el año 2000 por lo que en todas las alternativas que concibe aparece ese hecho. Es decir, si un individuo cree una proposición P entonces P es verdadera en todos los mundos que concibe. Para que la creencia sea conocimiento habrá, por supuesto, que agregar otras condiciones. Una que aparece con Hintikka es que obviamente para que P sea conocida tiene que ser verdadera en el mundo real. Sin embargo, en este modelo la justificación, tercer componente de la definición tradicional de "conocimiento", no aparece.

Ahora bien, aunque el problema de la omnisciencia lógica se presenta para varias actitudes proposicionales, y en la mayoría de los casos se ha intentado resolver el problema considerando solo al conocimiento, la tarea se facilita si consideramos únicamente a la creencia como un elemento que puede tener un desarrollo teórico propio e independiente del conocimiento. En lo que sigue trataré el problema en su versión doxástica. Dado que en cada mundo posible cada proposición es verdadera o falsa (pero no ambas cosas), podemos considerar a los mundos posibles como conjuntos de proposiciones máximamente consistentes. El agente cree una proposición α en el mundo real w, si α es un elemento de cada mundo concebido por él desde w, es decir, concebido como siendo posiblemente el mundo real.

Como lo hice en la introducción, agreguemos al lenguaje del cálculo proposicional con sus usuales conectivos lógicos, un símbolo $\mathcal{B}$ para representar las creencias de un agente (o $\mathcal{K}$ si se trata del conocimiento), de tal manera que si α es una fórmula del lenguaje, también lo sea $\mathcal{B}\alpha$, la cual simbolizará que la proposición que α expresa es creída por dicho individuo. Es usual introducir también un operador dual de $\mathcal{B}$, a saber, $\hat{\mathcal{B}}$ y que se define como $\hat{\mathcal{B}}\alpha \equiv_{def} \neg\mathcal{B}\neg\alpha$. Lo llamaremos "el diamante". El sistema modal más básico se suele denominar por la letra **K** (a no confundir con el operador de conocimiento) y puede ser presentado semántica o sintácticamente. Empezaré por el aspecto semántico que, aunque no es el originario, es más intuitivo.

Definición 1. Un modelo de Kripke es una terna $M = \langle W, R, V\rangle$ donde:

1. W es un conjunto no-vacío cuyos elementos serán llamados "mundos posibles".
2. $R \subseteq W^2$, es decir, R es una relación binaria de elementos de W.
3. $V : WXLP \Rightarrow \{1, 0\}$, donde LP es el conjunto de letras proposicionales del lenguaje.

En lugar de escribir $(w, v) \in R$ o $R(w, v)$ escribiré wRv. En ese caso se dice v es accesible desde w, que w ve a v o que v es un R-sucesor de w. V es una asignación de valores de verdad (1 representa el verdadero; 0, el falso) a las letras proposicionales en cada mundo. Otra manera de representarlo es concebir a la función V como asociando a cada mundo un conjunto de letras proposicionales (las que son verdaderas en ese mundo) o, aún, como asignando a cada letra proposicional un conjunto de mundos (los mundos en que esa letra proposicional es verdadera). Dado un modelo M y un mundo w, en él las letras proposicionales ya tienen un valor de verdad asignado. La definición de "enunciado verdadero en un mundo w" para enunciados compuestos extiende esta asignación. Cuando se trata de una fórmula cuyo símbolo lógico principal es un conectivo proposicional, la cláusula que determina su evaluación en términos de sus subfórmulas es la usual, solo que, esta vez, la evaluación es local, interna al mundo. Por ejemplo, si α y β son fórmulas del lenguaje, tendremos que $M, w \models (\alpha \wedge \beta)$ si y solo si $M, w \models \alpha$ y $M, w \models \beta$ (donde $M, w \models \alpha$ simboliza que α es verdadera en w). Solo la presencia del operador $\mathcal{B}$ en una fórmula requerirá que para su evaluación busquemos información del valor de sus subfórmulas en otros mundos. Más precisamente, la cláusula correspondiente será:

$$M, w \models \mathcal{B}\alpha \text{ si y sólo si para todo mundo } v \text{ tal que } wRv\ M, v \models \alpha.$$

De aquí se deriva que: $M, w \models \hat{\mathcal{B}}\alpha$ si y sólo si hay un mundo v tal que wRv y $M, v \models \alpha$.

Definición 2. Una fórmula α es verdadera en un modelo M (en símbolos, $M \models \alpha$; o $M \models_{\mathbf{K}} \alpha$, si fuera necesario precisar el sistema), si es verdadera en todos los mundos de ese modelo.

Definición 3. Una fórmula α es válida (en símbolos, $\models \alpha$), si es verdadera en todos los modelos; y es satisfacible, si es verdadera en un mundo de un modelo.

Aquí es muy importante observar que una vez aceptada esta semántica, la omnisciencia lógica aparece inmediatamente. Supongamos que el individuo cree α en un mundo w de un modelo M ($M, w \models \mathcal{B}\alpha$). Entonces para todo mundo v del modelo tal que wRv, $M, v \models \alpha$. Ahora bien, si β es consecuencia lógica de α (lo crea el individuo o no), $M, v \models \beta$. Por lo tanto, $M, w \models \mathcal{B}\beta$, es decir, el agente creerá β; y esto sucederá aún si β no es consecuencia de α, pero el sujeto se equivoca al respecto. Así el conjunto de creencias del agente será cerrado bajo consecuencia lógica, pero también contendrá todas las que él cree consecuencias de sus creencias.

Desde el punto de vista sintáctico $\mathbf{K}$ puede ser caracterizado axiomáticamente de la siguiente manera:
Son Axiomas de $\mathbf{K}$:

1). Todas las fórmulas que sean instancias de tautologías,

2). Las fómulas de la forma $\mathcal{B}(\alpha \to \beta) \to (\mathcal{B}\alpha \to \mathcal{B}\beta)$,

y que tiene como reglas de inferencia:

3). De $\vdash \alpha$ y $\vdash \alpha \to \beta$, se sigue $\vdash \beta$

4). De $\vdash \alpha$ se sigue $\vdash \mathcal{B}\alpha$

Como siempre $\vdash_{\mathbf{K}} \alpha$ (o simplemente $\vdash \alpha$ si no hay riesgo de confusión) significará que α es un teorema de $\mathbf{K}$. Este sistema axiomático es sólido y completo con respecto a la semántica dada, es decir, $\vdash \alpha$ si y solo si $\models \alpha$.

Es muy sencillo probar que si $\alpha_1, \alpha_2 \ldots \alpha_n \models_{CT} \beta$ (es decir, Si β es consecuencia tautológica de $\alpha_1, \alpha_2 \ldots \alpha_n$) entonces $\mathcal{B}\alpha_1, \mathcal{B}\alpha_2 \ldots \mathcal{B}\alpha_n \vdash_{\mathbf{K}} \mathcal{B}\beta$. Es decir, cualquier razonamiento del cálculo proposicional que podamos hacer, también lo puede realizar el agente.

Otras características de la creencia o del conocimiento pueden ser modeladas por sistemas que son extensiones de $\mathbf{K}$ (llamados también "sistemas normales") y que pueden ser obtenidos agregando axiomas a $\mathbf{K}$ o, equivalentemente, imponiendo restricciones a la relación R en la definición de "modelo". Por ejemplo, si el operador epistémico que agregamos al lenguaje proposicional pretende representar el conocimiento, en cuyo caso suele usarse el símbolo $\mathcal{K}$ para representarlo (de nuevo, a no confundir con el nombre del sistema), en lugar de $\mathcal{B}$, es necesario agregar al menos el (esquema de) axioma T, a saber, $\mathcal{K}\alpha \rightarrow \alpha$; o, equivalentemente, la restricción de que la relación R debe ser reflexiva. Así obtenemos el sistema $\mathbf{KT}$ que no es adecuado para la representación de la creencia puesto que un individuo puede creer algo falso.

De manera similar el sistema $\mathbf{KD}$ se obtiene agregando a $\mathbf{T}$ el axioma D, es decir, $\neg\mathcal{K} \perp$ (o $\neg\mathcal{B} \perp$) donde $\perp$ representa una contradicción elegida de antemano (por ejemplo, $(P \wedge \neg P)$) y la condición de que la relación R sea serial, es decir, que para cada mundo w exista otro v (aunque puede ser el mismo w) tal que wRv. Este sistema se emplea más para modelar la creencia. En él la fórmula $(\mathcal{B}\neg\alpha \rightarrow \neg\mathcal{B}\alpha)$ se vuelve válida. Es decir, el agente modelado no tendrá creencias contradictorias. El axioma D introduce una permeabilidad de la que resulta que si el agente tuviese creencias contradictorias, todo el sistema sería inconsistente.

Asimismo $\mathbf{K4}$ resulta de agregar al sistema $\mathbf{K}$ el esquema axiomático 4, a saber, $\mathcal{K}\alpha \rightarrow \mathcal{K}\mathcal{K}\alpha$ (o $\mathcal{B}\alpha \rightarrow \mathcal{B}\mathcal{B}\alpha$) llamado de instrospección positiva y a la relación R la condición de ser transitiva $(\forall w \forall v \forall z(wRv \wedge vRz) \rightarrow wRz))$. Por último $\mathbf{K5}$ se obtiene de K por la adjunción del esquema axiomático 5, es decir, $\neg\mathcal{K}\alpha \rightarrow \mathcal{K}\neg\mathcal{K}\alpha$ (o $\neg\mathcal{B}\alpha \rightarrow \mathcal{B}\neg\mathcal{B}\alpha$)) y la condición euclideana a R (o $wRv \wedge wRz) \rightarrow vRz)$). Uno de los sistemas más empleados para representar el conocimiento es $\mathbf{S_5}$ que se obtiene de $\mathbf{K}$ agregando como axiomas los esquemas T y 5, con lo que es posible obtener el esquema 4 como teorema.

En este caso, el agente tiene exactamente la misma lógica que el sistema. Más específicamente, si en lugar del sistema $\mathbf{K}$, hubiésemos tomado un sistema que tuviera como axiomas los del cálculo de enunciados y, como reglas de inferencia, tanto el *modus ponens*, como la regla que permite pasar de $\vdash (\alpha \rightarrow \beta)$ a $\vdash (\mathcal{K}\alpha \rightarrow \mathcal{K}\beta)$, entonces los conocimientos del agente serían cerrados bajo consecuencia lógica, pero él no lo sabría. Si a ese sistema hubiéramos adjuntado la regla que lleva de $\vdash \mathcal{K}\alpha$ a $\vdash \mathcal{K}\mathcal{K}\alpha$, entonces cada vez que el agente supiera algo sabría que lo sabe pero, de nuevo, no sabría este hecho. En cambio, en S_5 si el agente ignora P, entonces sabe que lo ignora, y también sabe que si ignora P, entonces sabe que lo ignora.

Después de esta breve formalización del problema de la omnisciencia lógica exploraré los aspectos filosóficos que desde mi perspectiva se omiten en las soluciones ofrecidas. Comenzaré por una afirmación común que hacen los autores al formular sistemas lógicos no omniscientes o que pretenden resolver el problema, a saber, que es inaceptable la idealización que Hintikka hace de las capacidades cognitivas del ser humano. En la siguiente sección pretendo revisar la literatura pertinente al tema y con una perspectiva más extensa que incluya la pregunta por la justificación de la idealización en los sistemas de lógica epistémica, así como el tipo de idealización que utilizan.

2.3. ¿Idealizar es el problema?

Distintos autores enfatizan que el problema con la lógica epistémica está en la idealización del agente humano. Acusan a Hintikka de presentar un prototipo de las capacidades cognitivas que no corresponde a la realidad. Por ejemplo, Levesque señala que la lógica de Hintikka hace "demandas computacionalmente irreales" (Levesque, 1984, p.198). En la actualidad se sigue considerando que el problema de la omnisciencia lógica es la idealización del agente humano: "La objeción principal hacia este acercamiento hecho por los lógicos epistémicos es que en esta se formula una imagen demasiado idealizada del razonamiento humano" (R.Rendsving y J. Symons, 2019). Considero que este diagnóstico es hasta cierto punto acertado, pero también trivial si no profundizamos en la manera en que se idealiza al agente y la justificación que tenemos para ello, si es que la hubiera. En esta sección revisaré algunos argumentos que defienden el uso de una herramienta como la idealización tanto en la lógica como en la lógica epistémica. Parece imposible que podamos prescindir de la idealización en una labor como la de modelar la creencia y el conocimiento, sea que se trate de describirlos o de normarlos. Entonces, ¿por qué no aceptamos que la omnisciencia lógica es una idealización que adoptamos para hacer más práctica la tarea del lógico y no como un problema que debemos resolver o en el mejor de los casos evitar? En efecto, podría decirse que dicha idealización en sí no constituye una objeción. Finalmente, la ciencia ha logrado considerables avances simplificando los problemas a los que se enfrenta. Si al estudiar la caída de los cuerpos Galileo eliminó la fricción y Turing supuso que sus máquinas podían disponer de papel y tiempo ilimitados, ¿por qué la lógica no podría igualmente soslayar ciertos factores que suelen perturbar las actividades cognitivas de agentes reales?

Voy a referirme enseguida a dos formas de justificar el recurso a la idealización en lógica, para finalmente oponerme a ambas. Por un lado están quienes piensan que la lógica es normativa y que, por ello, está justificada en idealizar al agente humano. Toda vez que debería decirnos cómo debemos razonar, la lógica requiere de un conjunto de presuposiciones que no necesariamente tienen que corresponder con la realidad. Por otro lado, encontramos autores que afirman que, aún si adoptamos una postura descriptivista de la lógica, el uso de la idealización está justificado en tanto se aplique a la construcción de sistemas lógicos que capturan algunos rasgos de nuestras habilidades inferenciales. No obstante que estas posturas no se contradicen ni se excluyen entre sí, el marco teórico que cada una plantea es distinto y digno de consideración. Ilustraré cada una de ellas con las propuestas de Colyvan (2012), de un lado; y de Yap (2014), del otro, aunque otros autores hubiesen podido servirme para el mismo fin.

2.3.1. Idealizaciones Normativas. La Postura de Colyvan

Colyvan en su artículo "Idealizations in Normative Models" (Colyvan, 2012) contrasta los modelos científicos con los modelos normativos. Los primeros son teorías científicas que se sustentan en lo que conocemos como el método científico, es decir, tienen sus fuentes en la experimientación y en la observación. En este punto yo adoptaré la palabra "modelo", pero no haré distingos entre modelos y teorías[1]. Podemos considerar como ejemplos de modelos científicos a la teoría de la evolución en biología o la teoría dual de sistemas en psicología cognitiva. Por lo general, se considera que los modelos científicos pueden ser explicativos cuando intentan informar por qué un fenómeno o evento ocurre de una manera determinada, sin que se involucren juicios de valor. Asimismo los modelos científicos pueden ser descriptivos si capturan cómo funciona el fenómeno en cuestión y pueden revelarnos detalles desconocidos en el momento de la investigación. Los modelos evolucionistas suelen ser explicativos en la medida en que, por medio de la selección natural, pueden dar cuenta, por ejemplo, de por qué aparece un rasgo en una especie determinada. En cambio, un mapa de una ciudad describe, pero no explica, la disposición espacial de calles y avenidas. Como la explicación y la descripción (e incluso la predicción) son objetivos afines en las teorías, a este tipo de modelos lo denominaré simplemente "descriptivos".

En cuanto a los modelos normativos, su esencia consiste en decirnos cómo debería ser determinado fenómeno. Como referentes de estos modelos Colyvan

[1]Tampoco Colyvan hace al respecto ninguna distinción

toma a la teoría bayesiana de la probabilidad, a la teoría de la decisión y a la lógica clásica. Aunque el autor no se extiende lo suficiente en caracterizar a los modelos normativos, está claro, por sus ejemplos, que siempre se trata de modelos del comportamiento humano, sea individual o de grupo. Si bien estos modelos también pueden ser útiles en la descripción o explicación, tienen una cualidad especial, a saber, establecen un requerimiento de racionalidad que les permite diseñar cómo debería ser la conducta humana, si debe contar como racional. Ahora bien, para Colyvan la lógica es normativa, aún si a veces pueda usarse (como parte de otras disciplinas) para la descripción: "la lógica puede ser considerada como una teoría de cómo deberíamos sacar inferencias. Así construidos, los axiomas y presuposiciones de esas teorías [las normativas] están motivados por cánones de racionalidad y aparentemente tienen poco que ver con la evidencia empírica" (Colyvan, 2012. p. 1337-1338). Colyvan afirma que sólo podrían ser descriptivas de un agente ideal pero, con respecto al ser humano real, son normativas.

Una de las tesis centrales de Colyvan es que hay dos tipos de idealizaciones en los modelos científicos, aunque también podemos encontrarlos en los modelos normativos. La primera es la de conveniencia matemática. Esta idealización se emplea cuando ciertas características del fenómeno en cuestión son redondeadas o desdeñadas para poder utilizar un aparato matemático en su estudio. Así por ejemplo, el crecimiento de la población, que es un fenómeno discreto, puede considerarse como continuo. El segundo tipo de idealización es el que Colyvan llama "cercanía con el jazz" *(Close enough for jazz)*. Este tipo de idealizaciones son "invocadas porque son suficientemente cercanas al modo en que las cosas son" (Colyvan, 2012, p. 1339). Pero no se refiere a que el modelo se asemeje mucho al fenómeno estudiado, sino a que la idealización es hecha para que el modelo produzca resultados cercanos a los que realmente se observan. Una idealización de este tipo, dice Colyvan, es la supocisión de que las tasas de crecimiento son constantes. Desde luego que no son constantes y tal vez tengan grandes oscilaciones, pero esa simplificación genera predicciones bastante acertadas.

Aparte de estos dos tipos de idealización compartidas por los modelos normativos y científicos, hay una tercera que es propia y exclusiva de los primeros, a saber, el ya mencionado requerimiento normativo, que es una condición de racionalidad: "Estas son idealizaciones impuestas por la racionalidad; una falla en tales condiciones aparantemente podría ser una falla de racionalidad"(Colyvan, 2012, pag.1342). Colyvan puntualiza qué postulados de la lógica clásica considera que están comprometidos con los tipos de idealización que ya vimos. Al respecto señala lo siguiente:

> "(. . .) Tenemos idealizaciones por conveniencia matemática tales como que los conectivos son veritativo funcionales. Hay idealizaciones que son suficiente cercanas al jazz tales como que a cada proposición se le asigna un valor de verdad. Entonces hay condiciones normativas tales como el comportamiento de al menos uno de los conectivos (tal como la tabla de verdad para la conjunción). Y como antes, muchas de las idealizaciones podrían estar motivadas por una mezcla de más de una de las tres consideraciones." (Colyvan,2012, pag. 1342)

Su idea parece ser la siguiente. Por ejemplo, hay usos del condicional en lenguaje ordinario que no son veritativo-funcionales y muchos otros que sí lo son. Quedémonos con estos últimos y con ello podemos aplicar un teoría combinatoria muy sencilla. Por eso es una idealización de conveniencia matemática. Por otro lado, muchos enunciados aparententemente carecen de valor de verdad (por ejemplo, los de ficción), pero tratémoslos como siendo verdaderos o falsos y la lógica nos arrojará resultados bastante acertados. Es una idealización de cercanía con el jazz. No me parece tan claro que estas dos primeras idealizaciones caigan en las categorías que Colyvan les asigna, pero sí que difieren de la última. Antes enfatizó que la lógica prescribe cómo debemos inferir (al menos para ser considerados racionales). Ahora considera que la tabla de verdad de la conjunción es producto de una idealización normativa. Aparentemente es porque la tabla nos dice, por ejemplo, que si creemos la conjunción de dos proposiciones estamos autorizados a creer cada una de ellas. Esto valida un teorema particular del sistema **K** como un tipo de idealización normativa. Es decir, en la medida en que no estamos describiendo el modo en que inferirnos sino prescribiendo cómo debemos hacerlo (para ser racionales), este caso de omnisciencia lógica estaría justificada. Lo que me importa subrayar aquí son dos puntos: a) que la lógica, como toda ciencia, contiene algún tipo de idealización; b) aunque Colyvan se refiere a la lógica, y no a la lógica epistémica, siguiendo sus criterios, la omnisciencia lógica debería ser clasificada como una idealización normativa.

Aunque otros autores han defendido puntos de vista similares, hay quienes, como Yap, sostienen específicamente que la lógica epistémica es una disciplina descriptiva y que, en su tarea de modelar al tipo de agente humano, requiere de cierto tipo de idealizaciones. En lo que sigue revisaré su postura contrastándola con la de Colyvan.

2.3.2. Idealizaciones Descriptivas. La Postura de Yap

En contraste con Colyvan, aunque no necesariamente en oposición con él, Yap defiende que, aún si la lógica epistémica es descriptiva, está justificada en emplear idealizaciones de los mismos tipos que son comunes a otras disciplinas científicas. Una de tales idealizaciones es la omnisciencia lógica y, como tal, no es ningún defecto que haya que enmendar. En todo caso, otros enfoques pueden proponerse para estudiar el razonamiento. Aunque dice explícitamente que no atacará la tesis de que la lógica epistémica es normativa, menciona varios argumentos contra esta tesis. Por ello, es interesante contrastar su punto de vista con el de Colyvan.

Según Yap, una de las objeciones que se hacen a la lógica epistémica es que sus aportaciones a la epistemología son de poca utilidad, puesto que modela el conocimiento de un agente idealizado. Como resultado obtenemos una imagen del conocimiento que no es del conocimiento ordinario. Retoma principalmente una objeción que hace Hocutt en un artículo de 1972 títulado *Is Epistemic Logic Possible?*(Hocutt 1972). En él el autor se plantea la siguiente pregunta: ¿en qué sentido la llamada lógica epistémica merece tal título? Hocutt busca en los formalismos de la lógica de Hintikka alguna definición de conocimiento y se da cuenta que no hay una noción como tal de conocimiento, sino que Hintikka idealiza demasiado ciertas capacidades de los agentes. Haré una muy breve exposición de los principales argumentos de Hocutt en favor de su tesis, porque no son demasiado ajenos a mi tema, especialmente el que Yap nombra *el dilema de Hocutt*, cuya conclusión puede resumirse de la siguiente manera: "o la lógica epistémica falla en ser lógica o falla en ser epistémica" (Yap, 2014, pag. 3353). Hoccut se pregunta si hay tal cosa como (una) lógica epistémica. Que haya formalismos así llamados no es suficiente para responder a la pregunta. Tome una lógica modal e interprete sus operadores modales como epistémicos. ¿Basta eso para tener un sistema de lógica epistémica? Claro que no. Un buen inicio sería comenzar con la siguiente definición: una verdad lógica es un enunciado verdadero que permanece como tal tras la sustutición de todos sus téminos no esenciales[2]. "Si el cielo es azul y la luna es de queso entonces el cielo es azul" es una verdad lógica, si consideramos "si... entonces" e "y" son sus únicos términos esenciales. Los otros vocablos que figuran en ese enunciado son irrelevantes a su verdad. Una verdad de la lógica epistémica es una verdad lógica que contiene entre sus términos esenciales términos epistémicos. Por ejemplo, "Si Pedro sabe que el cielo es azul, entonces el cielo es azul" es una verdad de la lógica epistémica, si consideramos también como esencial el término "sabe". Sin embargo, en otros ejemplos puede verse

[2]Definición que remonta a Quine (1950).

que esta definición no funciona porque los términos epistémicos son opacos. El enunciado "Si Cicerón es idéntico a Tulio y Pedro sabe que Cicerón es idéntico a Cicerón entonces Pedro sabe que Cicerón es Tulio" es falso. ¿Cómo podemos resolver este problema? Imagina Hoccut que tenemos dos agentes. El primero, denominado LPK (el hombre perspicaz), tiene la característica de que que sus conocimientos son cerrados bajo consecuencia lógica, tal vez porque puede hacer deducciones de forma automática. El segundo, que llama LOM (el hombre confundido), tiene las limitaciones usuales del agente humano y no es lógicamente omnisciente. Si solo tomamos en cuenta a este segundo, la lógica de Hintikka ni siquiera es lógica. Las verdades de la (o de una) lógica deben presevar su verdad ante toda sustitución de su léxico no-esencial, pero hay proposiciones P y Q y al menos un sujeto X tales que "si Q se sigue de P y X sabe que P, entonces X sabe que Q" es falso. Lo mismo ocurre si sustituimos "si Q se sigue de P" por "si X sabe que Q se sigue de P". Dicha de otra manera, los teoremas de la lógica de Hintikka no serían verdaderos[3]. Otra posibilidad sería decir que hay dos tipos de conocimiento: el del hombre confundido y el del hombre perspicaz. El sistema de Hintikka trata de conocimiento en este segundo sentido. Pero ¿qué sabemos de este tipo de conocimiento? Que es el que detenta el hombre perspicaz, y ¿quién es este? Aquel al que se aplica la lógica de Hintikka. Esas respuestas, por supuesto, no son satisfactorias. El conocimiento que conocemos es el del hombre confundido, pero de este no trata la lógica epistémica. Ergo, no es epistémica. El siguiente es otro argumento de Hoccut: adoptemos la definición dada de lógica epistémica y admitamos que trata de un conocimiento tal que si $a = b$ (o si P es lógicamente equivalente a Q) entonces a puede ser sustituido por b (y P por Q) en cualquier contexto epistémico *salva veritate*). Por un lado, podría decirse que la lógica no trata del conocimiento en el sentido ordinario del término. Por otro, podría decirse que los operadores epistémicos no tienen ninguna incidencia en las verdades lógicas. Por ejemplo, decir "X sabe que P es lógicamemte equivalente a X sabe que Q" sería una forma rebuscada de decir que P es lógicamente equivalente a Q. En ambos sentidos la lógica epistémica no sería epistémica. Volvamos a Yap quien retoma el problema de Hoccut. Apela a Hintikka quien escribe lo siguiente: "... las leyes de la lógica no son necesarias en el sentido de ser inevitables. No son verdades que *debamos* saber sino verdades que *podemos* saber sin hacer uso de ninguna información fáctica... no son leyes del pensamiento en el sentido de leyes naturales... [ni] mandatos, excepto quizás del pensamiento más agudo po-

[3]Mejor dicho serían esquemas con algunas instancias de sustitución falsas.

sible[4] "(Hintikka 1962, Pag. 37). Es decir, la lógica epistémica trata de la noción ordinaria de conocimiento y, por lo tanto, de agentes reales. Simplemente hace algunas idealizaciones necesarias, como las otras disciplinas científicas. Para ello distingue formas de idealización.
Antes de describir los tipos de idealización a los que apela, es importante advertir que para Yap la idealización es una actividad, tanto en ciencias como en los sistemas formales de lógica. Cada tipo de idealización representa una actividad distinta y, por ende, están justificadas de acuerdo con los fines que persigue cada una. Distingue tres tipos:

a) Galileana: Descompone las teorías en sus partes más simples para hacerlas computacionalmente más tratables.

b) Minimalista: Construye modelos teóricos que sólo incluyen a los factores más relevantes del fenómeno en cuestión.

c) Modelos múltiples: Construye modelos incompatibles entre sí, en donde cada uno representa ángulos diferentes del fenómeno.

Para Yap idealizaciones de estos tipos son admisibles por su utilidad en la descripción de algún fenómeno en especial. Al respecto, la idealización galileana está pensada con fines más pragmáticos en donde la descomposición de las teorías en sus partes más simples puede entenderse como una abstracción de algunos elementos de un fenómeno concreto, mientras que otros factores son dejados de lado, aunque más tarde, con la evolución de la ciencia, puedan ser integrados. Por el contrario, el tipo de idealización minimalista busca describir los elementos más esenciales al fenómeno. Como tal, también es el resultado de una abstracción, pero no hecha en aras de una simplificación, sino de atrapar lo que es la médula del fenómeno. No suele cambiar con el advenimiento de mejores instrumentos de cálculo, a diferencia de lo que ocurre con la galileana. Finalmente, de la idealización por modelos múltiples dice Yap que se da en pro de la predicción y porque un mismo modelo suele perseguir fines distintos.

Recorriendo diferentes sistemas de lógica epistémica, Yap encuentra en ellos los tres tipos de idealizaciones. Muy en particular, señala que, dada la complejidad que comporta el fenómeno del conocimiento, es natural que se presenten en la lógica epistémica modelos múltiples, incompatibles entre sí, pero que destacan, cada cuál a su manera, diferentes aspectos del tópico que abordan.

[4]Las itálicas no están en el original.

No es fácil hacer corresponder las idealizaciones descriptivas de Colyvan con los tipos de modelos de Yap. En el sentido en que la idealización galileana se hace con propósitos computacionales podría equipararse a la de conveniencia matemática. En cambio, por ejemplo, la de cercanía con el jazz definitivamente no corresponde a la minimalista. Tampoco me parece claro que los elementos que estos autores encuentran en la lógica se deban clasificar como lo hacen ellos. En todo caso, estos me parecen aspectos de menor peso para mi tema. En cambio, es importante subrayar que solo pueden detectarse estas idealizaciones (incluidas las normativas) en la lógica, si consideramos a esta disciplina bajo cierta óptica, a saber, en la perspectiva según la cual la lógica está ligada a actividades cognitivas. Para Colyvan la lógica norma nuestras inferencias (p. 1337). Para Yap la lógica epistémica describe ciertos aspectos del conocimiento (p. 3357). Así, por ejemplo, establecer que cada variable proposicional tiene un valor de verdad es una idealización de la forma en que funcionan los lenguajes naturales (donde muchas oraciones carecen de valor de verdad). Pero, como veremos también podríamos pensar que la lógica (a secas) es descriptiva en otro sentido. No es insólito considerarla como una disciplina que tiene como función principal estudiar la relación de consecuencia lógica, que se da o no entre proposiciones (independientemente de cualquier lenguaje en que sean expresadas). En ese caso, no hay tal idealización. A esta cuestión retornaré en la siguiente sección.

Aún si la lógica es parcial o totalmente una ciencia descriptiva en alguno de los dos sentidos mencionados, ¿podrá la omnisciencia lógica ser considerada una idealización descriptiva, como piensa Yap? A mi juicio la respuesta es negativa. La idealización por modelos múltiples consiste en el estudio de un fenómeno desde diversos ángulos que pueden arrojar solo vistas parciales de su objeto de investigación. En cada uno de esos modelos puede haber idealizaciones galileanas o minimalistas. ¿Cae la omnisciencia lógica en alguna de estas dos categorías? En la galileana, definitivamente, no. No se trata de una aproximación que con mejores instrumentos de cálculo, o con la consideración de otros factores, pueda reducirse. La distancia que media entre el ser humano confundido y el perspicaz es inconmensurable. Por lo pronto, aún una computadora con la mayor rapidez de procesamiento conocida no se parecerá en nada a LPK. ¿Es una idealización minimalista? ¿El modelo de Hintikka captura lo esencial de la creencia (o del conocimiento)? Veremos que hay autores que defienden el uso de la semántica de los mundos posibles como el mejor medio de representación de la creencia. No se trata de un asunto sencillo, pero la omnisciencia lógica no solo no es esencial al conocimiento, sino que lo desfigura por completo. Si yo tengo una creencia (que es conocimiento) y mi interlocutor no la tiene, puedo (en algunas ocasiones) proporcionarle un argumento que lo lleve de premisas que cree a la conclusión de la que era escéptico. Eso supone que podemos adquirir conocimiento por medio de un proceso dialéctico. Un ser omnisciente no necesita ni de argumento ni de dialéctica alguna. Su única forma de adquirir "conocimiento" es a través de fuentes perceptuales o intuitivas. El no requiere de ninguna lógica (es decir, de procesos inferenciales). Lo que Yap llama la lógica de LPK es un instrumento recursivo con el que describimos los conocimientos de ese individuo, no la forma en que los adquiere.

¿Puede entonces la omnisciencia lógica ser, como Colyvan pretende, una idealización normativa? Yo sostendré que no. La razón es que no puede ser la norma pedir algo que no solamente es imposible de alcanzar, sino una meta a la que no podemos acercarnos porque siempre estaremos a una distancia infinita de ella. Sin embargo, esto quedará más claro cuando veamos las objeciones de Harman. Si esto es así, la omnisciencia lógica sí es un problema que la lógica epistémica debe resolver. Pero, como mostré antes, es un problema insuficientemente planteado porque no está claro qué debe contar como una solución del mismo. Algunos autores antes citados delimitaron insuficientemente la cuestión y así es imposible saber si cumplieron su objetivo al proponer un sistema para resolver el problema

Aunado a la problemática y al debate sobre el uso de la idealización aparece otra cuestión sobre la cual poco se ha teorizado a saber, la creencia. ¿A qué se refiere el lógico que propone un sistema de lógica epistémica cuando supone que el agente cree P en el mundo w? En la siguiente sección desarrollaré este elemento de profundas raíces filosóficas y más adelante abordaré la cuestión de cómo debemos entender el adjetivo "lógica".

2.4. ¿Qué es la Creencia?

En esta sección abordaré el tema de la creencia siguiendo el patrón filosófico que me he propuesto. El resultado del analisis de la creencia se verá en la disputa sobre la normatividad de la lógica y en los siguientes capítulos cuando analicemos los sistemas de lógica doxástica.

¿A qué nos referimos cuando decimos que creemos tal o cual cosa? Esta pregunta puede remitirnos a un ambigüedad en el uso que hacemos de la palabra "creencia". Esta adopta un significado distinto dependiendo del enfoque disciplinario en que se use. Entre varios sentidos, podemos destacar una acepción cotidiana de "creencia" la cual se emplea para señalar alguna suposición acerca de algún evento en particular. Por ejemplo, cuando yo digo "creo que va a llover" no tengo la garantía de que esto va a ocurrir, simplemente estoy especulando. En este mismo plano podemos explicar ciertas conductas de otras personas afirmando que actuaron de acuerdo con ciertas creencias. Este sentido ordinario de "creencia" es laxo porque no requiere de otros elementos cognitivos ni teóricos para su uso. Otro empleo de esta palabra se da en el plano filosófico en donde la creencia pasa a ser objeto de analisis y un elemento central del conocimiento, junto con la justificación. Hay ciertas características propias de la creencia que hacen que por sí misma sea un tema relevante para la filosofía, por ejemplo, las cuestiones de cómo se justifican las creencias y de cuál es la relación entre creer algo y conocerlo. Por su parte los filósofos han dedicado gran parte de su estudio a definir de forma adecuada a la creencia. Sin embargo, es importante hacer una aclaración. En el uso ordinario, las más de las veces la creencia se opone al conocimiento. Solemos decir "esto no lo creo, lo sé de cierto". En cambio, en filosofía es lugar común definir el conocimiento como una creencia (con ciertas propiedades). Esta discrepancia tal vez pudiera explicarse por una implicatura conversacional. Nadie responde que hoy es lunes o martes, sabiendo que el primer disyunto es el correcto. De la misma manera decir que solo creo algo cuando en realidad lo sé, sería faltar a una de las reglas implícitas de la conversación. Sea como fuere, es importante aclarar que trato aquí de la creencia como se hace en epistemología, es decir, como un componente del conocimiento. Así como en el uso cotidiano tomamos a la creencia como un acto o como el objeto de ese acto, en el plano filosófico el acto de creer y su objeto son temas de estudio de disciplinas como la epistemología, las ciencias cognitivas y, por supuesto, la lógica.

Para poder entender qué es la creencia más allá del plano cotidiano, voy a revisar cómo distintas teorías filosóficas la definen. Como ya lo mencioné al inicio de esta sección adoptaré la clasificación hecha por Schwitzegebel (2021). Veremos que aunque cada una de las categorías de la siguiente clasificación es defendida por varios autores, hay ciertos elementos comunes que nos permitirán reclasificarlas.

2.4.1. Representacionalismo

La primera, el representacionalismo, es la que mejor expresa el concepto común, el que es usual encontrar como implícito en una gama amplia de la literatura filosófica; no obstante lo cual, hay defensores de esta postura que le agregan matices y peculiaridades propias de su pensamiento. La noción común es que la creencia es una representación albergada en la mente acerca de algún hecho posible. Además una creencia es susceptible de entrar en relaciones causales y, en particular, de generar conductas bajo ciertas circunstancias.

Para el representacionalismo la creencia es principalmente un estado mental que tiene un contenido proposicional. Esto significa que no puede ser caracterizado más que en relación con una proposición. Por ejemplo, para determinar el contenido de una creencia tal como la de Luis cuando cree que el helado que compró ayer es de fresa con chocolate, se requiere del lenguaje. Este es un instrumento necesario para la identificación de creencias. Ahora bien, cuando hablamos de un estado mental nos referimos a algo distinto de un estado físico. Por ejemplo, se dice que los estados mentales más recurrentes son el miedo, el deseo y la creencia. Un estado físico puede ser el hambre. Otro elemento distintivo de la creencia en relación con otras actitudes proposicionales es su vínculo con la verdad. Es decir que alguien cree P, es porque su actitud es de tomar P por verdadera.

Otro elemento relevante de la teoría representacionalista es que se concibe a la creencia como una representación mental. Hay ciertos elementos mentales que representan una proposición o un hecho posible porque están estructurados de la misma forma que los componentes de esa proposición. Esta es la idea defendida por Wittgenstein en el *Tractatus* (Wittgenstein, 1921). Otra posibilidad es que cierto estado A de un mecanismo mental M represente un hecho B, porque M entra en el estado A cuando B se produce. Por ejemplo, el perro de Pavlov fue llevado, por el condicionamiento operante, a salivar cada vez que se le anunciaba su alimento con el sonido de una campana. Podemos decir que entra en un estado que representa el hecho de recibir su comida. Dretske (1988), quien defiende esta postura, solo acepta que se adscriban creencias a organismos que tienen una compleja red interconectada de este tipo de representaciones, si bien no establece un límite claro que separe a los individuos susceptibles de tener creencias de los que no pueden tenerlas. En este caso no hay similaridad estructural entre la creencia de que P y P. Quienes defienden que la creencia representa a través de rasgos estructucturales pueden explicar fácilmente que un sujeto pueda creer una gran variedad de cosas. Por ejemplo, si el sujeto cree que el libro está sobre la mesa a través de elementos mentales a y b que entran en una relación R (aRb), entonces también puede creer que la mesa está sobre el libro o que ninguno está sobre el otro. Esta creatividad es más difícil de explicar en una concepción como la de Dretske.

En general, ¿cómo se genera esta combinatoria? Una posibilidad, es que se consiga a través de un lenguaje como cualquiera de los que conocemos y que justamente se caracterizan por ser gramaticales, es decir, por tener reglas recursivas que permiten generar una cantidad ilimitada de oraciones nunca antes producidas.

Puede postularse que hay un lenguaje del pensamiento, privado e interno, como lo hizo Fodor. Otra posibilidad es que sea través del lenguaje materno del sujeto que se estructuren sus creencias. La última posibilidad es que no sea por medio de un lenguaje, sino por otro tipo de mecanismo estructurado. Un ejemplo de este último caso serían los mapas. El sujeto tiene una configuración gráfica en la mente que representa, por ejemplo, las relaciones espaciales de algunos elementos de la realidad. Esta postura tiene una interesante consecuencia para nuestro tema. Alguien podría pensar que Jericó estaba al sur del Mar Muerto y después se entera de que, en realidad, estaba al norte. Este cambio de creencias solo requiere desplazar hacia el norte un elemento de su mapa mental. Automáticamente en la nueva configuración el mar Muerto quedará al sur de Jericó, sin necesidad de que el sujeto haga ninguna inferencia. Dos desventajas parece tener esta concepción. La primera es que puede representar más de la cuenta. Alguien puede tener la creencia de hay autos estacionados en su calle, sin que tenga la creencia de que hay exactamente tres, cuatro o cinco automoviles allí afuera. Pero el mapa tendrá que tener un número determinado de objetos que correspondan a los autos. Es decir, que el individuo no podrá creer que algunos hombres han llegado a la luna, sin creer que exactamente *n* individuos han alunizado, para *n* un número determinado. Por otro lado, el mapa no puede representar una contradicción y ni siquiera una configuración geómetrica imposible. En cambio, la versión lingüística de la creencia no tiene estas desventajas, pero sí que en el ejemplo de Jericó, el agente tendrá que partir de la información que adquirió y hacer una serie de inferencias para llegar a su nuevo estado de creencia, lo que supondrá un proceso más elaborado y tardado. A este respecto, Dretske piensa que las representaciones mentales no tienen ninguna estructura representacional. Es solo por su referencia a eventos externos que se relacionan. Más adelante analizaré con más detalle las propuestas de Dretske y Fodor como conspicuos representantes de esta postura.

2.4.2. Disposicionalismo

Estas teorías afirman que alguien cree una proposición P si tiene una cierta disposición a actuar de acuerdo con el contenido de P. Por ejemplo, el que una persona crea P significa que está dispuesta a aseverar P en condiciones adecuadas y a responder afirmativamente si se le pregunta por la verdad de P. Es decir, para esta postura la creencia se agota en las disposiciones para el comportamiento. Aún cuando pueda haber un estado mental, incluso de tipo repesentacional, que ocurra en el momento en que el sujeto manifiesta su creencia, este no forma parte de la creencia. Esta idea proviene del conductismo y, en última instancia, de un afán de reducir la cientificidad a eventos objetivos y observables. Así a un sujeto podríamos atribuirle una creencia en un tiempo t, no por estados mentales ocurrentes en ese momento, sino en términos de su conducta anterior o posterior a t. Sin embargo, en su versión más estricta el disposicionalismo es muy difícil de defender. Por un lado, solemos atribuir creencias a un sujeto que, por circunstancias externas, no puede ofrecer ninguna manifestación conductual de las mismas. Por ejemplo, imagine a alguien que está encadenado y con la boca cubierta, pero que cree que lo mejor para él es huir de ahí por cualquier medio. Por supuesto que el conductista dirá que tiene la disposición de hacerlo, pero no está en las circunstancias que le permitan actualizar esa potencia. El defensor de esta postura tendrá que recurrir a una explicación de las disposiciones en términos contrafactuales que se avenga con su verficacionismo, lo cual no es sencillo. En segundo lugar, es fácil imaginar que dos personas tengan exactamente el mismo comportamiento pero podamos atribuirles diferentes creencias, porque suponemos que tienen distintos fines o intenciones. Dos personas pueden ir a vacunarse, una porque busca salvarse y otra porque quiere morir (y cree que la vacuna es letal). Una posibilidad para evadir estas dificultades es relajar la restricción de atenerse a lo que puede ser directamente observable. Por ejemplo, admitir también en la determinación de la creencia de un sujeto su disposición a tener ciertos pensamientos o deseos, aunque estos no se manifiesten conductualmente. Sentir sorpresa al enterarse de que P es falso o aseverar mentalmente que P, o tener la disposición a ello, contaría como parte de la creencia de un individuo en P. Sin embargo, al adoptar esta postura se abandonan algunas de las motivaciones originales para aceptar el disposicionalismo (por ejemplo, una defensa del fisicalismo). Otra posibilidad es admitir una variante más refinada que es la que esbozo a contiuación.

2.4.3. Interpretacionalismo

Esta postura también reduce la creencia provee patrones de acciones y reacciones observables del sujeto y no a una representación mental (como lo hace el representacionalismo). A diferencia del disposicionalismo, las teorías de esta categoría adoptan la idea de que las creencias corresponden a un comportamiento del sujeto interpretado por un observador externo. Más adelante pondré como ejemplo de esta categoría el caso de Dennett. Para este autor, pueden adscribirse plenamente creencias a un individuo S, cuando la conducta de S puede ser explicada y prevista con frecuente éxito usando un tipo de estrategia que llama intencional. Otro de los defensores de este tipo de teorías es Davidson quien caracteriza a la creencia en términos de nuestras prácticas de atribución de creencias, en donde las expresiones lingüísticas o conductas de otros cobran sentido. Davidson imagina un lingüista que realiza una traducción radical, es decir, de un idioma completamente desconocido basado solo en la conducta observable de sus hablantes. La determinación de intenciones, creencias y significado de las proferencias lingüísticas de otro se da simultáneamente al explicar su conducta. Para Davidson esto es lo que hacemos al interpretar la conducta de nuestros vecinos. Schwitzgebel presenta esta postura como una variante o un refinamiento moderno del disposicionalismo. Como dije, la diferencia esencial entre ambas es que para el interpretacionista la creencia es atribuida a un sujeto por un observador externo que interpreta su conducta.

2.4.4. Funcionalismo

Estas teorías toman a la creencia como un estado mental, pero no en el sentido de una estructura interna, sino como una entidad caracterizada por el papel que juega en ciertas relaciones, más específcamente por entrar en relaciones causales principalmente con estímulos sensoriales o con otros estados mentales. Esta dualidad entre estado interno y función puede ser aplicada a la definición de artefactos. Por ejemplo, para un químico un jabón es una sustancia sólida o líquida compuesta de sales de sodio o potasio y que es producto de la descomposición de un ácido graso. En cambio, desde el punto de vista funcionalista un jabón es un producto que se usa para la higiene personal y para lavar ciertos enceres o ropa. En ese sentido, cualquier cosa que cumpla esta función, aunque sea hecho de cebo de ganado, es un jabón. El funcionalista aplica esta idea a la definición de los estados mentales. Así no importa mucho qué ocurra en el cuerpo y en la mente de un sujeto para decir que siente calor. Lo importante es que ese estado fue originado, por ejemplo, por exceso de radiación solar, y que produce que el individuo busque hidratarse. Lo mismo hace el funcionalista con la creencia. Esta postura no es incompatible con el disposicionalismo. Por ejemplo, la creencia de un individuo de que la salida de emergencia de una sala se encuentra en una cierta dirección causa que, cuando se encuentra en ella y escucha una alarma, corra en esa dirección. También podría decirse que tenía esa disposición. En cambio, el funcionalismo parece menos compatible con un representacionismo estricto (como el de Fodor). Si creer que P es tener una representación mental determinada, entonces no es esencial a la creencia el que pueda entrar en relaciones causales con estímulos y respuestas. Sin embargo, algunos autores han adoptado ambas posturas.

2.4.5. Eliminativismo

Según esta postura el hablar de creencias y nuestras teorías intuitivas sobre ellas forman parte del habla y la psicología populares, pero no es algo que la ciencia seria finalmente retendrá. Es decir, según el eliminativismo una teoría acabada del mundo no contendría términos como “creencia”. Una corriente de esta postura es el instrumentalismo la cual considera que las atribuciones de creencia son útiles para ciertos propósitos prácticos en la vida cotidiana. Sin embargo, afirma que no hay algo como las disposiciones, conductas o representaciones que, en última instancia, constituyan las creencias. En su versión más “estricta” esta postura niega absolutamente la existencia de tales cosas como las creencias. Una versión débil admite que las creencias son reales, pero no en un sentido robusto. Un simil es el siguiente. Es usual, por ejemplo, explicar el éxito de un basquetbolista que anota varias canastas frecuentes durante uno o varios partidos consecutivos diciendo que está en una buena racha. Sin embargo, para un estadístico con una gama amplia de datos al respecto, esa es una ilusión poderosa, pero sin ningún fundamento. Él puede explicar por leyes estadísticas la concurrencia de varios fenómenos que para una mente irreflexiva serían producto de relaciones causales. De la misma manera, el eliminativista supone que, aunque el recurso a la noción parezca explicativo, e incluso, lo sea, no hay nada el mundo que sea una creencia. En tal sentido, esta postura no es incompatible, por ejemplo, con el disposicionalismo.

2.5. Atribución y Representación

Ahora bien, hay dos aspectos de las distintas nociones de creencia que me parecen relevantes para su representación en un sistema lógico. Primero, hay un elemento que describe a la creencia como un evento interno tal como una representación contenida en un estado mental con un contenido proposicional, como lo señala el representacionalismo. El segundo elemento es la idea de creencia como algo que puede atribuirse a un sujeto basado en su conducta observable. Es decir, en esta concepción la creencia no es un acto privado que solo puede ser conocido plenamente por el creyente, sino algo que, en ultima instancia, siempre puede reducirse a patrones observables o intersubjetivos. En la primera categoría queda el representacionalismo como paradigma porque, aunque el defensor de esta postura puede aceptar que la creencia se manifiesta o puede manifestarse a través del comportamiento, nunca se reduce a ello. En la segunda categoría entran el disposicionalismo y, mejor aún, el interpretacionismo. Por ejemplo, según el disposicionalismo, un sujeto que tiene disposiciones de cierto tipo es un creyente, aunque descubramos más tarde que es un robot. De nuevo, el disposicionalista puede admitir que esas disposiciones están causadas normalmente por eventos cerebrales o mentales, pero estos no constituyen la creencia. El interpretacionista aceptaría bajo ciertas condiciones que una máquina tiene creencias, no como una figura o metáfora, sino en el sentido pleno de la palabra. La propuesta del funcionalismo es compatible con una noción de creencia tanto interna como externa, puesto que no define propiamente qué es la creencia sino que le asigna un rol causal que puede variar y, en este sentido, la creencia puede ser tanto una representación mental como una disposición conductual. Para el eliminativista simplemente no hay creencias aunque, como ya dije, su posición es compatible con el dispocisionalismo o el interpretacionismo. En resumen hablaré de creencias como representaciones y de creencias como atribuciones.

En lo que sigue ejemplicaré estas dos nociones más generales de creencia. Por un lado, expondré brevemente las teorías de Fodor y Dretske para ampliar la definición de creencia como representación mental. Por otro lado, expondré a Dennet y a Davidson como precursores de las creencias como atribuciones.

2.5.1. La Creencia como Representación Mental

Para profundizar en las posiciones de Dretske y Fodor respecto de la creencia es pertinente mencionar una crítica que ambos hacen a las teorías disposicionalistas e interpretacionistas de la creencia. Recordemos que para las dos últimas teorías la creencia se refleja en determinados patrones conductuales propios de los seres humanos. Además, en tanto que tales patrones son observables, quienes los percibimos podemos adscribir ciertos estados mentales al sujeto en cuestión. Por ejemplo, si alguien cree que más tarde lloverá entonces tomaría la decisión de llevarse el paraguas al trabajo. Estas formas de concebir a la creencia resultan ser bastante intuitivas y eficaces pues en la mayoría de los casos nos permiten atribuir estados mentales y conocer los efectos que tiene la creencia en las acciones humanas. Es decir, son informativas y predictivas, además la información y las predicciones son verdaderas en la mayoría de los casos. Pero para Fodor y Dretske esta teoría tiene una deficiencia explicativa en tanto que no establece con claridad la relación entre los estados mentales como la creencia o el deseo y la conducta. En lo que sigue veremos como tanto Fodor como Dretske intentan subsanar este vacío.

La creencia según Dretske

Dretske (1988) va a sostener que las creencias se encuentran agrupadas en sistemas. De manera particular, las creencias son representaciones que tienen un contenido y una referencia, pero además, tienen un rol causal en el comportamiento observado y en la producción de impulsos. Para tener una visión más amplia de la creencia en Dretske es necesario revisar su idea de sistemas representacionales y ver a cuál de ellos pertenece la creencia.

Dretske define un sistema representacional como: "Un sistema cuya función es indicar cómo se encuentra un estado de cosas relativo a una condición o magnitud". (Dretske, 1988, Pág. 52). El autor distingue tres tipos de sistemas representacionales, a los cuales simplemente llama I, II, III. Estos sistemas se clasifican de acuerdo con la función que realizan. En cierto sentido todos tienen la función de indicar, pero no todos pueden referirse a las mismas entidades, pues cada uno tiene un mecanismo diferente. La función de los sistemas tipo I le es asignada desde afuera y, por lo tanto sus poderes representacionales no son intrínsecos (con "poderes representacionales" me refiero a la capacidad de cada sistema para desempeñar su rol indicativo). Un ejemplo de este tipo de sistema puede ser el siguiente: tenemos a un entrenador de basketball que en el medio tiempo de un importante partido instruye a sus jugadores sobre ciertas jugadas estratégicas. Como no debe exponer dichas jugadas al equipo rival, en una base de cartón dibuja una cancha y para darle más realismo utiliza algunos objetos para representar a los jugadores contrarios y otro tipo de objetos para representar a sus propios jugadores, luego se dispone a enseñar las estrategias. Lo que es relevante es que ninguno de los objetos que usa fueron diseñados para representar una estrategia de juego sino que es una función que el estratega les asigna.
Los sistemas representacionales de tipo II tienen las mismas funciones que el sistema I, a saber, indicar un estado de cosas. Sin embargo, una diferencia importante entre ambos es que en el sistema II una vez que le asignamos una función este mecanismo la realiza sin necesidad de maniobra o supervisión alguna. Aún así no podemos decir que tenga una función intrínseca. Un ejemplo de este tipo de sistema son los objetos que empleamos para medir algún valor, puede ser el medidor de gasolina o un termómetro. Una vez que estas herramientas han sido programadas trabajan sin que haya alguien al pendiente de ellas. No obstante, tampoco la función del sistema II es completamente intrínseca. La regla para medir en metros depende de un acuerdo social. Por esta razón Dretske llama a este tipo de mecanismos, sistemas convencionales. Ahora bien, los sistemas de tipo III son mecanismos naturales que básicamente están en los seres vivos e indican un

estado de cosas, independientemente de intereses o propósitos externos. Por ello se consideran sistemas naturales pues su función es intrínseca. Un ejemplo de este tipo de sistema es la vista, la cual resulta ser un mecanismo muy complejo que no requiere que alguien le asigne su función. Otra propiedad que tienen el sistema representacional III es que es intencional y esta propiedad de intencionalidad es propia de *lo mental*:“Los filósofos han considerado a la intencionalidad como una marca de lo mental. Una dimensión importante de la intencionalidad es la capacidad de tergiversar”. (Dretske,1988. Pág 64). Cabe señalar que los sistemas I y II también son intencionales y tienden a tergiversar la información. Pero a diferencia del sistema III, en los sistemas I y II la intencionalidad es una propiedad asignada por quien diseña el sistema representacional, en cambio el sistema III tergiversa la información como un componente intrínseco.

Dretske apunta que la creencia es un sistema representacional tipo III y como tal tiene una función que no es asignada por nadie, puesto que es un mecanismo natural, además de ser intencional. Pero, ¿cual es la función de la creencia? Dretske acepta la analogía entre un sistema de creencias y un mapa. En el mapa encontramos ubicaciones de puntos muy específicos, pero además la finalidad de un mapa es guiarnos en un terreno muchas veces desconocido. Por ello es que Dretske acepta que, aunque el conjunto de creencias es como un mapa, no se distingue solo por señalar puntos localizables sino, además, por contener información que nos indique la dirección que debemos tomar. Este es el rasgo que distingue a las creencias de otro tipo de representaciones, a saber, que contienen información que nos orienta, de preferencia en la dirección correcta. Pero, no solo eso, las creencias también deberían tener un valor explícativo del comportamiento. Recordemos que, precisamente este último punto era el que le hacía falta a la explicación disposicionalista. Para Dretske, la creencia tiene un rol causal que es llevar información al agente. Digamos que este es su papel y la manera en que incide en el comportamiento.

La concepción de la creencia de Fodor

De acuerdo con Fodor (1987), la teoría psicológica del sentido común es una teoría intuitiva sobre estados mentales que permite explicar o predecir el comportamiento. ¿Cómo funciona esta teoría? De manera muy simple pues basta con observar alguna conducta o escuchar alguna proferencia o expresión para atribuir creencias y explicar conductas. Por ejemplo, yo escucho que Alejandra dice: “el Popocatépetl hizo erupción”. De escuchar esta expresión en voz de Alejandra yo puedo inferir que Alejandra cree que el Popocatépetl hizo erupción y que buscará alejarse de su hogar si este se encuentra en la cercanías del volcán. En la vida cotidiana empleamos esta teoría de manera casi automática y sin darnos cuenta para predecir y explicar la conducta de los otros, de la misma manera que hacemos predicciones acerca de la inflación de los precios sin hacer alusión a una teoría económica en particular. Otra de las bondades de esta teoría psicológica del sentido común es que es predictiva e informativa, similar a una teoría científica; no obstante, al igual que ésta, puede llegar a fallar en sus predicciones, es decir, puede arrojar informaciones falsas.

Fodor (1987) especificó algunas características de esta teoría y dio razones que explican su uso: primero, en la mayoría de los casos que la empleamos funciona bien en tanto que hacemos predicciones efectivas con ella y su mecanismo es, hasta cierto punto, simple porque basta con escuchar alguna proferencia u observar algún evento para poder interpretar las creencias y deseos del otro. En segundo lugar, no es una teoría trivial, es decir, tiene cierta profundidad y podemos establecer algunas generalizaciones a las que subyacen ciertas deducciones similares a las de la ciencia real. En tercer lugar, es una teoría indispensable en tanto que no tenemos otra manera de interpretar las acciones de los otros.

Hasta aquí la concepción de Fodor ve la creencia como atribución, pero eso no es lo esencial. Ahora me referiré a las entidades que postula la teoría. De acuerdo con (Fodor, 1987) lo que está en el trasfondo de nuestras atribuciones de creencia y disposiciones conductuales son los estados mentales. Estos son estados de la mente que representan a un objeto, es decir, son acerca de ese objeto. Hay un tipo especial de estados mentales que son las actitudes proposicionales. El contenido de una actitud proposicional es una proposición. Como ejemplos de actitudes proposicionales encontramos a la creencia, el deseo y el miedo. También hay estados mentales que no necesariamente tienen un contenido proposicional, por ejemplo, el hambre, el sueño, el dolor. Cabe señalar que no todas las teorías acerca de la creencia aceptan definirla en términos de una actitud proposicional ni como estado mental. Veamos qué propiedades le atribuye Fodor a las creencias:

1) Son semánticamente evaluables. Cuando se habla de evaluación semántica nos referimos a aquello que hace que una creencia sea verdadera o falsa. Es decir, el valor de verdad de una creencia depende su relación con el mundo real. Una creencia será verdadera si se corresponde con un evento o si hay algo en el mundo que la haga verdadera. Por ejemplo, si Juan tiene la creencia de que esta corriendo un maratón, esta será verdadera si efectivamente se encuentra participando en dicha prueba. Yo puedo tener la creencia de que el hombre que veo se comporta inadecuadamente porque esta borracho y esta creencia será

verdadera si el hombre ha ingerido bebidas alcoholicas. Ahora bien, para saber el valor de verdad de una creencia es necesario tener acceso a su contenido. Entonces, la creencia tiene un contenido cuya evaluación depende de que haya algo en el mundo que le corresponda.

2) Tienen poderes causales. La teoría psicológica del sentido común reconoce tres tipos de causación mental, a saber: la causación de la conducta por los estados mentales, la causación de los estados mentales por los eventos del entorno que los afecta y, finalmente, la causación de los estados mentales entre sí. Observemos que hasta aquí las actitudes tienen dos propiedades importantes: primero, tienen un contenido semántico que se evalúa de acuerdo a la situación en el mundo. Segundo, los estados mentales, en particular cuando estos estados son actitudes proposicionales, tienen poderes causales que convergen no sólo de las creencias hacia la conducta, sino también hacia otros estados mentales, además de que hay eventos en el mundo que causan estados mentales. Por ejemplo, mi creencia de que debo salir con abrigo de casa puede estar causada por que el cielo esta nublado. Pero también hay casos en que tenemos creencias que causan que se forme otra creencia que no necesariamente se refleja en la conducta.

3) Las generalizaciones implícitas de la teoría del sentido común, son, en gran medida, verdaderas de tales estados. Las generalizaciones que hace la teoría tienen un sustento en el sentido común y se realizan asignando contenidos a los estados mentales.

Es importante mencionar que para Fodor las actitudes proposicionales son también representaciones mentales interconectadas entre sí. Al igual que las palabras en el lenguaje natural, las representaciones mentales tienen su propio lenguaje que es el lenguaje del pensamiento. En consonancia con las actitudes proposicionales, las representaciones mentales también tienen contenido y referencia, pero lo más importante es que juegan un rol explicativo en la conducta humana y esta última afirmación es lo que distingue a un representacionalista de un disposicionalista.

La creeencia como representación. Recapitulación

En esta sección voy a mencionar los aspectos del representacionalismo que quisiera retener:

1) Para el representacionalismo hay dos elementos definitorios de la creencia. Primero, la creencia es un estado mental similar al miedo o al deseo. Estos estados se instancían mediante una proposición. Es decir, todo estado mental es acerca de una proposición. La creencia en P y la creencia en Q son diferentes si P y Q son proposiciones distintas. El otro elemento que distingue a la creencia de otros estados mentales como el deseo es que es una representación mental, es decir, un ítem que habita al interior de nuestra mente y que puede ser evaluado por corresponder o no con el mundo. Tanto Fodor como Dretske están de acuerdo con estas características. Ahora bien es un problema filosófico complejo qué debe entenderse por "representación". En esto, Fodor y Dretske no coinciden.

2) La idea central de Fodor que me gustaría rescatar es la forma en la que logra vincular estados mentales, actitudes proposicionales y comportamientos. Recordemos que una de las premisas básicas en la teoría de este autor es que la creencia no puede reducirse al mero acto conductual, sino que solo se manifiesta en él. Como todas las actitudes proposicionales, las creencias son estados de la mente y tienen como objeto una proposición. Las características de las creencias son las siguientes: a) que son semánticamente evaluables, b) tienen poderes causales y c) con ellas podemos aspirar a establecer generalizaciones sobre estados mentales propios de una teoría científica. Insisto en que aunque, según Fodor, tenemos una teoría intuitiva que nos ayuda a explicar la conducta ajena y contiene una noción de creencia, la creencia primordialmente es un estado mental.

3) Dretske incluye a la creencia en uno de los tres sistemas representacionales. Los sistemas representacionales contienen mecanismos que permiten indicar un determinado estado de cosas, por ejemplo, un procedimiento o un valor específico. La creencia como tal corresponde al sistema representacional III que es un sistema natural porque realiza sus funciones sin necesidad de supervisión y además contiene una programación de origen. Lo que es esencial a la creencia como elemento de este tipo de sistemas es que contiene información acerca del mundo y esa información es imprescindible para guiarnos. La forma en la que Dretske ilustra la función de las creencias es en analogía con un mapa que contiene los nombres de los lugares que debemos ubicar y sus posiciones respectivas. Cuando hay un cambio de creencia o algún proceso inferencial en curso supondremos que el agente se desplaza hacia ciertos lugares en el mapa. La creencia representa, pero no toda representación mental es una creencia.

No es mi objetivo entrar aquí en este debate de filosofía de la mente, sino únicamente apuntar la relación que las creencias tienen con la lógica desde esta perspectiva, aunque luego ahondaré en este tema. Para Dretske la creencia debe contener información que nos oriente y, por ello, el sujeto debería guiarse por la leyes lógicas en su búsqueda de una representación verdadera de su entorno. Por lo pronto, no debería tener un cuerpo de creencias inconsistente. Ahora bien, si las creencias son como mapas una cierta porción de esa lógica es consitutiva de la creencia. En el caso de Fodor, la creencia es una representación mental semáticamente evaluable. Cabe suponer que al individuo le conviene tener creencias verdaderas y, por ello, debe seguir ciertas reglas lógicas (al alcance de sus limitadas facultades cognitivas). En ambos casos, la lógica tiene un rol normativo.

2.5.2. La Creencia como Atribución

Ahora vamos a revisar las bases teóricas de la concepción de la creencia como un elemento atribuible a los seres humanos. La idea central, dicha en sus mínimos términos, es que atribuimos creencias a otros basados en su conducta en circunstancias visibles, de acuerdo con una psicología intuitiva que se forma en nosotros como resultado de la experiencia. Por supuesto, esto es lo que Fodor describe como teoría de sentido común. La diferencia entre ambas reside en que, para este último, la creencia es un estado mental y la atribución es una mera hipótesis, mientras que para la postura "atribucionista" la creencia, antes que nada, es un instrumento conceptual de esa teoría intuitiva. Puede después descubrirse que la atribución acertada de una creencia corresponde con cierto estado mental, pero la creencia no se identifica con este estado. Brevemente podemos decir que la postura "atribucionista" define a la creencia como un elemento accesible para el ser humano que no posee la creencia. Para explicar mejor esta idea voy a contrastarla con el postulado central del representacionalismo. En el representacionalismo la creencia es una representación mental que contiene una proposición acerca del mundo y sólo tiene acceso a ella el creyente que tiene la representación en su mente, aunque en ciertas circunstancias es de esperar que se manifieste en el comportamiento del sujeto. Por el contrario, en la postura atribucionista la propiedad central de la creencia es que puede ser observada y analizada desde el punto de vista de alguien externo al creyente. En esta sección me propongo ampliar los presupuestos teóricos de la teoría atribucionista tomando como referencia dos teorías acerca de la creencia que, con algunos matices, ilustran esta posición, y que fueron propuestas respectivamente por Dennett y Cherniak.

Sistemas intencionales de Dennett y la creencia

Para Dennett (1987) la noción de creencia está íntimamente ligada a lo que denomina la actitud o estancia intencional. Hay tres enfoques, actitudes o estancias, en las cuales podemos situarnos para predecir o explicar el comportamiento de alguna entidad compleja o sistema. La primera es la estancia física. Consideramos al objeto en cuestión como formado por elementos físicos y recurrimos únicamente a las leyes de la física o de la química para predecir lo que le sucederá en una cierta circunstancia. Desde luego, esto puede ser hecho a diversos niveles, por ejemplo, subatómico o molecular. Es lo que hace un químico cuando explica por qué el aceite y el agua no se mezclan, en términos de la densidad de ambas substancias. Pero también es lo que hace una cocinera cuando explica por qué la gelatina no cuajó. Otra es la estancia o actitud del diseño. A veces no es posible predecir un fenómeno desde la perspectiva física, sea por la complejidad a que tal nivel nos obligaría, sea por nuestra ignorancia, y recurrimos a considerar al sistema como algo que está diseñado para cumplir una determinada función. Un ejemplo común es el manejo de un celular. Sería retorcido recurrir a propiedades físicas para explicar por qué se apaga después de un tiempo de inactividad. Asimismo, podemos predecir que la alarma de un reloj de pulsera sonará a las 6.45 solo mirando su carátula, confiando en que está bien construido y que no se ha averiado. Hay ocasiones en que las dos actitudes mencionadas son muy poco factibles y entonces adoptamos la actitud intencional. Consiste en considerar al sistema cuyo comportamiento estamos tratando de predecir como si fuese un agente racional. Le atribuimos creencias, dada su posición en el mundo, e intenciones y deseos fundados en ciertos principios (que dependen del tipo de objeto de que se trate). Entonces suponemos que el sistema operará de la mejor manera posible (según sus capacidades) para conseguir sus deseos de acuerdo con sus creencias. Esta proyección de creencias e intenciones a un objeto procede según ciertas reglas implícitas de sentido común. Por ejemplo, si la entidad en cuestión es un ser vivo supondremos que tiene el deseo de preservar su vida y la de su especie y, por lo tanto, de alimentarse, protegerse de peligros ambientales, reproducirse, y que tiene una propensión hacia lo que le causa placer y que evitará lo que le origina dolor. Además, le atribuiremos ciertas creencias a este respecto. Por ejemplo, podemos predecir muy bien qué hará un ratón en un túnel en una de cuyas extremidades hay un gato y en la otra un pedazo de queso. No es necesario para ello hacer ningún experimento previo.

Para aclarar más esta estrategia intencional, Dennett da algunas otras indicaciones sobre cómo procede. Por ejemplo, cuando se trata de seres humanos considerados como sistemas intencionales, damos por sentado que el que un individuo A haya estado expuesto sensorialmente a X un tiempo adecuado es normalmente una condición suficiente para atribuirle ciertas creencias verdaderas acerca de X. Por supuesto habrá fallas por las limitaciones cognitivas de A (por ejemplo, olvidará algunas cosas) porque, debido a sus intereses, no atenderá a todo lo que estimulativamente le ofrece el ambiente. Por supuesto la atribución de creencias más abstractas procede según criterios más intricados y está más sujeta a controversias. En general, Dennett supone que la mayor parte de las creencias de un individuo tienen que ser verdaderas, pues incluso las creencias falsas se apoyan en creencias verdaderas. Yo no estoy convencida de esto y creo que pudo limitarse a la afirmación más modesta de que las creencias de un agente (vivo) están adaptadas, vía la selección natural, para ser útiles a la supervivencia de su especie. Otro axioma de la atribución de creencias es que, en la mayoría de los casos, un agente parlante que asevera una proposición P la cree, es decir, que la mentira es menos frecuente que la expresión honesta de creencias, y este parece ser un presupuesto de la comunicación, que tiene también un valor de supervivencia. Estos "axiomas" tienen más que ver con cómo un sistema de lógica pudiera ser aplicado a casos reales. En lo que se refiere a cómo debe estar constituido importa mucho la estructura del entramado de las creencias que atribuimos a un sujeto. A este respecto, Dennett es menos específico. Señala, por ejemplo, que no puede atribuírsele a un sujeto una creencia aislada, pero no se pronuncia sobre si es posible suponer que un individuo tiene creencias contradictorias o muy contrarias al sentido común. Se pregunta: ¿puede alguien creer que un conejo es un pájaro? La cuestión no es si alguien tiene la capacidad para representarse algo así, sino si habría circunstancias en las cuales haríamos tal aseveración.

Ahora bien, la creencia es correlativa a la estancia intencional. Sólo podemos adscribir creencias a un sistema cuando adoptamos con cierto éxito la actitud intencional para explicar su comportamiento: "Lo que es ser un verdadero creyente es ser un sistema intencional cuyo comportamiento se puede predecir en forma confiable y amplia por medio de la estrategia intencional" (1987, p. 27)

Aquí surgen algunas cuestiones esenciales. Dennett plantea un dilema entre una concepción realista de la creencia, según la cual un individuo tiene o no una creencia en P (por ejemplo, cree o no que hay un billete de 100 pesos en su bolso) y una concepción relativista, que se ilustra mejor en preguntas del estilo de ¿cree usted que Lula fue un buen presidente? y a las cuales se tiende a responder que depende del punto de vista que se adopte. Esta es la que llama Dennett concepción interpretacionista. Su propia posición rechaza este dilema y combina algo de cada polo. Para adelantar un poco la solución: que a un sistema se le atribuyan creencias determinadas dependerá, en más de un sentido, del punto de vista que se adopte, pero esas atribuciones tendrán algún grado de objetividad.

En un primer momento la adscripción de creencias a un sistema es meramente un movimiento estratégico:

> "La definición de sistemas intencionales que yo he dado no dice que un sistema intencional realmente tiene creencias y deseos, sino que uno puede explicar y predecir su comportamiento adscribiéndoles creencias y deseos. La decisión de adoptar esta estrategia es pragmática y no es intrínsecamente correcta ni incorrecta." (Dennett, 1987, p7)

Sin embargo, es natural que surja la cuestión de cómo distinguir entre sistemas intencionales a los que simplemente atribuimos creencias y aquellos que realmente las tienen. ¿Hay realmente tal diferencia? Recuerde que estamos ante una disyuntiva sobre cómo pensar la creencia, si como un fenómeno interno del cual la conducta es únicamente una manifestación, o bien algo que es atribuido por un observador externo. Para Dennett la respuesta es esta segunda opción, como lo explicaré enseguida.

Empecemos con la siguiente pregunta: ¿hay algo que no pueda ser considerado como un sistema intencional? Por ejemplo, ¿no podríamos pensar que un bulto de tierra es un sistema intencional? Podríamos atribuirle la creencia aristotélica de que debe estar en el centro del universo, además de la creencia de que realmente está allí. De tales hipótesis podemos concluir que no se moverá, que es lo que efectivamente sucede. A esto Dennett opone que en este caso la estancia intencional no ha servido para nada porque no hemos concluido algo novedoso que no pudiéramos prever con anterioridad. Es decir, la actitud intencional no es aquí exitosa y, por tanto, no deberíamos considerar al bulto en cuestión como un creyente. Pasemos a un segundo ejemplo. De un termostato que hace que de repente se apague una caldera podemos decir que cree que la temperatura de ese artefacto ha llegado al límite máximo que debe tener y que es conveniente que ahora se enfríe absorbiendo un poco de la temperatura ambiente, más baja. Pero parece que con ello nos excedemos pues no hay razón para atribuirle la posesión de conceptos como "caldera" o "calor" a un termostato. Sería mejor concederle creencias más esquemáticas, por ejemplo, no sobre esta caldera en particular, sino sobre un objeto y una cierta propiedad, etc. Sin embargo, conforme el sistema es más complejo y está mejor diseñado es factible atribuirle creencias más precisas. Por ejemplo, a un GPS (localizador geográfico) podemos atribuirle la creencia de que el lugar en que se encuentra en un momento dado está cerca de Acolman, pero que no hay un camino directo para llegar a ese lugar. En el extremo de esta gradación, a un sistema tan complejo como un ser humano podemos atribuirle referencia a objetos concretos y representaciones de estados de cosas. Al respecto dice Dennett:

> "No se trata de que atribuyamos (o debamos atribuir) creencias y deseos solo a las cosas a las que encontramos representaciones internas, sino que cuando descubrimos algún objeto para el cual la estrategia intencional funciona, nos esforzamos por interpretar algunos de sus estados o procesos internos como representaciones internas. Lo que hace que algún rasgo interno sea una representación solo podría ser su papel en la regulación de un sistema"(Dennett, 1987, p. 32).

Es decir, del termostato al ser humano hay una gradación, en algún momento vago a partir del cual cual ya podemos hablar, dada la complejidad alcanzada por el sistema, de referencia a ciertos objetos particulares del entorno, y de representaciones. A mayor complejidad de un sistema cuya conducta es predecible y explicable por la actitud intencional, mayor es la propiedad con que podemos decir que realmente tiene creencias. En el límite podemos hablar de que el sistema es un creyente.

Aquí topamos con una cuestión planteada previamente y que ahora podemos formular como una objeción. Si adoptar una actitud intencional para explicar la conducta de un sistema es resultado de su complejidad o de nuestra ignorancia y la atribución de creencias es parte de la actitud intencional, entonces parece que el atribuir o no creencias a un agente depende más de quién se las atribuye y menos del agente en sí. Podemos imaginar observadores muy inteligentes, marcianos dice Dennett, para quienes la conducta humana fuese predecible desde la estancia física. Ellos explicarían todas nuestras acciones considerándonos como sistemas de partículas elementales sujetos a las leyes de la física. Para tales criaturas los seres humanos no seríamos creyentes. La respuesta de Dennett es que la actitud intencional seguiría explicando mejor la situación y prediciendo con más economía las acciones humanas. Consideremos un editorialista que desea criticar una política gubernamental por medio de un artículo que escribirá en su computadora. El marciano hará complejísimos cálculos para predecir cada movimiento de los dedos sobre el teclado, pero no comprenderá que otras muchas posibles combinaciones de movimientos digitales entran bajo el mismo rubro de criticar la política del estado. A su lado, el colega del editorialista que sigue la estrategia intencional y que no se sorprenderá el ver el artículo en el periódico parecerá un genio. La conclusión es que hay un sentido en que la estancia intencional es la más adecuada y por ello, y por la complejidad del sistema, podemos hablar de que el individuo realmente tiene creencias.

Sin embargo, en lo que respecta a mi tema, un punto central es el siguiente: si la estrategia intencional funciona para predecir lo que hará el mencionado ratón que se encuentra en el tubo con dos salidas y eso presupone que es un ser racional, entonces ¿debo adscribirle un conocimiento de todas las tautologías? Parece absurdo. La primera parte de la respuesta de Dennett es muy sencilla: por supuesto que el ratón no puede determinar que una expresión escrita es una tautología, pero se comporta conforme a las leyes de la lógica, según las creencias y deseos que le atribuimos. Si corre hacia el gato tendrá una muerte atroz y no desea morir. Por tanto, no se desplaza en esa dirección. Sin atribuirle esta "obediencia" a las leyes de la lógica sería imposible predecir lo que hará. Pero la segunda parte de la cuestión es qué tanta lógica podemos suponer que "conoce" (en el sentido de ser capaz de seguir). Allí Dennett no es muy claro, pero podemos guiarnos por el ejemplo mismo. No vamos a atribuirle al ratón la capacidad de seguir una estrategia adecuada para salir de un laberinto muy complejo, aunque quien observa su comportamiento sepa esa estrategia. Ni menos podemos predecir que el roedor hallará una solución que ni el observador puede prever dadas sus limitadas capacidades cognitivas. En cualquier caso, racionalidad no conlleva omnisciencia lógica.

Vemos aquí que en la perspectiva de Dennett ciertas leyes lógicas aparecen no como normativas de la creencia, sino como constitutivas de la misma. Esta distinción se remonta a Kant y puede ilustrarse de la siguiente manera. Una receta de cocina es una serie de prescripciones para producir un platillo. Son normativas en el sentido en que el no seguirlas al pie de la letra puede conducir a resultados no deseados. El que las obedece más estrictamente obtiene un producto de mejor calidad, dado el fin propuesto, que el que las sigue laxamente. Por otro lado, las reglas de un juego son constitutivas del mismo. Un juego con tablero y piezas de ajedrez en que las torres se desplazan en diagonal no es ajedrez. No es que la persona que realiza tales movimientos juegue mal, sino que la actividad en la que está involucrada no puede ser descrita como una partida de ajedrez. De la misma manera, el individuo con una conducta impredecible, que dice una cosa y hace otra o que se contradice palpablemente, no puede ser considerado como sujeto de creencias, porque considerarlo como un sistema intencional no es útil. En conclusión, si seguimos la perspectiva de Dennett, la lógica no tendría un rol normativo sobre las creencias que atribuimos a un individuo, sino un papel constitutivo en el sentido dicho. Ahora bien, ¿por qué lógica debe cumplir esta tarea? Claramente no se trata de una lógica que obligue al agente a ser omnisciente. Como diré más adelante, es una lógica que, al menos proscribe ciertos patrones como no propios de un creyente. Un punto de vista distinto, pero en el mismo enfoque que ahora analizo puede hallarse en la obra de Cherniak, que examinaré enseguida.

Cherniak y la Racionalidad Mínima

A diferencia de los autores revisados anteriormente, en Cherniak (1986) no encontramos una definición de creencia ni en el sentido disposicional ni de representación mental. Sin embargo, Cherniak ofrece una teoría de la racionalidad que analiza cómo atribuimos creencias, habilidades inferenciales y de memoria a seres a los que consideramos racionales. Su obra es relevante para mi tema desde dos puntos de vista. Por un lado, está en consonancia con la de Dennett en la medida en que se ocupa de la atribución de creencias (y de otras capacidades cognitivas), y, por otro, toma muy en consideración las limitaciones cognitivas del ser humano para oponerse a las idealizaciones tales como la omnisciencia lógica.

Cherniak comienza con un principio psicológico básico: "... sin racionalidad, no hay agencia".(1986, p. 3) Podemos nosotros agregar: sin racionalidad, no hay creyente. En nuestra vida diaria realizamos atribuciones de creencias que nos permiten predecir conductas de otros seres humanos. Para dicha atribución es necesario presuponer un cierto grado de racionalidad en ellos. Pero ¿qué tanta racionalidad debemos considerar para que un sujeto sea un agente cognitivo y, por ende, atribuirle creencias? La respuesta del autor es que sólo se requiere un grado mínimo de racionalidad: "... aunque en las explicaciones psicológicas diarias del comportamiento requerimos de la racionalidad del agente, de hecho, sólo requerimos racionalidad mínima, como distinta de la idea" (pag. 3). Cherniak encuentra que la concepción de racionalidad que ha sido adoptada comúnmente en filosofía no describe ni funciona de manera correcta para los seres humanos, puesto que no considera sus limitaciones cognitivas, de memoria y de tiempo para razonar. En el marco de la vida cotidiana atribuimos creencias con éxito, pero sin considerar a un razonador ideal, sino únicamente presuponiendo que el agente tiene un conjunto de creencias con las cuales realiza algunos procesos mentales. De acuerdo con Cherniak, es necesaria otra teoría de la racionalidad que tome en cuenta que los seres humanos son limitados en sus habilidades de razonamiento y con la cual se puedan explicar las atribuciones de creencia.

Cherniak describe condiciones necesarias para la racionalidad mínima, es decir, requerimientos sin los cuales no podríamos considerar que un agente es racional en el grado mínimo en que podemos atribuirle creencias. No obstante, dichos requerimientos no pueden considerarse como condiciones suficientes: "Estas serían sólo condiciones necesarias para la agencia. La principal razón por la que no pueden ser suficientes es porque son sólo prospectos para acciones causadas por creencias"(Cherniak, 1986, p.5).

Veamos ahora cuales son las condiciones para la racionalidad mínima, pero antes revisemos algunas delimitaciones teóricas. Cherniak va a considerar que un agente mínimo[5] es un ser humano con habilidades cognitivas e inferenciales imperfectas. Puede algunas veces actuar de forma incongruente con sus creencias, pero la mayoría de las veces sus decisiones y deseos van de acuerdo con su conjunto de creencias. Estos dos presupuestos son el punto de partida para la racionalidad mínima. Otro elemento importante es que Cherniak establece un contraste entre racionalidad mínima y racionalidad idealizada. "Racionalidad idealizada" es una expresión que el autor usa para referirse a la concepción tradicional de racionalidad, la cual se caracteriza por describir a un ser humano ilimitado en sus habilidades cognitivas, tanto inferenciales como de memoria. En la concepción clásica de la racionalidad opera una idealización excesiva de cómo funciona la mente humana. La diferencia entre ambos conceptos radica en lo que Cherniak llama la dificultad finitaria (*the finitary predicament*), esto es limitaciones de memoria y de tiempo para razonar. En general, la dificultad finitaria es una propiedad de los seres humanos pero que en la versión idealizada de la racionalidad no se toma en cuenta. En cambio, Cherniak formula los requerimientos de racionalidad mínima tomando en cuenta esta característica.

Ahora bien, ¿cuáles son los requerimientos para la racionalidad mínima? Cherniak enumera una serie de condiciones aún vagas y programáticas pero que le permiten distinguir la racionalidad mínima de la ideal y, de paso, excluir ciertas exigencias excesivas sobre el agente. El primero es una condición general que consiste en describir el conjunto de acciones que un ser humano mínimamente racional debe realizar:

a) Si A tiene un conjunto particular de creencias y deseos, A puede realizar algunas, pero no necesariamente todas aquellas acciones aparentemente apropiadas. (Cherniak, 1986, p. 9)

   ¿Cuál es la diferencia entre la condición general de racionalidad mínima y aquella que se describe en la racionalidad idealizada? En la condición general de racionalidad idealizada lo que tenemos es un agente que puede realizar

[5]Es decir, una persona que cumple las condiciones mínimas para ser considerado un agente.

todas aquellas acciones que sean apropiadas para él. Es decir, tenemos a un agente capaz de actuar de una manera congruente con sus creencias y deseos todo el tiempo. Ahora bien, si pensamos en términos de atribuirle creencias a ese agente idealizado seguramente podríamos predecir su comportamiento de manera sencilla. Sabemos, primero, que siempre va a actuar de acuerdo con sus creencias y, segundo, su conjunto de creencias es consistente. No podríamos afirmar lo mismo de un agente mínimo el cual no siempre actúa de acuerdo con sus deseos.

Un segundo requerimiento para el agente mínimo es que tenga una cierta capacidad inferencial basada en una habilidad deductiva tal que cumple lo siguiente:

b) Si A tiene un conjunto particular de creencias y deseos, A podría hacer algunas pero no todas las inferencias de su conjunto de creencias que son aparentemente apropiadas. (Cherniak, 1986, p. 10)

Nuevamente se hace un contraste con la habilidad deductiva de un razonador ideal, pues este puede hacer todas las inferencias del conjunto. Nótese que esta formulación de la habilidad deductiva ideal nos lleva a un planteamiento teórico de la omnisciencia lógica. En última instancia el sistema de Hintikka supone que el agente puede inferir todas las consecuencias lógicas de sus creencias, pero también nosotros se las podemos atribuir. En los términos de la agencia mínima eso se supone está controlado al señalar que el agente mínimo hace solo algunas de estas inferencias. De nuevo, la condición es vaga, pero permite eliminar ciertas exigencias. Por ejemplo, cuando de su creencia en que el agua es H_2O deduce que el agua H_2O o el sistema solar tiene 7 planetas no es apropiada.

De la capacidad inferencial deductiva se desprenden otros dos requerimientos:

c) Requerimiento heurístico mínimo : A podría realizar algunas de las inferencias sólidas del conjunto de creencias que podría ser aparentemente apropiado hacer por A. (pag. 13)

d) Requerimiento deductivo mínimo : A debe tener éxito para realizar algunas de las inferencias aparentemente apropiadas que A ha emprendido.

El requerimiento heurístico le permite al agente elegir cuáles son las inferencias adecuadas del conjunto disponible. Un último requerimiento es el de consistencia mínima que señala lo siguiente:

e) Si A tiene un conjunto de creencias-deseos, entonces A podría algunas veces eliminar algunas de las inconsistencias que surjan en dicho conjunto. (Cherniak, 1986, p. 13)

Sería excesivo pedirle al agente que constantemente revise sus creencias y que elimine del conjunto aquellas que resulten inconsistentes con el resto pues eso puede requerirle un gasto cognitivo exagerado.

Consideremos ahora si las cualidades de la teoría de Cherniak nos dan una base sólida para aplicarse a la atribución de creencias. La teoría nos ofrece una concepción más realista del agente humano puesto que sí considera la dificultad finitaria. Me parece viable la idea de que el ser humano cuenta con una habilidad lógica que le permite hacer sólo algunas inferencias. Pero esta idea me parece problemática desde el punto de vista de la atribución de creencias ¿Cuáles creencias le puedo atribuir al sujeto? Tal vez las más obvias y prácticas pero aún así en las ideas de "algunas inferencias" y de "algunas creencias" hay bastante vaguedad. La teoría debería poder circunscribir un límite pasado el cual no podemos seguir considerando racional al agente que con su conducta lo rebasa. En ese sentido, la propuesta de Cherniak es insuficiente. Sin embargo, este autor puede replicar que ese límite, aunque vago, existe porque lo empleamos de manera implícita en nuestras prácticas cotidianas de atribución de creencias. Es decir, la teoría de Cherniak es una reflexión sobre la forma en que adscribimos creencias en nuestra vida diaria, pero las condiciones que impone a la racionalidad son solo necesarias y programáticas.

2.6. Atribución de creencias. Recapitulación

Ahora bien, ¿cuáles son los rasgos generales de esta teoría intuitiva de atribución de creencias? Enseguida enumero algunos que tomo de los dos mencionados autores, así como de Quine y Davidson. Algunos de ellos atañen a la construcción de sistemas formales; otros, en cambio, solo a su aplicación:

1) Nuestra teoría intuitiva presupone que el creyente no es lógicamente ignorante ni lógicamente omnisciente. Como dije antes, suele suponerse, por un lado, que todo lector de Euclides creerá que sus postulados son verdaderos, pero, por otro, que no creerá inmediatamente la verdad de todos sus teoremas. La prueba lógica, al menos concebida como instrumento de persuasión, presupone que el interlocutor, que al principio está convencido de la verdad de las premisas y no de la verdad de la conclusión, puede ser llevado a creer en esta precisamente por medio de la prueba. El ofrecer un argumento tiene como presuposición que el interlocutor no es lógicamente omnisciente (ignora al principio la conclusión), ni lógicamente ignorante (finalmente se convence de la verdad del teorema). Esta competencia lógica mínima está presupuesta también en nuestros códigos morales y legales. No es necesario que la ley especifique que matar a un individuo determinado está prohibido, basta con que explicite la interdicción general de privar de la vida a seres humanos. Por otro lado, se puede excusar a un legislador de no haber previsto una consecuencia lejana y recóndita de alguna ley que propuso.

2) La atribución de creencias se realiza a la par de otras atribuciones que también realizamos intuitivamente. Como hace notar Davidson esta adscripción está íntimamente ligada con la atribución de deseos e intenciones al agente, entre otras cosas. Puedo suponer que un individuo sabe conducir un automóvil (bajo ciertas circunstancias) a pesar de haber arrollado a una multitud, si a la vez le atribuyo la intención de realizar un acto terrorista. Sin embargo, además de intenciones y creencias otras actitudes están íntimamente vinculadas a ellas. Esta operación es hasta cierto punto subjetiva, pero también obedece a ciertos canones comunes gracias a los cuales nos ponemos de acuerdo en ciertos procesos (por ejemplo, los judiciales).

3) Las creencias, así como otras actitudes proposicionales, no pueden ser atribuidas individualmente. No se puede adscribir una creencia a un sujeto sin atribuirle otras y algunos deseos e intenciones. Al respecto retomo un ejemplo diseñado por Dennett en otro contexto pero que me sirve para ilustrar este punto. Imagina que, en un futuro, tal vez no muy remoto, fuese posible a través de técnicas de micro-cirugía cognitiva insertar en el cerebro de Tom la sola creencia (falsa) de que tiene un hermano mayor que vive en Cleveland. Ahora imaginemos que Tom está sentado en un bar y un amigo suyo le pregunta si tiene un hermano mayor que vive en Cleveland. Tom responde afirmativamente. El interlocutor le pregunta cómo se llama su hermano. Consideremos dos posibilidades verosímiles de cómo podría seguir la historia. Puede ser que Tom, al no encontrar la respuesta en su mente, dijera “perdón, no, no tengo ningún hermano mayor que viva en Cleveland, no sé por qué dije eso”. O pudiera ocurrir que dijera “no sé su nombre” y, presionado por otras preguntas acerca de su familiar, terminara diciendo “yo soy hijo único, pero tengo un hermano mayor que vive en Cleveland”. Concluye Dennett que el cirujano no ha conseguido insertar la creencia en la mente de Tom. Aunque él utiliza el ejemplo a otro propósito, a mí me sirve para mostrar mi punto: no puedo atribuir una creencia aislada a un individuo. Al decir que Tom cree tener un hermano mayor en Cleveland debo atribuirle algunas otras creencias sobre ese individuo, aunque sean creencias acerca de por qué Tom no sabe más de él. Una idea similar podemos encontrar en Davidson. El punto central es que una presuposición, incorporada en nuestra gramática, es que las creencias (junto con otras actitudes concomitantes que atribuimos) forman una estructura racional, con ciertas relaciones de coherencia mínima, (en parte lógicas, en parte semánticas y de sentido común) y, como ya dije, se atribuyen de acuerdo con criterios holísticos.

4) Tenemos una idea intuitiva de cuán compleja es una inferencia, de tal manera que algunas que nos parecen sencillas supondremos que nuestro interlocutor las ha hecho, mientras que, de otras, que consideramos más complejas no se las atribuiremos a un agente ordinario.

Si bien es cierto que esos criterios holíticos son subjetivos, hay algunos comunes que regulan muchas de nuestras prácticas sociales. Por ejemplo, alguien podría alegar ignorancia diciendo que no creía que la mordida de un canino pudiese trasmitir rabia. Si es un niño, no tendremos reparo en aceptar lo que dice. Si es un estudiante avanzado de medicina, diremos que está mintiendo.

Como dije, en esta perspectiva de la creencia la lógica aparece como constitutiva, en el sentido en que al sujeto que no procede con un mínimo de coherencia lógica, no podemos atribuirle creencias. Por ejemplo consideremos el siguiente esquema que puede verse como una manera de formalizar la omnisciencia lógica: $\mathcal{B}(\alpha \rightarrow \beta) \rightarrow (\mathcal{B}\alpha \rightarrow \mathcal{B}\beta)$ en donde, supongamos, β es una fácil consecuencia de α. En esta perspectiva dicha fórmula podría interpretarse como dictaminando que al individuo que cree α y no β no puede atribuírsele la creencia en $(\alpha \rightarrow \beta)$ o, bien, si por otros criterios manifiesta su creencia en este condicional, entonces la regla nos diría que no podemos atribuirle racionalidad. Sin embargo, si esto pasara en un caso concreto, podríamos pensar que el sujeto no es irracional sino que cometió un error en su desempeño cognitivo (tal vez por factores circunstanciales) y, aún si no fuese verosímil esa posibilidad, Cherniak aún admitiría que si el sujeto no falla en otras inferencias tendríamos argumentos para considerarlo racional. Es más, en algunos casos podemos inclusive suponer que el sujeto en cuestión sigue una lógica distinta a la nuestra. Por esta vaguedad en nuestras atribuciones es que no podemos obtener más que pocas directrices para formalizar una lógica constitutiva (y normativa) de la atribución de creencias.

Entre estas directrices está el que no podemos concederle al individuo la creencia en las consecuencias lógicas de todas sus creencias ni, por el contrario, podemos restringir sus creencias solo a aquellas que manifiesta directamente. Como veremos, también se aplicarán ciertos principios que establecen que si un individuo tiene ciertas creencias no debe tener otras (so pena de ser considerado irracional). En ello la lógica evidentemente juega un papel constitutivo.

En la siguiente sección trataré la cuestión de cómo la lógica se relaciona con nuestras creencias. Pueden distinguirse al menos tres posibilidades: **a)** la lógica describe la competencia razonante del individuo normal, como la gramática de una lengua describe la competencia lingüística del hablante promedio de esa lengua, **b)** la lógica norma la forma en que debemos inferir para acomodar coherentemente nuestras creencias a nueva información y **c)** la lógica es parte constitutiva de nuestra red de creencias (es decir, no se puede creer algo contrario a la leyes de la lógica). La primera puede ser fácilmente descartada. La lógica no cambia como lo hace la gramática según la época y la geografía. Sus verdades o reglas han sido casi siempre consideradas como eternas y universales. Me concentraré en los otros dos casos.

2.7. Lógica y Normatividad

La cuestión que voy a tratar en esta sección es la de si la lógica es una ciencia que rige la manera en que el ser humano debería razonar y contribuye así a la forma en que este debería formar sus creencias. A la respuesta afirmativa a esta pregunta se le conoce como la tesis de la normatividad de la lógica. Sus defensores tienen una larga tradición que los respalda, mientras que quienes se oponen a ella comparan a la lógica con cualquier otra ciencia que describe un aspecto o ámbito de lo real. Enseguida revisaré los argumentos a favor y en contra de la normatividad de la lógica. Para los primeros recurriré a John MacFarlane (2004) y a Graham Priest (2006). Para los del segundo tipo examinaré los argumentos propuestos por Gilbert Harman (Harman, 1986) y, más recientemente, por Gillian Russell (Russell, 2020). Finalmente sostendré las tres tesis siguientes:

1. La lógica es normativa, tal vez no por sí sola, sino combinada con otras normas que habitualmente damos por sentadas.

2. La omnisciencia lógica no es una idealización normativa y, como tal, no es tolerable.

3. La normatividad de la lógica puede proporcionarnos ciertos criterios para el desarrollo de la lógica epistémica.

2.7.1. Argumentos en Favor de la Normatividad de la Lógica.

Comencemos con los argumentos en favor de la tesis de normatividad de la lógica. Que la lógica nos dice cómo debemos razonar y no cómo, de hecho, lo hacemos ha sido una opinión bastante común en la historia de la filosofía. Podría mencionar a Arnauld y Nicole, Kant, Frege y a Priest, entre otros que han sostenido esta idea. Por ejemplo, Kant dice "en lógica no nos interesa cómo el entendimiento es y piensa ni cómo ha procedido hasta ahora, sino cómo debería proceder" (Kant, 1800, p. 4). Asimismo podemos encontrar muchos pasajes en la obra de Frege defendiendo la misma tesis. Por ejemplo, el siguiente:"Como la ética, la lógica también puede ser llamada una ciencia normativa. ¿Cómo debo pensar para alcanzar el objetivo, la verdad?"[6] (Frege 1979, p. 128).

Pero, además de estos recursos a la autoridad, ¿habría otro tipo de argumentos a los que se pueda apelar para sostener esta idea? ¿Qué razones pueden darse a favor de esta tradición? Consideremos estas tres evidencias:

1) El argumento del sentido común. Tenemos la tendencia censurar a quienes sostienen opiniones contradictorias o a quienes fallan al no sacar una consecuencia obvia de sus creencias.

2) El argumento de la demarcación. Este argumento es sugerido por MacFarlane (2004, p. 1). Si pensamos que la lógica se ocupa del pensamiento o de la inferencia, por ejemplo, entonces ¿por qué no es un fragmento de la psicología? La réplica ordinaria consiste en notar que esta última disciplina describe el pensamiento, pero la lógica lo norma, es decir, prescribe cómo debemos pensar o inferir para preservar la verdad. Es este caracter de la lógica el que es distintivo.

3) El argumento del incumplimiento (Priest, 2006, p. 179). Las leyes descriptivas, como las de las ciencias naturales, no pueden ser violadas. Las leyes de la lógica pueden ser infringidas. Por lo tanto, tienen que ser normativas.

El argumento del sentido común

Al primero puede objetarse que podemos censurar a quien viola las leyes de la lógica, pero siempre bajo el supuesto de que esa persona tiende a ciertos fines. Elogiamos al detective que usa la lógica para encontrar al asesino, porque creemos

[6]En otros párrafos Frege considera a la lógica como una cuencia descriptiva, pero no de la competencia razonante del ser humano. Ver Frege (1918). p. 289

que la justicia es un bien en sí. Análogamente, solemos desaprobar a un político que dice cosas aparentemente contradictorias porque suponemos que quiere ganarse la confianza de los ciudadanos. La química solo describe que cierta mezcla de elementos da lugar a una explosión. Un prejuicio en favor de la vida nos hace condenar a quien realiza tal combinación en un espacio habitado. Es decir, las ciencias por sí solas no son normativas, pero generan una normatividad cuando se persiguen ciertos fines. Aparentemente no es aventurado considerar de la misma forma a la lógica. En ese sentido la lógica sería descriptiva como lo es la química.

Como yo defenderé que la lógica es normativa, intentaré mostrar brevemente por qué considero que la objeción anterior es incorrecta. Frege equiparó la relación que tiene la estética con la belleza, a la que mantiene la lógica con la verdad. Es cierto que puede construirse una estatua espléndida para halagar a un poderoso, pero se puede sostener que la obra tiene un valor intrínseco por su hermosura. Análogamente aunque la verdad puede ser útil a otros fines, es sensato concederle un valor propio. En muchos casos censuramos a quien falla al seguir las reglas de la lógica porque pone en riesgo la verdad de sus aseveraciones. Es decir, reprobamos su conducta aun sin considerar sus objetivos. Por supuesto que esto podría dar lugar a una larga discusión, pero creo que el punto de vista que sostengo es razonable y desarticula la réplica anterior.

El argumento de la demarcación

Podría contrarrestarse el argumento expresado en 2) recurriendo a una réplica similar a la que Sócrates le da Gorgias (Platón 1980, p. 5) cuando este le dice que la retórica es el arte de persuadir. Sócrates contesta que todas las artes que pueden enseñarse producen persuasión. El médico convence sobre cuestiones de salud y, el arquitecto, sobre la construcción de casas. Kant distinguía entre la lógica general y las lógicas especiales atinentes a cada ciencia. Si se trata de cómo debemos inferir en materia de fuerzas, supuestos ciertos fines, es mejor recurrir a la física. Así, por ejemplo, la física nos autoriza a pasar de las premisas que nos dan la masa, velocidad y el radio de la trayectoria circular de un cuerpo a la conclusión de cuál es su momento angular. De la misma manera podemos considerar a la aritmética como una lógica que nos permite hacer transiciones entre valores numéricos asignados a un fenómeno. En ambos casos el valor verdadero de las premisas (si lo tienen) se transmite a la conclusión. Bajo esta perspectiva la lógica es tan normativa como la química. Su normatividad no la distingue. A ello podría contestarse que, puesto que no tiene objeto particular, la lógica es normativa del pensamiento (o de la inferencia) en cualquiera de sus aplicaciones. Sin embargo, esto podría ponerse en duda. ¿Realmente la lógica no tiene un tema particular? Para Gillian Russell (2017) la lógica tiene por objeto propio ciertas relaciones entre los portadores de verdad (sean estos proposiciones o tipos de proferencias o pensamientos fregeanos) cuyas relaciones atinentes a la preservación de la verdad describe.

Mi respuesta es que la réplica anterior es artificiosa. Aunque una verdad de la química pueda ser útil al hacer una transición inferencial, es una exageración considerarla una ley del pensamiento. Podría alguien pasar su vida entera prescindiendo de esta "regla" sin perjuicio suyo. Es una verdad provincial. En cambio, las leyes lógicas son indispensables incluso para el descubrimiento o corrección de las leyes químicas. Aun si no hubiese nada en el mundo, la lógica sería de utilidad en la búsqueda de la verdad. En ese sentido, sus reglas son normas para el pensamiento.

El argumento del incumplimiento

En lo atinente a 3), hay un sentido en que las leyes naturales sí son violadas. En esos casos solemos preservar la ley y desechar la excepción suponiendo que hay factores ignotos que explican la aparente falla o que hay leyes más generales, tal vez aún sin descubrirse, que subsumen el fenómeno en cuestión. Es decir, aún en esos casos, preservamos la verdad de las leyes naturales. En todo caso, estas juegan un papel muy diferente a las jurídicas y eso es lo que la comparación de Priest destaca. Así es que el argumento 3) puede ser sostenido.

Si mi posición con respecto a la normatividad parece muy radical, puedo matizarla ligeramente de la siguiente manera. G. Russell y Steinberger sostienen que la lógica es normativa cuando se combina con otras normas. Yo diría que lo es cuando se combina con la norma de la verdad, que es una norma que damos por sentada. Esta norma establece que:

"Para cualquier proposición P, es correcta la creencia de que P si y solo si P es verdadera"

Es decir, la verdad de P otorga una virtud a la creencia en P. Yo no entraré ahora en el debate de si esta norma forma parte esencial del concepto de creencia o es un rasgo accidental. Velleman (2000) suscribe el primer punto de vista:

> "Yo tomo como una verdad conceptual que las creencias son correctas cuando son verdaderas e incorrectas cuando son falsas. Es

> más, yo no pienso que el defecto en las creencias pueda consistir en su tendencia a extraviar nuestra conducta... aún algunas creencias falsas pueden dirigirnos bastante bien. Las creencias falsas son defectuosas en sí mismas antes e independientemente de consecuencias prácticas adversas" (Velleman, 2000, pp. 277-8)

Otra posibilidad, como lo menciona Velleman, sería sostener que la creencia verdadera es correcta porque nos conduce a fines deseables, por ejemplo, a la supervivencia de la especie. Llamemos al primer punto de vista deontológico; al segundo, consecuencialista. Una polaridad similar podemos hallar con respecto al razonamiento. De acuerdo con el acercamiento deontológico, lo que es constitutivo del buen razonamiento es solamente razonar en concordancia con ciertas reglas o principios. En esta perspectiva un razonamiento es adecuado por el simple hecho de seguir el conjunto de reglas que determina lo que significa razonar correctamente. No se toman en cuenta los fines ulteriores a que esta forma de razonar conduce. La racionalidad en sí misma tiene un valor intrínseco. Por otro, el acercamiento consecuencialista sostiene que razonar correctamente es hacerlo de manera tal que verosímilmente se obtengan ciertas metas o resultados. En esta perspectiva lo aceptable se define a partir de las consecuencias. Las reglas que para ello se sigan no son determinantes (Foley, 1993). En el caso de la inferencia, ambos acercamientos conducen a otras dificultades. Por el lado deontológico subsiste el problema de la fuente de nuestro conocimiento de las reglas correctas, dado que no están determinadas por una finalidad ulterior. Por el lado consecuencialista, ¿juzgaremos la calidad de un argumento hasta saber los resultados que produjo? Parece absurdo. Más bien, una forma de inferencia será juzgada correcta si, a partir de la experiencia o de otras intuiciones, podemos *inferir* que tendrá consecuencias favorables. Pero entonces se cae en una circularidad. Las señalo solo como dificultades, no necesariamente como obstáculos infranqueables.

Me parece que hay otra opción que, aunque no exenta de dificultades, es atractiva. Podemos adoptar una postura consecuencialista respecto a la lógica y un enfoque deontológico con respecto a la norma de la verdad: nos conviene seguir los preceptos lógicos porque conducen a creencias verdaderas que tienen un valor en sí mismas (como lo dice la cita de Frege más arriba).[7]

Apunto otro argumento que parecería conducir a que la lógica tendría que ser normativa. Recordemos que Aristóteles dice que se da un silogismo cuando sentadas ciertas cosas, otras se siguen exclusivamente de ellas por necesidad (An. pr.I

[7]Sostener que la verdad conduce a la supervivencia de la especie o del individuo no parece acertado. Ver García (2009), Stein (2012)

24 b, 18-23). Esta modalidad ha dado pie a mucha discusión, pero una posibilidad es interpretarla como un imperativo epistémico. Al menos, así lo hace Quine quien dice: "De entre todas estas relaciones de enunciados a enunciados hay una de especial importancia: la relación de un enunciado a todo lo que se sigue lógicamente de él. Si un enunciado *debe ser mantenido* como verdadero, cada enunciado implicado por él también *debe ser sostenido* como verdadero" (Quine 1950, 1972, pp. 3-4) En este caso, la normativdad sería esencial a la lógica[8], aunque, ciertamente, esta interpretación de la consecuencia lógica requeriría una justificación.
.

Yo supondré la norma de la verdad y con ello la norma: "deberíamos esforzarnos por tener una representación correcta del mundo y, por lo tanto, por tener creencias verdaderas". Ahora bien, si pretendemos creer solo o mayormente lo verdadero, debemos atender a la lógica, que se encarga de estudiar la preservación de la verdad a través de la forma. Aunque ambas ideas, la referente al tema de la lógica y la norma de la verdad, han sido cuestionadas[9] (como casi todo en filosofía), intuitivamente parecen correctas, como asimismo, la tesis que de ellas se deriva, a saber, que la lógica, aunque en sí misma podría no ser normativa, combinada con normas (como la de la verdad), genera proposiciones normativas. Este es el matiz que deseaba introducir en mi posición.

2.7.2. Argumentos Contra la Normatividad de la Lógica

Ahora expondré tres argumentos dados contra la normatividad de la lógica. Del primero solo me ocuparé brevemente. Los otros dos han dado materia a un largo debate que analizaré con detalle.

1) (De Bruin, 2008, p. 124) El adoptar una postura normativista de la lógica nos llevaría a un absurdo. ¿Cómo la lógica podría ordenarnos creer algo, si tener una creencia no es algo voluntario? Además [la lógica epistémica] nos ordenaría hacer de esa creencia conocimiento, en particular, esforzarnos por hacerla verdadera.

   La primera crítica de De Bruin depende cómo seleccionemos el operador deóntico que constituiría la norma lógica. De Bruin elige "ordenar" y parece interpretarlo como la fuerza que ejercería un capataz contra un subalterno rebelde.

[8] Las itálicas no están en la cita original. En realidad, Quine enuncia el esquema de un principio puente, como diré más adelante

[9] Con respecto a la norma de la verdad, véase la discusión en Chan (2014)

Sin embargo, hay un sentido en que la tortuga de Lewis Carroll sí está obligada a creer la conclusión, y un sentido en que Smith sí tiene justificación para creer que el hombre con las diez monedas en el bolsillo obtendrá el puesto. Lo mismo puede decirse respecto al párrafo de Hintikka ya citado: "... las leyes de la lógica no son necesarias en el sentido de ser inevitables. No son verdades que *debamos* saber sino verdades que *podemos* saber sin hacer uso de ninguna información fáctica... no son leyes del pensamiento en el sentido de leyes naturales... [ni] mandatos, excepto quizás del pensamiento más agudo posible[10] "(Hintikka 1962, Pag. 37). No me parece que provea una defensa de la postura descriptivista. Si la lógica nos proporciona verdades que *podemos*, aunque no *debamos*, saber (sin información fáctica adicional), entonces, en algún sentido sí es normativa. Dadas ciertas premisas que el agente cree, la lógica lo autoriza a tener otras; o bien, le recomienda tenerlas, si quiere ser suficientemente agudo. Si esto es así, entonces la lógica no es una ciencia que describe simplemente cómo razonamos los seres humanos. La segunda parte de la objeción de De Bruin no atañe a la creencia, sino al conocimiento y, por ello, no me concierne.

2) Los enunciados verdaderos de la lógica (digamos las tautologías) no tienen ningún término deóntico. No son imperativos sino enunciados con un valor de verdad.

3) Si interpretamos los enunciados verdaderos de la lógica como normativos, siguiendo nuestras intuiciones, resultan en normas absurdas.

Los dos argumentos 2) y 3) anteriores, con algunos matices, fueron planteados por Harman (1986) y por Goldman (1986) en los años ochentas del siglo pasado. El segundo fue desarrollado principalmente por Harman. Expondré y analizaré el debate subsecuente de que fueron objeto.

Insisto en que lo que diré enseguida parece ateniente a la lógica en general, pero una de las aportaciones de este trabajo será mostrar que las variantes epistémicas y doxáticas de la lógica tienen conexión con los tópicos siguientes. Recordemos que, en los sistemas formales de Hintikka, si β es consecuencia lógica de α, entonces tenemos como teorema $(\mathcal{B}\alpha \rightarrow \mathcal{B}\beta)$. ¿Debemos leer esta fórmula como describiendo a un razonador ideal o como normando el modo en que debemos adquirir creencias? Vimos que los sistemas normales (extensiones de $\mathbf{K}$) atribuyen al agente toda nuestra lógica clásica, así es que suponen que esta es adecuada para

[10]Las itálicas no están el original

modelar alguno de esos dos casos, pero si los interpretamos como descripciones de las creencias del agente son claramente falsos. Por ello, la pregunta sobre la normatividad de la lógica resulta esencial para mi tema.

La falta de operadores deónticos

En 1986, dos autores muy influyentes, Alvin Goldman (1986, p. 82) y Gilbert Harman (1986), pusieron en duda que la lógica fuese prescripitva del razonamiento humano. Ambos comienzan por la observación trivial de que las proposiciones de la lógica no contienen términos doxásticos ni deónticos. Se refieren a la lógica en general (digamos clásica). Sin embargo, como ya dije la observación sigue siendo pertinente respecto a la lógica de Hintikka por al menos dos razones: a) El sistema $\mathbf{S_5}$ (y en menor medida también $\mathbf{K}$ y, por lo tanto, todos los sistemas normales) concede al agente el conocimiento completo de nuestra lógica. En segundo término, ninguno de estos sistemas contienen operadores deónticos. Aun si aceptamos, por ejemplo, $\mathbf{S_5}$, como un sistema que debe normar las creencias del agente, y no simplemente describirlas, queda por resolver el punto de cómo interpretar sus fórmulas como prescriptivas.

Si consideramos un sistema de lógica clásica presentado semánticamente, en efecto, encontraremos establecidas ciertas relaciones entre proposiciones o conjuntos de proposiciones, tales como consecuencia lógica, consistencia, compacidad, etc. Como observa G. Russell (2020, pp. 382-3), ningún operador deóntico figura allí. Por otro lado, si consideramos la presentación de teoría de la prueba en su versión más canónica, a saber, el cálculo de deducción natural, podríamos sostener que consiste en reglas para inferir o deducir, las cuales pueden ser consideradas como conteniendo implícitamente operadores deónticos, aunque no esté claro de qué tipo sean. Parecerá natural leer, por ejemplo, la regla de eliminación de la conjunción como autorizándonos a aseverar P si ya aseveramos $(P \wedge Q)$. O tal vez como afirmando que si tenemos razones para afirmar $(P \wedge Q)$ también las tenemos para afirmar P. Si se trata de la lógica intuicionista este punto parece más evidente. Recordemos que en la interpretación Brouwer-Heyting-Kolmogorov, la misma regla puede ser leída como garantizando que si tenemos una prueba de $(P \wedge Q)$, asimismo la tenemos de P.

Harman reconocería este punto, pero ello no lo forzaría a cambiar su tesis. El distingue (aunque no con estos términos) el razonamiento formal del informal. Puede ilustarse el primero con un libro de matemáticas que procede de modo deductivo. El lector constata que de un conjunto de proposiciones otras se pueden inferir, pero en ningún momento se pretende o se requiere que crea esas proposiciones. El razonamiento informal, en cambio, es el que usamos a diario y que nos conduce a modificar nuestro conjunto de creencias a la luz de nueva evidencia o de algunas consideraciones. Para este proceso -dice Harman- la lógica no tiene relevancia: "Mi conclusión es que no hay ningún modo claramente significativo en el cual la lógica sea especialmente relevante para el razonamiento [informal]" (Harman, 1986, p. 20). Las siguientes objeciones caen (también) bajo el rubro 3) de la sección anterior, es decir, la observación de que si interpretamos de manera intuitiva las proposiciones verdaderas de la lógica como normativas los resultados son normas absurdas. Con ello veremos, de paso, que la lógica epistémica estándar no describe a un ser idealmente racional.

Constreñimiento

La segunda observación que hacen tanto Harman como Goldman es que aunque Q implique P, y el agente lo reconozca, no siempre sería correcto (racional) que si cree Q crea P:

> "Suponga que *q* implica *p* y que S cree *q*, ¿se sigue que S debe creer que *p* o que le esté permitido creer que *p*? En absoluto... ¡tal vez lo que S debería hacer, una vez que advierte que *q* implica *p*, es abandonar su creencia en *q*! Después de todo, a veces aprendemos cosas que nos hacen aconsejable desechar creencias que teníamos". (Goldman, 1986, p. 83).

Para referencia posterior denominaremos a esta observación "constreñimiento". Las tres siguientes son objeciones de Harman a la tesis de que la lógica es relevante al razonamiento humano, entendido "razonamiento" como dije antes.

Demandas excesivas

Una alude precisamente a la dificultad finitaria y a la observación de que aquel que no puede algo, no debe hacerlo (observación que suele atribuirse a Kant). Se le llama "demandas excesivas": es imposible para el agente creer las consecuencias lógicas de todas sus creencias porque son infinitas. Por lo tanto, la lógica no puede obligarlo a hacerlo. Como vimos, ésta es la observación más común contra los sistemas de Hintikka. Suele formularse diciendo que el sujeto modelado por esos sistemas es ideal (veremos enseguida que esta última afirmación es incorrecta y nos llevaría por un camino equivocado). La objeción descansa en la idea de que la creencia tiene un cierto contenido proposicional o representativo y que ocupa un cierto “espacio en el almacén de la memoria”. Dado el caracter finito de nuestros sistemas cognitivos, este almacen tiene limitaciones de espacio. Si entendemos la creencia de manera representacional o en un sentido disposicional, entonces también debemos considerar limitado nuestro conjunto de creencias, al menos si seguimos la autoridad de Kripke (1989, p. 32). Incluso si pensamos la creencia en tanto que es atribuida, es claro que las creencias que atribuimos a un agente no son cerradas bajo consecuencia lógica.

Trivialidades

La siguiente objeción de Harman es que, aunque un individuo contara con una memoria indefinidamente extensible que le permitiera almacenar cualquier cantidad de información, no sería racional o aconsejable que creyera en ciertas consecuencias triviales de sus creencias. Una vez que cree que la tierra es redonda ¿de qué manera podría serle útil o conveniente creer la disyunción “la tierra es redonda o Venus es el planeta más alejado del sol”? Claramente, debe conservar solo algunas creencias que son pertinentes a sus intereses. Se llama a esta objeción “trivialidades” y muestra cómo el problema de los sistemas de Hintikka no es que modelen a un razonador ideal.

Inconsistencia

La última objeción de Harman (llamada “inconsistencia”) es la observación de que a veces la inconsistencia es una opción más racional que la consistencia, como lo había señalado Cherniak (1986), y como se ilustra en la paradoja del prefacio. Supongamos que un entomólogo escribe un tratado muy completo sobre los lepidópteros. Todo lo que en él dice está sólidamente respaldado. Es decir, cree cada una de las oraciones del libro. Sin embargo, sabe por experiencia que suele equivocarse y que aún en ocasiones en que ha estado muy convencido, algo de lo que ha escrito ha resultado falso. Por ello, a la hora de escribir el prefacio le aclara al lector que seguramente habrá alguna parte del texto que no resulte verdadera. Esto también es algo que él cree sinceramente. Así es que sabe que su conjunto de creencias es inconsistente, es decir, que tiene una contradicción como consecuencia, pero es incapaz, al menos por el momento, de decidir a qué creencia o creencias debe renunciar. Adviértase que, además, es racional y cuerdo proceder así, porque de todo lo que escribió (incluido el prefacio) tiene fuerte evidencia. Es probable que puedan atestiguarse ejemplos en la historia de la ciencia en que una teoría inconsistente es aceptada a título provisional, pero se ha puesto en duda que una teoría científica pueda ser creída. Seguir esta cuestión me alejaría de mi tema.

Rigor

Esta es una exigencia que agrega Broome (1999), más que una objeción. La idea es que la normatividad que la lógica impone no es laxa, sino exigente y rigurosa. Por ejemplo, a veces condenamos la conducta de alguien porque pone de manifiesto que no sacó una consecuencia de sus creencias que era obvia. La lógica garantiza con toda certeza la verdad de un argumento correcto con premisas verdaderas. Llamaremos "rigor" a esta condición.

Más adelante aparecerán otras objeciones contra la normatividad de la lógica o, más precisamente, argumentos en favor de la tesis de que la lógica no es especialmente relevante al razonamiento humano (informal).

2.8. Los Principios Puente

Como observa Florian Steinberger (2020), una posibilidad sería aceptar los argumentos y la conclusión de Harman. Sin embargo, la opción más favorecida en la literatura a este respecto, y por la que yo me inclino, es aceptar las objeciones anteriores y eludir la conclusión "escéptica". Para ello, la estrategia seguida por importantes autores parte de la observación, hecha tanto por Goldman como por Harman, de que los enunciados de la lógica no tienen operadores deónticos. La idea es que efectivamente la lógica no cuenta con proposiciones normativas, pero cada enunciado de consecuencia lógica podría convertirse de alguna manera sistemática en un principio de razonamiento. El mismo Harman muestra que algunos enunciados de consecuencia interpretados de una cierta manera dan lugar a proposiciones normativas que, en ciertas circunstancias, son absurdas. De $\models (\alpha \rightarrow \beta)$ y de que el agente lo reconoce (lo que podemos formalizar así $\models \mathcal{B}(Cons(\alpha, \beta))$, o bien, en una versión más débil $w \models \mathcal{B}(\alpha \rightarrow \beta)$ donde w representa una situación particular) no podemos deducir $w \models \mathcal{B}(\alpha) \rightarrow \mathcal{O}\mathcal{B}(\beta)$ donde $\mathcal{O}$ es un operador deóntico de obligación o de conveniencia. De sus objeciones Harman concluye que no hay manera verosímil de interpretar los enunciados válidos de la lógica como prescriptivos de lo que una persona racional debe hacer en cualquiera de los casos descritos en el antecedente.

Varios autores, MacFarlane (2004), Field (2009), Sainsbury (2002), Steinberger (2020), entre otros, se dieron a la tarea de responder al reto de Harman partiendo de sus premisas. Emprenden la búsqueda de lo que el primero llamó "principios puente”. Se trata de principios normativos de la creencia que se deriven, de manera uniforme, de enunciados de consecuencia lógica. Una idea similar fue propuesta en epistemología por Samuels, Stich y Faucher (2004, p. 38) bajo el nombre de “reglas de conversión” (aunque su propuesta abarca, además de los principios de la lógica, algunos de la probabilidad). MacFarlane y Field intentan, a través del análisis y refinamiento de los principios puente, avanzar en la determinación de qué logica es la correcta. Formulado de esta manera parece presuponer un monismo lógico muy estricto. De manera menos contenciosa podemos decir que se trata de ver de si en la consecuencia lógica hay, por ejemplo, un elemento de relevancia o de necesidad, etc. Sainsbury (2002) dedica menos espacio a la cuestión de los principios puente que a tratar de mostrar que hay un terreno común en que los proponentes de lógicas diversas pueden debatir. Yo, en cambio, me enfocaré solamente en la cuestión de qué principios puente podrían considerarse correctos o, al menos, de cuáles deben descartarse, y luego en la forma en que esta discusión puede arrojar luz en la construcción de sistemas de lógica doxástica.

2.8.1. La propuesta de John MacFarlane

John MacFarlane (2004) hizo una clasificación de los principios puente que, aunque dista mucho de ser exhaustiva, nos será de gran utilidad. Podemos partir de un enunciado de consecuencia lógica del tipo $(\alpha, \beta \models \gamma)$, y derivar de allí un enunciado normativo relativo a las creencias (u otras actitudes proposicionales) en α, β y γ. Para ilustrarlo recordemos el ejemplo que usa Harman.

Si γ es consecuencia lógica de α y β, entonces, si el sujeto cree α y β, entonces debe creer γ.

Al condicional en itálicas lo llamaré “principal” y “subordinado” al otro. En este caso tenemos el operador deóntico de obligación, afectando solo al consecuente del condicional subordinado y se refiere a la creencia en γ y no a la prohibición de creer la negación de γ. Más específicamente MacFarlane identifica los parámetros siguientes:

Modalidad Deóntica

obligación	permiso	razón
$\mathcal{O}$	$\mathcal{P}$	$\mathcal{R}$
Debe (creer)	Le está permitido	Tiene razones

1). La modalidad deóntica:

- $\mathcal{O}$) obligación, tal como *debe* (creer),
- $\mathcal{P}$) permiso, *le está permitido* (creer), o
- $\mathcal{R}$) razón, *tiene razones* (para creer).

Lo que distingue a esta última modalidad es que las razones en cuestión pueden ser superadas por otras. Es decir, yo puedo tener razones para no votar, pero también otras más poderosas para emitir mi voto.

2). La polaridad, que puede ser:

- **+**) positiva, o
- **-**) negativa.

Por ejemplo, creer α y β podría darme razones para creer γ o simplemente para no creer la negación de γ. MacFarlane no se refiere aquí explícitamente a la negación sino a *disbelief*, pero aclara que no se trata de la suspensión de la creencia. Es decir, no se refiere simplemente a no creer γ, sino aparentemente a creer la negación de γ, solo que no quiere utilizar aquí un conectivo lógico tan complejo, a saber, la negación, por dos razones. Una es que se propone seguir una metodología con la que piensa dirimir desavenencias entre lógicos de diferentes escuelas. Así es que debe evitar en lo posible el uso de conectivos lógicos susceptibles de ambigüedad. La otra es que para un dialetista no sería aceptable la identificación entre "descreer γ" y "creer no $\neg\gamma$"[11]. Sin embargo, podemos dejar de lado estas reservas en pro de una mayor claridad y aceptar, provisionalmente al menos, esta identificación.

3). El alcance del operador deóntico.

Puede afectar:

- $\mathfrak{C}$) solo al consecuente del condicional subordinado,
- $\mathfrak{B}$) al antecedente y al consecuente, o
- $\mathfrak{T}$) al condicional (subordinado) completo.

[11]Porque para el dialetista es posible creer γ y creer su negación, pero no lo es creer y descreer la misma proposición

Si partimos del mero hecho de consecuencia $\alpha, \beta \models \gamma$ obtenemos 18 principios. Doy a continuación algunos ejemplos de cómo queda el condicional subordinado con la combinación de las modalidades señaladas:

- $\mathfrak{C}\mathcal{O}+$ Si crees α y crees β, deberías creer γ.
- $\mathfrak{C}\mathcal{O}-$ Si crees α y crees β, deberías no creer no-γ.
- $\mathfrak{C}\mathcal{P}+$ Si crees α y crees β, puedes creer γ.
- $\mathfrak{C}\mathcal{P}-$ Si crees α y crees β, se te permite no creer no-γ.
- $\mathfrak{B}\mathcal{R}+$ Si tienes razones para creer α y creer β, tienes razones para creer γ.
- $\mathfrak{T}\mathcal{O}+$ Deberías ver que si crees α y crees β, tu creas γ.
- $\mathfrak{T}\mathcal{P}+$ Puedes ver que si crees α y crees β, tu creas γ.

Si representamos por $\mathcal{B}$ la modalidad doxástica y por $\mathcal{O}$, $\mathcal{P}$ y $\mathcal{R}$ las modalidades deónticas mencionadas, podemos representar las 18 modalidades de la siguiente manera:

$\mathfrak{C}\mathcal{O}+$	$(\mathcal{B}\alpha \wedge \mathcal{B}\beta) \to \mathcal{O}\mathcal{B}\gamma$
$\mathfrak{C}\mathcal{O}-$	$(\mathcal{B}\alpha \wedge \mathcal{B}\beta) \to \mathcal{O}\neg\mathcal{B}\neg\gamma$
$\mathfrak{C}\mathcal{P}+$	$(\mathcal{B}\alpha \wedge \mathcal{B}\beta) \to \mathcal{P}\mathcal{B}\gamma$
$\mathfrak{C}\mathcal{P}-$	$(\mathcal{B}\alpha \wedge \mathcal{B}\beta) \to \mathcal{P}\neg\mathcal{B}\neg\gamma$
$\mathfrak{C}\mathcal{R}+$	$(\mathcal{B}\alpha \wedge \mathcal{B}\beta) \to \mathcal{R}\mathcal{B}\gamma$
$\mathfrak{C}\mathcal{R}-$	$(\mathcal{B}\alpha \wedge \mathcal{B}\beta) \to \mathcal{R}\neg\mathcal{B}\neg\gamma$
$\mathfrak{B}\mathcal{O}+$	$(\mathcal{O}(\mathcal{B}\alpha \wedge \mathcal{B}\beta) \to \mathcal{O}\mathcal{B}\gamma$
$\mathfrak{B}\mathcal{O}-$	$(\mathcal{O}(\mathcal{B}\alpha \wedge \mathcal{B}\beta) \to \mathcal{O}\neg\mathcal{B}\neg\gamma$
$\mathfrak{B}\mathcal{P}+$	$(\mathcal{P}(\mathcal{B}\alpha \wedge \mathcal{B}\beta) \to \mathcal{P}\mathcal{B}\gamma$
$\mathfrak{B}\mathcal{P}-$	$(\mathcal{P}(\mathcal{B}\alpha \wedge \mathcal{B}\beta) \to \mathcal{P}\neg\mathcal{B}\neg\gamma$
$\mathfrak{B}\mathcal{R}+$	$(\mathcal{R}(\mathcal{B}\alpha \wedge \mathcal{B}\beta) \to \mathcal{R}\mathcal{B}\gamma$
$\mathfrak{B}\mathcal{R}-$	$(\mathcal{R}(\mathcal{B}\alpha \wedge \mathcal{B}\beta) \to \mathcal{R}\neg\mathcal{B}\neg\gamma$
$\mathfrak{T}\mathcal{O}+$	$(\mathcal{O}((\mathcal{B}\alpha \wedge \mathcal{B}\beta) \to \mathcal{B}\gamma)$
$\mathfrak{T}\mathcal{O}-$	$(\mathcal{O}((\mathcal{B}\alpha \wedge \mathcal{B}\beta) \to \neg\mathcal{B}\neg\gamma)$
$\mathfrak{T}\mathcal{P}+$	$(\mathcal{P}((\mathcal{B}\alpha \wedge \mathcal{B}\beta) \to \mathcal{B}\gamma)$
$\mathfrak{T}\mathcal{P}-$	$(\mathcal{P}((\mathcal{B}\alpha \wedge \mathcal{B}\beta) \to \mathcal{B}\gamma)$
$\mathfrak{T}\mathcal{R}+$	$(\mathcal{R}((\mathcal{B}\alpha \wedge \mathcal{B}\beta) \to \neg\mathcal{B}\neg\gamma)$
$\mathfrak{T}\mathcal{R}-$	$(\mathcal{R}((\mathcal{B}\alpha \wedge \mathcal{B}\beta) \to \neg\mathcal{B}\neg\gamma)$

Dos aclaraciones más se requieren. La primera es que, como ya dije, “deberías no” no se refiere a la ausencia de una obligación, sino más bien a una prohibición y por ello no la traduzco por la locución más natural, pero también más ambigüa, “no deberías”. La segunda es que las modalidades $\mathcal{P}$ y $\mathcal{O}$ no son interdefinibles. El permiso para creer γ no es lo mismo que la ausencia de obligación para no creer γ. Más bien es una suerte de autorización o facultad. Cuando decimos que el juez puede casar a los novios, no solamente decimos que no le está prohibido hacerlo.

Desde luego que otros muchos principios puente podrían proponerse, por ejemplo, dejando que el operador deóntico del consecuente en los principios de tipo $\mathfrak{C}$ y $\mathfrak{B}$ cayera bajo el alcance de la negación, pero no se trata de dar una lista exhaustiva, sino precisamente de ofrecer una ayuda a la imaginación.

Observa MacFarlane que podemos duplicar esos principios agregando la modalidad K, es decir, agregando principios que parten no simplemente de que γ es consecuencia lógica de α y β, sino que, además, el individuo lo sabe. Por ejemplo, el principio $\mathsf{K}\mathfrak{C}\mathcal{O}+$ es:

Si el sujeto sabe que γ es consecuencia lógica de α y β, entonces, si el sujeto cree α y β, entonces debe creer γ.

En ese caso, no cualquier condicional válido genera una obligación, permiso o justificación en las creencias del agente sino solo aquellos que el individuo sabe. Como dirá MacFarlane, esto suscita una dificultad[12] .

Desde luego que también podríamos tener principios que partan de la creencia en un hecho de consecuencia lógica del tipo:

Si el sujeto cree que γ es consecuencia lógica de α y β, entonces, si el sujeto cree α y β, entonces debe creer γ.

Estrictamente hablando, este no es un principio puente, pues no parte de un hecho de consecuencia lógica, pero será importante considerarlo dado que es de la lógica doxástica que aquí se trata.

Ni Field ni Steinberger se quedan con ninguno de esos principios, lo que muestra la insuficiencia de esta clasificación. Sin embargo, nos sirve para descartar rápidamente varias posibilidades y sugerir otras nuevas. Note que, por ahora, no se requiere distinguir entre principios diacrónicos y sincrónicos, excepto que, en el caso de los principios K, es después de que el individuo (re)conoce el hecho de consecuencia (suponiendo que esto ocurra en el tiempo) que adquiere obligaciones, permisos o razones para creer.

Algunas de las ideas de MacFarlane provienen de un artículo de John Broome (1999). Una de ellas es justamente la de distinguir los alcances de los operadores deónticos en proposiciones normativas en que figura un condicional. El tomar el alcance amplio del mismo (es decir, adoptar las versiones $\mathfrak{T}$ de los principios puente) anula la primera objeción de Harman (constreñimiento). Recordemos que esta era la observación de que no es cierto, en general, que un individuo deba creer γ, porque cree α y β[13]. La crítica de Harman se dirige contra las versiones $\mathfrak{C}$ de los principios puente, es decir las que tienen el operador deóntico en el consecuente del segundo condicional. Broome presenta un ejemplo particular que hace la objeción más punzante contra las versiones positivas de estos principios. Evidentemente γ se sigue de γ. Por lo tanto, los principios $\mathfrak{C}$ positivos, dirían que si el agente cree γ, entonces tiene razones para creer γ, o está obligado o autorizado a creer γ, lo cual evidentemente es falso. Alguien puede creer que tiene derecho a violar la ley, pero no por eso debe hacerlo. Otro puede creer en Dios, sin tener una justificación racional para hacerlo. Veremos en los siguientes capítulos que esta no es la única forma de evitar la objeción del constreñimiento.

[12]Por ejemplo, en el caso de la obligación, la lgica obligaría solo al que reconoce el principio lógico en cuestión.

[13]Siendo α, β y γ tales que $\alpha, \beta \models \gamma$, lo que supondré en lo sucesivo

Los principios $\mathfrak{B}$, es decir, los que contienen un operador modal tanto en el antecedente como en el consecuente, parecen más adecuados. Si S debe creer α y creer β, entonces parecería que, en algún sentido, también debe creer γ. Sin embargo, si es muy difícil percatarse de que γ se deriva de α y β, la regla podría estar obligando al sujeto a algo de lo que es incapaz (al menos por un tiempo). Por lo mismo tampoco parece del todo correcto el principio $\mathfrak{B}\mathcal{O}-$ que prohibe al individuo en cuestión la creencia en no-γ. Aquí cabría la posibilidad, no mencionada por MacFarlane, de pricipios "mixtos", es decir, en que el antecedente y el consecuente tuviesen diferentes modalidades. Ejemplo de tales es: si S debe creer α y creer β, entonces le está permitido creer γ.

El problema recién mencionado (las limitaciones cognitivas del agente) se resuelve o mitiga adoptando las versiones $\mathbf{K}$ correspondientes. Ya no pedimos al agente que sepa las consecuencias lógicas de todo lo que cree, sino solo en el caso en que reconozca las implicaciones correspondientes. Pero MacFarlane tiene otra objeción más general contra los principios $\mathfrak{B}$ que también se aplica al caso anterior. El problema es que parten de proposiciones que tienen una modalidad en el antecedente que se transmite al consecuente. Si las creencias de que el individuo parte tienen cierta características (por ejemplo, son razonables) también el consecuente las tendría. ¿Cuál es el problema? Que la lógica solo sería normativa para ciertas creencias que, por así decirlo, ya están bien. Esto le basta a MacFarlane para deshacerse de los principios $\mathfrak{B}$ (y es aplicable también a sus versiones mixtas). Creo que es sensato si queremos, como pide Harman, que la lógica sirva para guiar a un individuo aislado en su cambio razonado de creencias: Para aplicar los principios B, tendría este que partir de creencias que ya están bien, pero esta puede ser una información inaccesible al sujeto o que solo tiene en un número muy pequeño de casos. La condición de "estar bien" debería ser transparente al individuo para que pueda aplicar la lógica. Sin embargo, en otros contextos[14] los principios $\mathfrak{B}$ podrían ser adecuados.

Para MacFarlane solo quedan las versiones $\mathfrak{T}$. Enseguida elimina también las versiones $\mathfrak{T}\mathcal{P}$. El problema es que son demasiados débiles para que puedan ser consideradas como normas. No imponen ni la más mínima obligación al sujeto. No satisfacen la condición del rigor de Broome. Antes de verificar qué tan bien los 8 principios restantes enfrentan las objeciones de Harman, MacFarlane agrega dos objeciones o condiciones más.

---

[14] Por ejemplo, en el razonamiento formal.

2.8.2. Dos nuevas objeciones. Prioridad y terquedad

La primera se refiere a las versiones K. Como dije, con ellas se anula, o mitiga, la objeción de las demandas excesivas. En epistemología cuando es cuestión del principio de clausura (los conocimientos de un individuo son, o deben ser, cerrados bajo consecuencia lógica), se hace más verosímil limitándolo a las consecuencias que el individuo reconoce (lo que se llama “principio de clausura bajo implicación conocida”). Parece recomendable aquí hacer un movimiento similar y restringirnos a los principios K. Sin embargo, MacFarlane no lo hace. Por supuesto reconoce que un principio implica su variante K, pero prefiere el primero. La razón es que si solo aceptamos las versiones K estamos admitiendo que la lógica es más normativa para quien reconoce hechos de consecuencia lógica que para quien los ignora. No menos debe seguir la lógica el sabio que el lego. La normatividad de la lógica no debería hacer estos distingos:

> “Buscamos el conocimiento lógico de tal manera que sepamos cómo deberíamos revisar nuestras creencias; no solo cómo estaremos obligados a revisarlas cuando adquiramos este conocimiento lógico, sino cómo estamos obligados a revisarlas aún ahora, en nuestro estado de ignorancia". (MacFarlane, 2004, p. 12)

Llama a esta objeción “prioridad”. Sin embargo, parece difícil renunciar a las ventajas que ofrecen las versiones K. Como veremos más adelante la cuestión dependerá del tipo de normatividad que esté en juego. El mismo MacFarlane matizará su posición al final del artículo.

La segunda objeción que hace es la de la “terquedad”. Algo no está bien con la tortuga de Lewis Carroll. Su conducta al aceptar las premisas, pero no la conclusión de un argumento trivial (teniéndolas todas en mente) es reprobable. Un principio puente correcto no puede decirle simplemente que evite la inconsistencia, es decir, que no crea la negación de la conclusión. Debe ser más exigente y condenar su actitud. Un principio correcto debe decirnos que la tortuga está mal en su obstinación de negarse a admitir la conclusión obvia de un argumento correcto con premisas que acepta. Esta objeción es muy similar a la del rigor de Broome. Pueden encontrarse matices que las distinguen, pero ambas van en la misma dirección.

Los principios que han sobrevivido al análisis ($\mathfrak{T}\mathcal{O}+$, $\mathfrak{T}\mathcal{O}-$, $\mathfrak{T}\mathcal{O}+\mathsf{K}$, $\mathfrak{T}\mathcal{O}-\mathsf{K}$, $\mathfrak{T}\mathcal{R}+$, $\mathfrak{T}\mathcal{R}-$, $\mathfrak{T}\mathcal{R}+\mathsf{K}$, $\mathfrak{T}\mathcal{R}-\mathsf{K}$) guardan entre sí ciertas relaciones lógicas. Las versiones simples (sin K) implican las variantes K. De los principios con modalidad

$\mathcal{O}$ simples se siguen las correspondientes versiones $\mathcal{R}$[15] (porque es razonable hacer lo que se debe). Los principios simples positivos implican los correspondientes negativos. $\mathfrak{T}\mathcal{O}+$ es el que implica todos los demás y es justamente el único que no pasa el test de las demandas excesivas (MacFarlane parece incluir allí la objeción de las trivialidades). Dice textualmente "$\mathfrak{T}\mathcal{O}+$ implica que deberías dejar de creer los axiomas de Peano de la aritmética o venir a creer también todos los teoremas" (MacFarlane, 2004, p. 11). La cita me parece interesante porque revela cómo el autor interpreta el operador en alcance amplio, a saber, como presidiendo un condicional material. Como ya dije, la cuestión de la prioridad atenta contra las versiones K. El test de rigor de Broome debilita las versiones negativas $-$ porque solo se refieren a no creer la negación de una conclusión. La paradoja del prefacio cuestiona las versiones con modalidad $\mathcal{O}$, mientras que el problema de la terquedad se dirige contra los principios $\mathcal{R}$. Aquí hay un punto interesante: estas dos últimas objeciones van en sentidos contrarios y no hay ninguna versión que resuelva ambas. Por ende hay que elegir una u otra. MacFarlane razona que sea cual fuere la decisión que tome el autor del prefacio (admitir la inconsistencia o tratar de renunciar a alguna de sus creencias) algo estará mal. De ello sugiere un diagnóstico de la situación: una norma lógica choca aquí contra una norma de modestia epistémica. Ello equivale a dejar la paradoja del prefacio de lado. No habiendo ningún principio sobrevivido a todos los embates (lo que podría interpretarse en favor de la posición de Harman), MacFarlane adopta $\mathfrak{T}\mathcal{O}-$, aunado con $\mathfrak{T}\mathcal{R}+$ para combatir la terquedad. Sin embargo, al final del artículo cuando estos principios sirven de criterios para evaluar lógicas en competencia, MacFarlane defiende brevemente una postura interesante que más adelante comentaré. A continuación analizaré la propuesta de Field.

2.8.3. La Posición de Hartry Field

Field propone sucesivamente principios puente para enfrentar las objeciones de Harman. También responde a la objeción del constreñimiento siguiendo la observación de Broome. El operador deóntico debe tomar alcance amplio. Comienza con $\mathfrak{T}\mathcal{O}+$. Con respecto a las demandas excesivas, la solución más a la mano parece ser adoptar principios K, pero Field considera la objeción de MacFarlane contra ellos y se inclina por una versión intermedia, a saber:

(w) Si juntas α_1, α_2,... α_n implican obviamente a γ, entonces uno no debería creer α_1, α_2,... α_n sin creer γ.

[15]Esto no vale para las versiones $\mathfrak{B}$, pero a estas alturas solo quedan principios $\mathfrak{T}$.

Advierta que (w) no es ninguno de los 36 principios de MacFarlane. Una idea similar a esta de "consecuencias obvias" había sido explorada por Harman:

> "Estoy inclinado a tomar como fundamentales ciertas disposiciones a tratar proposiciones de una cierta manera, en particular, disposiciones a tratar ciertas proposiciones como inmediatamente implicando otras, y algunas como inmediatamente inconsistentes con otras" (Harman, 1986, p. 18).

Sin embargo, veremos un poco más adelante cómo Field explota una ambigüedad de la palabra "obviamente" en su figuración en (w). En contraste con Harman, para atacar el problema de la inconsistencia (y, en particular, la paradoja del prefacio) Field introduce grados de creencia en sus siguientes propuestas:

(w+) Si α_1, α_2,... α_n implican obviamente a γ, entonces el grado de creencia que uno tiene en γ debería ser al menos tan alto como el grado de creencia que uno tiene en $\alpha_1 \wedge \alpha_2, ... \wedge \alpha_n$[16].

El siguiente principio es una refinamiento del anterior:

(d) Si α_1, α_2,... α_n juntas implican obviamente a γ, entonces los grados de creencia de uno en α_1, α_2,... α_n y γ deberían estar relacionados según la fórmula $P(\gamma) \geq P(\alpha_1) + P(\alpha_2) + \ldots P(\alpha_n) - (n-1)$, (donde $P(\alpha)$ denota el grado de creencia en α).

(d) es preferible a (w+) porque permite hacer mayores precisiones en cuanto a los grados de cada una de las creencias que constituyen el antecedente del principio puente. Dos cosas son de notarse. La primera es que (d) consituye una generalización de (w), pues si el grado de creencia en cada α_i es 1, (d) prescribe que el grado de creencia de γ sea 1. La segunda observación es que tanto (d) como (w) son neutros con respecto a una gran variedad de relaciones de consecuencia lógica. Es un principio que parece correcto para una noción intuitiva de implicación. Sin embargo, es dudoso cómo se aplicaría si entre las premisas hubiera alguna contradicción.

Dejando estos problemas de lado, subrayo uno que ahora es más urgente, a saber, ¿debemos introducir grados de creencia? En general, yo no exploraré esta vía. Para ello tengo varias razones. La primera es la de tratar de simplificar al máximo el problema de que me ocupo y limitarme a los más importantes sistemas de lógica epistémica que no utilizan la teoría de la probabilidad. Me centro ahora en principios como (d) a sabiendas de que tal vez puedan generalizarse. La segunda cuestión es que, a pesar de que se recurre a ellos especialmente para tratar

[16]Por ejemplo, el grado de creencia de que Linda es es cajera y activista debe ser menos que el grado de creencia de que Linda es cajera.

las paradojas del prefacio y de la lotería, hay un sentido en que no resulta fácil aplicarlos aún a estos casos. En general no solemos atribuir grados de probabilidad a todas las proposiciones que creemos. Pero supongamos, por ejemplo, que las personas están dispuestas a otorgar un valor entre 0 y 1 a ciertas proposiciones según el grado de convicción con que las mantienen, donde 0 representa la creencia en la negación. Podríamos determinar un umbral más arriba del cual el sujeto está dispuesto a decir que cree algo. Entonces una propuesta como la de Field no resolvería la paradoja de la lotería. Supongamos que mostramos al agente un boleto de lotería en un concurso en el que hay 100 mil números. El individuo en cuestión otorga 0.99999 de probabilidad a que el boleto no ganará. Tendrá todo el derecho a decir que cree que ese boleto no será el vencedor. Sin embargo, su convicción de que ningún boleto ganará es de 0. La lógica lo llevará de proposiciones que cree a una conclusión en la que no cree. Por supuesto que el defensor de esa propuesta dirá que la conjunción de dos proposiciones con grado 0.9999 tendrá una probabilidad menor que ese decimal. Conforme los conyuntos vayan creciendo disminuirá la probabilidad otorgada a la conclusión ("ninguno ganará") va bajando. Pero si nos atenemos a la aserción, el sujeto afirmará todas las premisas individuales y negará la conclusión. Este argumento aparecía ya en Kyburg (1961) pero tenía otra conclusión. El punto es que si consideramos la creencia como ligada a la aserción (si alguien afirma P, podemos, casi siempre, atribuirle la creencia en P), entonces en la medida en que la aserción no suele estar ligada a ninguna gradación, la creencia tampoco lo estará. Los defensores de utilizar aquí la teoría de la probabilidad argumentarán quizás que la aserción o la dispocisión a afirmar una proposición no es más que una manifestación de la creencia, que esta es una actitud interna con diversas gradaciones como lo muestra la mayor o menor confianza con que un individuo defiende sus creencias o está dispuesto a renunciar a ellas. Pero no olvidemos que estamos hablando de creencias en el sentido ordinario de este vocablo y que en el uso que de él hacemos a diario no suele estar ligado a una escala graduada muy fina y precisa.

Reconozco que este argumento es discutible y que ha sido frecuente enfrentar ese tipo de paradojas con teorías probabilísticas de la creencia. En todo caso, recuerdo la razón que di antes para seguir con principios puente sin términos de probabilidad: me limito a algunos de sus casos particulares (donde el grado de creencia es 0 o 1) para simplificar el problema. Vuelvo a la propuesta de Field. ¿Cómo resolver el problema de las trivialidades? Field retoma una idea de Harman, a saber, distinguir entre creencias implícitas y explícitas.

> "Pero uno puede creer y cree una infinidad de cosas. Por ejemplo,

> uno cree que la tierra no tiene dos soles, que la tierra no tiene cuatro soles...Para acomodar este punto, asumo que podemos distinguir lo que uno cree explícitamente de lo que uno cree solo implícitamente. Entonces podemos aplicar el principio de evitar trivialidades solo a lo que uno cree explícitamente" (Harman, 1986, p. 12)

Como veremos esta idea ha aparecido con frecuencia en el tratamiento formal del problema de la omnisciencia lógica desde Levesque (1986). Sin embargo, no todos sus proponentes trazan la distinción en el mismo punto. Es más, la mayoría se contenta con una entendimiento intuitivo de lo que estos dos adjetivos significan. Levesque caracteriza lo que entiende por "creencias explícita" muy breve y vagamente ("un enunciado es explícitamente creído cuando es activamente sostenido por el agente" (1986, p. 198)), mientras que define las crencias implícitas como las consecuencias lógicas de las primeras. Es decir, que el problema de la omnisciencia lógica solo se plantea para las creencias explícitas. Harman no es mucho más elocuente a este respecto, pero aclara que la distinción implícito/explícito no coincide con la dicotomías inconsciente/consciente ni con disposicional/ocurrente (Harman, 18986, pp. 13-14). Para él las creencias implícitas tampoco son cerradas bajo consecuencia lógica, pero el problema que sus objeciones plantean es aplicable primariamente a las creencias explícitas (porque estas están sujetas a deliberación racional y claramente ocupan "espacio" en la mente).

Para Field las creencias explícitas de S son las que S tiene directamente almacenadas, mientras que S cree implícitamente algo cuando está dispuesto a creerlo explícitamente, si la cuestión resultara. Con ello enuncia un nuevo principio puente:

(d*) Si es obvio que α_1, α_2,... α_n juntas implican obviamente γ, entonces uno debería imponer la restricción de que $P(\gamma)$ debe ser al menos $P(\alpha_1) + P(\alpha_2) + \ldots P(\alpha_n) - (n-1)$ en cualquier circunstancia en que sea cuestión de α_1, α_2,... α_n y γ.

Note que la última cláusula toma en cuenta el caracter explícito de la creencia. Al ser cuestión de ellas, las premisas y la conclusión son expuestas a la conciencia del agente.

Field confiesa que su último refinamiento es aún muy tosco. Su objetivo era solo mostrar que se puede enfrentar el reto de Harman. Por ejemplo, queda el problema de qué debemos entender por "implicación obvia". Una posibilidad sería suponer que α_1, α_2,... α_n obviamente implican γ si γ puede ser obtenido de α_1, α_2,... α_n a través de ciertas reglas de inferencia. Aclara Field que esa implicación obvia no sería transitiva. De lo contrario se expondría a la objeción de que una inferencia compleja puede descomponerse en deducciones obvias, en previsión de lo cual observa asimismo que la complejidad inferencial no (siempre) tiene que ver con la longitud del argumento. Sin embargo, no está claro con qué criterios podríamos elegir estas reglas de inferencia. Veremos cómo algunos sistemas de lógica epistémica intentan una salida similar para el problema de la omnisciencia lógica.

2.8.4. Evaluaciones internas y externas. Field y MacFarlane

Volvamos al problema anterior. (d*) comienza con "si es obvio que α_1, α_2,... α_n juntas implican γ", pero ¿qué significa "obvio"? Field observa una ambigüedad en esta palabra. La pregunta es ¿obvio para quién? ¿Para nosotros o para el agente? Supongamos que el agente es creyente de la corrección de una lógica L' distinta de la lógica "correcta" L (la de quien lo evalúa). Supongamos que hace un razonamiento conforme a L no autorizado por L'. ¿Cómo debemos evaluar su desempeño? En algún sentido cometió un error porque no actúo según las normas de su propia lógica. En otro, no, porque finalmente hizo una inferencia correcta de acuerdo a nuestra lógica. Field no cree que haya un *fact of the matter* sobre qué lógica es la correcta y, por lo tanto, admite ambos tipos de evaluaciones (es decir de normatividades).

Es interesante observar que MacFarlane, al considerar la formalidad de la lógica, hace una consideración que puede llevar a una conclusión similar a la de Field. Observa que mientras de las premisas:

(a) Cicerón come

(b) Cicerón está en compañía de unos amigos

Puede lógicamente inferirse:

(c) Alguien come y está en compañía de unos amigos.

No ocurre lo mismo si la primera premisa es reemplazada por:

(a') Tulio come.

¿Qué hace la diferencia? Se dirá que el agente que sostiene (a) y (b) podría no saber que Tulio y Cicerón son la misma persona. O dicho a la inversa, el argumento original es lógicamente correcto si suponemos que las dos figuraciones de "Cicerón" se refieren a la misma persona. Pero ¿cómo lo podemos determinar? Si adoptamos una teoría internista de la referencia el asunto dependería de lo que ocurre en la intimidad del sujeto. Es decir, si suponemos que lo que determina la referencia de un término en una ocasión en que lo usa un agente S depende, en última instancia, de lo que sucede en la mente de S, entonces solo el individuo sabrá, si acaso, a qué hacía referencia. Si, por el contrario, adoptamos una teoría externista de la referencia, es decir, si suponemos que la referencia de un término en una ocasión de uso puede depender de factores físicos o sociales más allá del control del usuario, entonces puede ser que ni siquiera el agente sepa si su argumento es correcto. La solución de MacFarlane es la siguiente. La lógica produce esquemas de inferencia correctos (o de consecuencia lógica). Un esquema generará una norma N para el sujeto en una circunstancia concreta C, solo si él percibe que en C se da un ejemplo particular de N. Lo interesante de esta propuesta es que apunta, como la de Field, a la posibilidad de una evaluación desde el punto de vista del sujeto y, por ello, deja abierta la posibilidad de una evaluación externa. Sin embargo, MacFarlane no advierte que cae bajo la objeción de la prioridad: la lógica sería más normativa para el sujeto más diestro en reconocer patrones de inferencia.

2.8.5. La propuesta de Florian Steinberger

En una de sus exposiciones del tema, Florian Steinberger (2016) ha introducido una distinción de los operadores de obligación, que me será muy útil y puede extenderse a los otros dos tipos de operadores deónticos:

1) Objetivos: El operador deóntico objetivo evalúa las acciones de un individuo desde una perspectiva completamente externa. Por ejemplo, un individuo se encuentra extraviado en un bosque. Un observador que rastrea sus movimientos por medio de un satélite puede calificar una acción suya como errada porque no conduce a la salida.

2) Subjetivos: En una evaluación subjetiva del ejemplo anterior el observador podría considerar que el sujeto obró correctamente, aunque no se dirija a la salida, porque hizo lo que era mejor dadas sus creencias o conocimientos.

Recordemos el dilema que plantea Field. Si el observador tiene una lógica L y juzga a un agente que infirió de acuerdo a L, pero que profesa los principios de una lógica L' que, en este caso, prescribía otra conclusión, ¿cómo debe ser calificado? ¿Hizo lo correcto? La respuesta es doble: si el observador está evaluando objetivamente, la respuesta es positiva. Si se trata de una evaluación subjetiva, la respuesta es negativa.

Steinberger acusa a Field y a MacFarlane de no contestar al reto de Harman en los términos en que fue originalmente planteado. Harman dudaba que la lógica pudiese dar normas que sirviesen de guías para el razonamiento del individuo en su soledad. Steinberger se propone ofrecer principios puente que muestren que Harman estaba equivocado tomando en cuenta este requerimiento. Primeramente para justificar su posición apela la norma de la verdad. Aunque admite que la lógica deductiva y la teoría del razonamiento son distintas, trata de mostrar que hay un vínculo estrecho entre ambas. Un primer paso para establecer esa conexión es el siguiente: el conocimiento teórico nos provee una imagen correcta del mundo y nosotros lo representaremos correctamente teniendo creencias verdaderas y evitando las falsas. Tener conciencia de las relaciones lógicas entre los diferentes contenidos de nuestros estados doxásticos puede conducirnos finalmente a tener creencias verdaderas y, por lo tanto, será relevante al conocimiento teórico. En especial las nociones de consistencia y de consecuencia son relevantes.

Al igual que Field, ofrece un refinamiento sucesivo de principios puente para ir superando los retos de Harman. Su posición puede resumirse de la siguiente manera:

> ($\mathfrak{C}\mathcal{O}$+B) Si S cree que $\alpha_1, \ldots \alpha_n \models \gamma$ entonces si S cree todos los α_i's, entonces debería creer γ. (No responde a la objeción del constreñimiento).
>
> ($\mathfrak{T}\mathcal{O}$+B) Si S cree que $\alpha_1, \ldots \alpha_n \models \gamma$ entonces si S debería creer todos los A_i's entonces cree γ. (No resuelve la paradoja del prefacio).
>
> ($\mathfrak{T}\mathcal{R}$+B) Si S cree que $\alpha_1, \ldots \alpha_n \models \gamma$ entonces si S tiene razones para creer todos los α_i's entonces cree γ. (No resuelve el problema de las trivialidades).
>
> ($\mathfrak{T}\mathcal{R}$+B$'$) Si S cree que $\alpha_1, \ldots \alpha_n \models \gamma$ y considera o tiene razones para considerar γ, entonces si S tiene razones para creer todos los α_i's entonces cree γ.

Es muy interesante observar que estos no son realmente principios puente, si no partimos de que la creencia del sujeto consignada en el antecedente es verdadera, es decir, si no partimos de ningún hecho de consecuencia lógica sino únicamente de la creencia del agente en un hecho tal. Por ello, la B en la denominación de los principios anteriores. Es decir, los principios son internos al sujeto. Podemos hacer con ellos una apreciación subjetiva de sus inferencias, pero para él son guías. Con ello se anula la objeción de las demandas excesivas. Steinberger supera el reto del constreñimiento adoptando el alcance amplio del operador deóntico; enfrenta la paradoja de la lotería pasando de "obligación" a "tener razones"; y el problema de las trivialidades con la cláusula "y considera o tiene razones para considerar γ". Sin embargo, no soluciona el problema del rigor. ($\mathfrak{T}\mathcal{R}$+B') es un principio bastante laxo, pero Steinberger considera que es un problema menos grave que los otros.

Antes de hacer una comparación entre las propuestas anteriores de principios puente, mencionaré brevemente que una disputa similar se dio en epistemología en torno al principio de clausura, es decir, al principio según el cual los conocimientos de un individuo son cerrados bajo consecuencia lógica. El asunto, por supuesto, es mucho más complejo, entre otras cosas porque involucra la discusión que aquí revisamos con respecto a la creencia. Si el conjunto de creencias de un individuo no es lógicamente cerrado, el de sus conocimientos difícilmente lo será. Sin embargo, hay otros factores que complican aún más el debate en epistemología. Una manera de hacer verosímil el principio de clausura es restringirlo a su versión K (lo que suele llamarse "clausura bajo implicación conocida"). Otro expediente suplementario es adoptar una versión sincrónica del principio para asegurarse de que el conocimiento de las premisas y de la inferencia se de casi simultáneamente. Sin embargo, eso no basta cuando se trata del conocimiento porque puede ocurrir que el individuo crea las premisas, haga la inferencia, crea la conclusión, pero no por haberla derivado de las premisas, sino por alguna otra razón, en cuyo caso tal vez no tenga la justificación adecuada para calificar como teniendo conocimiento[17]. Hay además quienes consideran que la evidencia que tiene el agente para creer podría no trasmitirse a través de una inferencia correcta a la conclusión[18]. Como se ve, es mejor tratar con anterioridad el problema restringiéndolo a la creencia. Menciono esto para mostrar que la discusión anterior tiene ciertas implicaciones en el terreno epistemológico.

Expuse las propuestas de MacFarlane, Field y Steinberger que, además de ser de las más promientes en la literatura, sobre el tema, no involucran principios de probabilidad. Como dije antes, yo no abarcaré soluciones probabilísticas. Paso

[17]Para una presentación de estos argumentos, véase (Marian, D. Warfield, D. 2008).
[18]Para un panorama de la discusión en este tema, véase Collins (2008).

ahora a examinar los principios puente en las variantes de estos autores.

2.8.6. Análisis de los Principios Puente

Me parece que si una lógica epistémica quiere normar el cambio de creencias del agente entonces debe contar con algunos artificios que eludan o mitiguen las objeciones de Harman. Enseguida repaso transversalmente algunas de las propuestas anteriores. Para evitar la objeción del constreñimiento, hay consenso en adoptar principios $\mathfrak{T}$. La idea de considerar el alcance de operadores deónticos en proposiciones normativas fue propuesta por Broome (1999). La expresión "deberías ver que"[19] es un mero artificio para poder completarla con una oración. Por ejemplo, en lugar de decir "deberías bañarte", lo formularemos así: "deberías ver que tu te bañes". Broome distingue normas condicionales de requerimientos normativos. Las primeras son formuladas de manera condicional, pero admiten *modus ponens* ("si te piden el IFE, deberías reclamar", "si la prueba te sale positiva, tienes una razón para pensar que estás enfermo"). Es decir, se trata de principios $\mathfrak{C}$. En cambio, los "requerimientos normativos" se expresan con un operador deóntico que domina un condicional fuerte (no material). En particular, el enunciado "α requiere normativamente de γ" no es veritativo funcional ni admite *modus ponens*. Implica pero no es equivalente a un enunciado de la forma $\mathcal{O}(\alpha \to \gamma)$, donde $\mathcal{O}$ es un operador de obligación. ¿Por qué no es equivalente? Porque puedo obedecer $\mathcal{O}(\alpha \to \gamma)$ esforzándome para que no suceda α y eso no es lo que propone el requerimiento normativo. Lo esencial dice Broome es que "α requiere que γ" implique $\mathcal{O}(\alpha \to \gamma)$, pero no tenga como consecuencia "si α entonces deberías γ", ni "si α entonces tienes razones para γ". Uno de los ejemplos que ofrece es justamente el de la inferencia lógica:

> Supongamos que de Q se deriva de P por una inferencia válida... Ahora supón que tu crees P. Entonces un proceso de razonamiento te llevará a creer que Q... Sin embargo, no es necesariamente el caso que tu deberías creer Q, ni que tengas una razón para creer que Q... La relación entre creer que P y creer que Q es estricta. Si tu crees P y no crees Q, definitivamente no estás como deberías.... BP requiere normativamente BQ. (Broome, 1999, p.6)

Aquí vemos en qué difiere dicho requerimiento normativo, del simple $\mathcal{O}(\alpha \to \gamma)$. No es lo mismo decir que la creencia en las premisas de un argumento correcto

[19] Tal vez sería mejor la expresión "deberías procurar que" para recuperar el sentido de la expresión inglesa.

requiere normativamente la creencia en la conclusión que decir "deberías ver que si crees las premisas entonces creas la conclusión" porque esta última podría satisfacerse no creyendo alguna de las premisas. Es decir, el requerimiento normativo es más fuerte. ¿Qué le falta a $\mathcal{O}(\alpha \to \gamma)$ para expresar que α requiere normativamente a γ? Broome solo dice que una cierta "determinación" y que le basta con lo dicho, dejando el resto a la intuición. Sin embargo, si partimos de una lógica que admita *modus ponens* y condicionalización (para el condicional material), $\alpha \models \gamma$ será equivalente a $\models \alpha \to \gamma$ lo que dará pie a un principio puente $\mathfrak{T}\mathcal{O}+$ del tipo $\mathcal{O}(\alpha \to \gamma)$, donde el condicional inserto es material. Por otro lado, nuestros autores usan el alcance amplio del operador modal para permitir que un agente, ante una inferencia correcta cuyas premisas creía, tenga la opción de aceptar la conclusión o de renunciar a alguna de las premisas. Es decir, básicamente requieren que un prinicipio $\mathfrak{T}$ positivo no implique el correspondiente principio $\mathfrak{C}$, pero no exploran principios puente con un condicional subordinado más fuerte. Sin embargo, concuerdan en que los principios puente correctos son de tipo $\mathfrak{T}$. Así superan la objeción del constreñimiento. Esto no impide que algunos sistemas la interpretación en términos de principios puente de tipo $\mathfrak{B}$ o $\mathfrak{C}$ sea más adecuada, como diré enseguida. Como vimos, la única objeción a los principios $\mathfrak{B}$ es que limita el alcance de la normatividad de la lógica, pero es posible que fuesen útiles en contextos más restringidos.

En cuanto a la objeción de las demandas excesivas, varios expedientes fueron sugeridos. El primero es limitar los principios puente a las consecuencias o implicaciones obvias de las creencias del agente. Esto requiere de un procedimiento para medir el esfuerzo cognitivo necesario para obtener una conclusión a partir de un conjunto de premisas. Como recuerda Field, tal medida no coincide necesariamente con la longitud de una derivación formal. Quizás no sea un problema que ataña solo a la lógica, pero un sistema podría integrar algún artificio que tome en cuenta esa medida. En segundo lugar un argumento correcto puede ser normativo a un agente que cree las premisas solo en la medida en que se trate de una inferencia que él reconoce como correcta. Esto puede obtenerse tomando los pricipios puente K (o B). O, bien, como hizo Steinberger combinando ambas propuestas: el agente solo contrae una obligación o una autorización para creer β en la medida en que esta es una conclusión obvia de sus creencias, por medio de inferencias que reconoce como correctas. Pero en los sistemas que veremos aparece otra mejora ya mencionada. Podemos refinar aún más las obligaciones del agente limitándonos a la lógica que él reconoce o que se le concede, la cual no tiene por qué ser la nuestra.

Había yo antes hablado de una lógica externa, que es la lógica que permite hacer transiciones en el sistema y una lógica del observador que permite hacer transiciones entre sus creencias. Por ejemplo, supongamos que la lógica del sistema es clásica, mientras que la del agente es relevantista. Entonces en el sistema podemos inferir de $((\mathcal{B}\alpha \vee \mathcal{B}\beta) \wedge \neg\mathcal{B}\beta)$ que $\mathcal{B}\alpha$, pero no podemos deducir $\mathcal{B}\alpha$ de $\mathcal{B}(\alpha \vee \beta)$ y $\mathcal{B}\neg\beta$.

Es importante que esté abierta la posibilidad de otorgarle su propia lógica al agente (aunque no sea la que él reconoce, sino tal vez aquella con la que que será evaluado). Sin embargo, esta posibilidad también debe estar restringida. No otorgar ninguna lógica al agente contravendría la normatividad de la lógica y también sería contrario a las reglas por las cuales atribuimos creencias a los otros. Por otro lado, un sistema que nos permita concederle cualquier lógica nos deja en la obscuridad sobre el fenómeno de la creencia.

Tanto Field como Steinberger mitigan la cuestión de las trivialidades apelando a lo que podríamos llamar la conciencia; la lógica solo norma en situaciones en que al individuo se le hace consciente de las premisas y la conclusión del argumento. MacFarlane finalmente parece coincidir con ellos en este punto.

¿Qué modalidad deóntica es la adecuada? Mientras Field elige la modalidad de obligación, Steinberger opta por la de tener razones. Field pretende resolver la paradoja del prefacio apelando a las probabilidades y deja en pie el rigor de la lógica. Steinberger recurre a la modalidad $\mathcal{R}$ para resolver la paradoja del prefacio (el autor del libro solo tiene razones para revisar sus creencias, pero no está obligado a ello en la medida en que también podría tener una justificación más fuerte para no emprender esa revisión). En esto no hay consenso ni un tipo de principio que obviamente se imponga.

Los diversos principios puente que hemos revisado, u otros posibles, nos abren varias posibilidades para interpretar sistemas de lógica epistémica como normativos.

2.9. Recapitulación

En este capítulo mostré que, aunque el problema de la omnisciencia lógica es claro en su aspecto negativo, a saber, el ser humano ordinario no es omnisciente como lo representan los sistemas clásicos de lógica epistémica, también hay un aspecto positivo que hay que iluminar: ¿qué deseamos que cumpla un sistema de lógica epistémica? ¿Cuál es el objetivo para construir un sistema sin esa característica? Ello da pie a una pregunta previa obligatoria, a saber, ¿los sistemas de Hintikka no satisfacen ya sus objetivos? Para responder a esta pregunta en este capítulo desarrollé una serie de elementos filosóficos con los cuales pretendo analizar los sistemas de lógica epistémica que se han propuesto resolver el problema de la omnisciencia lógica.

En un primer acercamiento al problema noté que muy frecuentemente se ha considerado a la omnisciencia lógica como siendo producto de una idealización de las capacidades cognitivas del ser humano. Pero, ¿no es la idealización consubstancial a toda empresa científica y, por lo tanto, está justificada? Para responder a esta cuestión resumí las posiciones de Yap y Colyvan. Parecen ambos coincidir en que hay idealizaciones descriptivas y las hay normativas. No obstante mi postura es que ninguna de estas dos justificaría la omnisciencia lógica. Supongamos que usamos la idealización para describir al agente humano. De acuerdo a la lógica de Hintikka este agente estaría a una distancia infinita del individuo de carne y hueso. En ese sentido no representan ni lo esencial del fenómeno de la creencia, ni un objetivo al cual haya esperanza de acercarse algún día. Eso ha sido reconocido en la literatura sobre el tema. También concluí que la idealización no puede valer como norma, pues entonces exige de un ser humano algo absolutamente imposible. El ser omnisciente no requiere de razonamiento ni de lógica alguna y si alguna vez intentáramos imitar sus hazañas cognitivas, la derivación de conclusiones inútiles y triviales nos consumiría. Sin embargo, acepto que algún grado de idealización es conveniente y finalmente aparecerá en cualquier sistema de lógica epistémica.

El segundo planteamiento filosófico que exploré tiene que ver con un elemento central en el problema de la omnisciencia lógica, a saber: la creencia. ¿Qué tan bien representan los sistemas lógicos la proposición de que el agente cree que P o cree que no P? Para contestar a esta pregunta revisé algunas concepciones de la creencia que son abanderadas por ilustres filósofos del siglo pasado (Fodor, Drestke, Dennett y Cherniak). Destilé dos posiciones centrales en este tema. La primera, defendida por Fodor y Dretske, concibe la creencia como una representación interna a la mente del agente. En el caso de Fodor la creencia no se limita a ser una instancia que aparece en la mente del agente y que puede corresponder a un hecho en el mundo. Fodor argumenta a favor de una teoría de las actitudes proposicionales que le permite explicar de qué manera las creencias como representaciones internas influyen en la conducta del agente. Dretske también señala un aspecto práctico de la creencia: debe ser guía para la acción y el pensamiento. La segunda concepción de la creencia, ejemplificada por Dennett y Cherniak, define a la creencia como un concepto que utilizamos para explicar y predecir la conducta de los demás. En el caso de Dennett la creencia puede ser correctamente atribuida al otro en la medida en que podemos predecir su conducta a un grado razonablemente bueno usando la estancia intencional. Para Cherniak, la posibilidad de predecir y explicar con ciertos principios racionales el comportamiento de un individuo es lo que nos autoriza a concebirlo como creyente. Sugerí que estas dos concepciones de la creencias se corresponden con sendos roles que la lógica puede desempeñar en relación con la creencia.

En una perspectiva representacionista la lógica contribuye a la preservación de la verdad a través de la inferencia y con ello propongo que al modelar este tipo de creencia el rol que adquiere la lógica es más bien normativo. Alguien que sigue una lógica inadecuada tendrá ideas falsas sobre su entorno y la creencia no cumplirá cabalmente su función en tanto que representación. Por el lado del atribucionismo la lógica es constitutiva de la creencia (aunque también pueda tener un rol normativo). Por ejemplo, si atribuimos una creencia en la santidad de la vida a un sujeto también le adscribimos la creencia de que la eutanasia no es totalmente aceptable. Si el individuo defiende a la vez la santidad de la vida y el aborto y la eutanasia, empezaríamos a dudar de que realmente lo que dice corresponde con sus creencias. Ahora bien, tanto la concepción ordinaria de la creencia como algunas más filosóficas suelen combinar estas dos perspectivas.

En la última parte de este capítulo abordé una discusión bastante reciente que se cuestiona sobre si la lógica es un una disciplina normativa del razonamiento humano. Para ello revisé los argumentos para responder negativamente a esta cuestión y que van de Gilbert Harman y Alvin Goldman a Gillian Russell. Su punto es que la lógica estudia ciertas relaciones entre conjuntos de enunciados (consecuencia lógica, consistencia, validez lógica) y no tiene nada que ver con el razonamiento. Más precisamente la lógica guía un proceso inferencial que se puede llevar a cabo en las ciencias, pero que no necesariamente está relacionado con creencias. En cambio, no guía, ni pretende hacerlo, el cambio razonado de creencias que llevamos a diario cuando recibimos nueva información y modificamos a partir de ella nuestras convicciones. Uno de los resultados que ha arrojado esta investigación sobre la normatividad de la lógica alude a la concepción de principios normativos llamados principios puente, que se derivan de manera uniforme de enunciados de validez lógica. En este contexto MacFarlane, Field y Steinberger nos presentan algunas propuestas de principios puente. Estos principios abren un canal para, por un lado aceptar la postura de Harman de que las verdades de la lógica no son normativas pero, por otro, admitir que sí pueden dar lugar a principios normativos. Esta es la posición que yo defiendo.

Capítulo 3

Sistemas sin omnisciencia lógica I.

3.1. Criterios de evaluación de sistemas de lógica doxástica

En el capítulo anterior desarrollé los criterios filosóficos para la clarificación de la omnisciencia lógica y para la evaluación de los sistemas que se han propuesto resolverla. En el capítulo 3 y 4 mostraré la pertinencia de esta discusión. ¿Qué criterios derivo de la discusión anterior?

Antes que nada, debo aclarar que no pretendo que mi propuesta sea la única que conteste a la cuestión de cuáles deberían ser los objetivos de los mencionados sistemas. Como veremos algunos autores dan algunos criterios, aunque un poco vagos, que también permitirían evaluar sus sistemas en la medida en que lograron cumplir lo que intentaron. Más bien lo que pretendo es mostrar que las concepciones filosóficas de la creencia y del papel de la lógica permiten dar una mayor precisión al problema de la omnisciencia lógica (en su versión doxástica) y con ello proveer criterios para determinar los sistemas que veremos resuelven el problema.

En lo que se refiere a la creencia como atribución, me parecen muy pertinentes las observaciones de Cherniak, muy en particular la idea de que las reglas que seguimos en la adscripción de creencias a los otros se aplican de manera holística y contextual y tal vez sea imposible abstraer reglas lógicas generales que nos rijan en esta práctica. Sin embargo, hay al menos una idea que me parecen indubitable y es que las creencias vienen en cúmulos organizados de acuerdo a ciertas relaciones lógicas.

Más relevante para mi investigación será el papel normativo de la lógica. La estrategia será tomar algunas fómulas doxásticas representativas de la omnisciencia lógica y ver qué es lo que su validez o invalidez en cada sistema nos dice sobre este, cuando interpretamos esas fórmulas como principios puente. Esto me llevará también a considerar cuál es la concepción de la creencia que subyace a la construcción de cada sistema

Se me podrá objetar inmediatamente que no es el objeto de esos sistemas responder a las objeciones de Harman, sino resolver el problema de la omnisciencia lógica. A ello respondo primeramente que las objeciones de Harman y MacFarlane se traducen efectivamente en criterios que un sistema de lógica debe satisfacer, si pretende normar el cambio de creencias de agentes reales. Por otro lado, es notable cómo muchos proponentes de estos sistemas leen sus fórmulas como si tuvieran operadores deónticos implícitos.

Ahora bien, debo aclarar que mi enfoque será liberal, es decir, que no pretendo que haya una única manera correcta de interpretar normativamente dichas proposiciones, sino que tomaré diversos elementos de la discusión anterior. Por ejemplo, he dicho que hay al menos dos estrategias para enfrentar las demandas excesivas. Algunos sistemas disponen de principios para medir o considerar la dificultad cognitiva que conlleva una inferencia. En ese caso, solo las que sean de muy bajo costo cognitivo impondrán una obligación o un permiso al agente. Otros, en cambio, limitan las obligaciones del sujeto a las creencias que reconoce (es decir, adoptar algo similar a las versiones K de MacFarlane, solo que para la creencia). Veremos que esas estrategias han aparecido efectivamente en la literatura lógica, y a veces un sistema se adaptará más a una que a otra.

En cuanto al constreñimiento, yo adoptaré principios $\mathfrak{C}$, que me parece ofrecen una lectura más natural. Es verdad que los tres autores antes considerados adoptaron principios puente de la modalidad $\mathfrak{T}$. Como vimos, en la concepción de Broome los requerimientos normativos no pueden traducirse por un condicional material gobernado por un operador deóntico ($\mathcal{O}(\mathcal{B}\alpha \rightarrow \mathcal{B}\gamma)$. Sin embargo, no está claro cómo podrían ser formalizados. Yo adopto una solución semejante a la de los autores mencionados. Leo las fórmulas (con cierta forma) como principios puente $\mathfrak{C}$, pero con la posibilidad abierta al sujeto de rechazar el consecuente o revisar los antecedentes. Así $(\mathcal{B}\alpha \rightarrow \mathcal{B}\gamma)$ podrá ser visto como una prescripción de lo que el agente debe creer si ya cree α. Sin embargo, retengo la lectura $\mathfrak{T}$, en el sentido en que si la conclusión es inadmisible o poco aceptable, la regla prescribiría renunciar a su creencia en α. Al igual que Field, favorezco la lectura con la modalidad de obligación para responder a los requerimientos del rigor y de la terquedad. Sin embargo, un sistema podría interpretarse mejor a partir de una lectura de principios $\mathfrak{B}$ o con una modalidad $\mathcal{R}$. Este es el caso de la lógica clásica en su guía de la inferencia correcta en una ciencia deductiva.

Ahora bien, podemos conceder al agente una lógica diferente a la nuestra, pero esa lógica forzosamente limita la manera en que debe cambiar sus creencias. Si un sistema permite que el agente cambie sus creencias arbitrariamente no supera la objeción del rigor. Como vimos, ni siquiera podemos sensatamente atribuir racionalidad a un individuo cuyas "creencias" se modificaran sin ninguna regla. Normalmente un sistema que tolera esta libertad supone que deberá agregarse otro artificio para enfrentar esta dificultad. En esa medida el sistema es insuficiente para resolver el problema tal y como yo lo planteo.

Veremos que varios sistemas proponen, al atribuir una lógica más débil (que la nuestra) al sujeto, resolver así el problema que nos ocupa, pero está claro que si el agente es omnisciente en esa lógica entonces no se supera la objeción de las demandas excesivas. Hay que agregar algo más.

La objeción de la inconsistencia suscita varias dificultades que analizaré en detalle en las siguientes secciones y, posteriormente en el capítulo 4, donde estas consideraciones quedarán más claras a la luz de su aplicación a sistemas concretos.

3.2. Manifestaciones de la omnisciencia lógica

Recordemos que el problema de la omnisciencia lógica consiste en que en todos los sistemas normales (como los sistemas $\mathbf{KD}$ y $\mathbf{KT}$ empleados para representar respectivamente la creencia y el conocimiento) la creencias del agente representado son cerrradas bajo consecuencia lógica. En particular, en ello es verdadera la siguiente afirmación:

$$\text{si } \alpha \models \beta \text{ entonces } \mathcal{B}(\alpha) \models \mathcal{B}(\beta))$$

donde $\mathcal{B}$ es el operador de creencia y donde $\models$ se refiere aquí a la consecuencia lógica en el sistema respectivo, la cual incluye las consecuencia tautológica. En general, para culquier n:

$$\text{si } \alpha_1, \alpha_2 \ldots \alpha_n \models_{\mathbf{K}} \beta \text{ entonces } \mathcal{B}\alpha_1, \mathcal{B}\alpha_2, \ldots \mathcal{B}\alpha_n \models_{\mathbf{K}} \mathcal{B}\beta$$

Es decir que si en el sistema podemos realizar una inferencia, también el agente puede llevarla a cabo. Por eso digo que la "lógica del sistema" es la misma que "la lógica del agente". Aquí me expreso de manera dinámica al referirme a las inferencias que el agente puede realizar. Estrictamente hablando la situación es que las creencias del individuo modelado son cerradas bajo la propia lógica en cuestión. Como hice ver en el primer capítulo, con referencia a la lógica K (o a cualquiera de sus extensiones) si consideramos a la lógica como descriptiva, este resultado es inaceptable. Aún suponiendo que la lógica es normativa y que en fórmulas del tipo $(\mathcal{B}\alpha \to \mathcal{B}\beta)$ se halla implícito un operador deóntico inicial (de alcance amplio), $\mathbf{K}$ sigue adoleciendo de varios defectos. Por ejemplo, si tomamos el operador $\mathcal{O}$, el sistema $\mathbf{K}$ no toma en cuenta las limitaciones del agente que podría no ser consciente del hecho de consecuencia lógica, ni la cuestión de las trivialidades, etc. Si pasamos al operador $\mathcal{R}$ entonces la lógica pierde casi toda fuerza preceptiva.

En lo que sigue revisaré algunos de los muchos sistemas formales que han sido propuestos para tratar el problema de la omnisciencia lógica en los últimos cuarenta años. En este capítulo trataré de sistemas no dinámicos o, podría decir, atemporales. En el próximo me enfocaré en sistemas que describen o norman el cambio de creencias de un individuo después de que realiza alguna acción epistémica. En sentido estricto se trata de sistemas en que entra la temporalidad. Algunos de estos sistemas son representativos de alguna estrategia general que otros siguen con variantes menores.

A imitación de algunos autores, tomaré algunos enunciados metalingüísticos acerca de teoremas de $\mathbf{K}$ que manifiestan de manera muy intuitiva la omnisciencia lógica y veré si son verdaderos para los sistemas en cuestión. No se trata de fórmulas del sistema mismo. No pretendo con ello evaluarlos porque no es fácil determinar si algunos de estos enunciados representan, o no, un rasgo correcto, es decir, si describen o norman adecuadamente las creencias del agente. Me servirán, en cambio, como una forma de describir a los creyentes que los respectivos sistemas modelan.

1) Si $\models$ [1] β entonces $\models \mathcal{B}\beta$

Atribuye al agente la creencia en todas las verdades lógicas de ese sistema (en particular, en todas tautologías), lo que, por supuesto, también es inaceptable. El siguiente enunciado es:

2) $\models \mathcal{B}(\alpha \to \beta) \to (\mathcal{B}\alpha \to \mathcal{B}\beta)$

La fórmula representa la cerradura de las creencias del agente bajo implicación (material) creída.

[1] El símbolo "$\models$" se referirá en cada sección a la verdad lógica o a la consecuencia lógica del sistema que entonces esté en consideración.

3) Si $\models (\alpha \rightarrow \beta)$ entonces $\models (\mathcal{B}\alpha \rightarrow \mathcal{B}\beta)$

Creo que la elección de los tres esquemas anteriores es clara. Los siguientes, en cambio, son casos particulares de 1) y de 3). Desde luego que 1) y 3) tienen una infinidad de instancias (aún si nos restringimos a esquemas), así es que explicaré enseguida por qué elijo esos esquemas. Antes de ello recordaré lo que Dretske llamó operadores semipenetrantes y daré un ejemplo de por qué considerar esta posibilidad.

Dretske (1970) considera ciertos operadores que concatenados a un enunciado dan como resultado otro. Por ejemplo, "es verdad que…" o "Juan cree que…". Algunos de ellos tienen la siguiente característica: si P implica que Q, entonces $O(P)$ implica $O(Q)$. Dretske llama a estos operadores completamente penetrantes. Entre ellos se encuentran "es un hecho que…","es verdad que…","si R, entonces…", etc. Otros operadores no son de este tipo. Por ejemplo, aún si P implica Q no siempre es cierto que si deseas que P, deseas que Q, o que si es extraño que P, entonces debe ser extraño que Q. Por ejemplo, puede ser extraño que Pedro vaya a una escuela y vaya descalzo, pero no que vaya a la escuela. Ahora bien, Dretske sostiene que el fenómeno de la penetración tiene grados y de entre los operadores que no son completamente penetrantes distingue dos grupos. A los primeros los llama semipenetrantes; y, a los segundos, no penetrantes. Un operador $\mathcal{O}$ es un semipenetrante si, cuando β es consecuencia lógica de α, $\mathcal{O}\beta$ no necesariamente es consecuencia lógica de $\mathcal{O}\alpha$, pero hay cierto tipo de inferencias muy sencillas (pero no triviales)[2] tales que, si la inferencia de δ a γ es de este tipo, $\mathcal{O}\gamma$ se sigue de $\mathcal{O}\delta$. Por ejemplo, puede ser que $\mathcal{O}$ no sea completamente penetrante pero, en cambio, $\mathcal{O}\delta$ sea consecuencia lógica de $(\mathcal{O}(\delta \wedge \epsilon))$, es decir,$\mathcal{O}$ penetra en el caso de la inferencia que procede por eliminación de la conjunción. En ese caso, $\mathcal{O}$ es semipenetrante. La definición de "conocimiento" que daba Dretske (1971) por ese entonces y que pretendía capturar nuestras intuiciones al respecto, correspondía a un operador semipenetrante. Muy conocido es su ejemplo: el padre va con su hijo al zoológico y este le pregunta cómo se llaman ciertos animales blancos con rayas negras que están en una jaula en la que se lee "cebras". El padre le responde que son cebras. El niño lo cuestiona de si no serán más bien mulas ingeniosamente disfrazadas. A esto el padre le responde que no lo sabe. Es decir sabe una proposición, pero no una de sus consecuencias lógicas evidentes. En la definición de Dretske el conocimiento se da contra un fondo de alternativas plausibles que puede variar de las premisas a la conclusión de un argumento correcto, por lo que el conjunto de conocimientos de una persona, por

[2] Es decir, no de la forma de α se sigue α

definición, no es lógicamente cerrado. Sin embargo, su operador de conocimiento es semipenetrante.

El problema de la omnisciencia lógica podría reformularse diciendo que el operador de conocimiento (o de creencia) en los sistemas normales es completamente penetrante, lo cual es inadecuado. Pero cabe la posibilidad de que un operador semipenetrante fuese correcto para modelar o normar cierto tipo de creencia. Por eso es conveniente ver si, en los sistemas que expondré, los operadores epistémicos penetran en cierto tipo de inferencias. Muy especialmente me interesan la introducción de la conjunción y la disyunción, el *ex falso quodlibet* (o principio de explosión), y si es posible que un agente crea una contradicción explícita (es decir del tipo $P \wedge \neg P$). Ahora explicaré por qué elijo estas inferencias.

4) $\models (\mathcal{B}\alpha \wedge \mathcal{B}\beta) \rightarrow \mathcal{B}(\alpha \wedge \beta)$

Esta parece una afirmación problemática porque es justamente el tipo de inferencia que se viola en la paradoja del prefacio o de la lotería, De cada boleto de la lotería puedo estar casi segura que no ganará el premio (parece inverosímil que este boleto que tengo en la mano, y que es uno de 100000 similares, resulte el ganador). Debería entonces pensar que ninguno ganará, pero esto es absurdo. En segundo lugar, algunos autores afirman, fundados en evidencia empírica, que es posible atribuir a un individuo la creencia en una contradicción particular. Pero no es lo mismo decir que individuo cree P y también cree no P (tal vez no en el mismo momento) y otra es atribuirle o exigirle la creencia en el enunciado P y no P. Veremos que algunos sistemas distinguen estos dos casos. Como dije, no me ocuparé de la paradoja de la lotería, pero este segundo punto, relativo a las creencias inconsistentes, será importante en lo que sigue.

5) $\models \mathcal{B}\alpha \rightarrow \mathcal{B}(\alpha \vee \beta)$

Este principio es dual de 4) y es problemático porque enfrenta dos de las críticas de Harman. La primera es la de las demandas excesivas. Podría ocurrir que el individuo ni siquiera entendiera la proposición β por lo que sería arbitario atribuirle la creencia en $(\alpha \vee \beta)$ a partir de su creencia en α. Otra objeción que aparece aquí es la de las trivialidades: no es cierto que un individuo deba procurar creer que Sócrates fue un gran filósofo solo si cree que Sócrates fue un gran filósofo o que los dragones existen. Además la tautología que da pie a 6) como principio puente es una de las premisas del argumento de Pseudo-Escoto (Kneale, Kneale (1980): 261)[3] para sostener que de una premisa contradictoria se sigue cualquier propo-

[3]A saber, de $P \wedge \neg P$ se derivan (1) P y (2) $\neg P$. De (1) se deduce (3) $(P \vee Q)$, y de (2) y (3), Q

sición, así es que 5) (junto con otros principios) podría emplearse para sostener 6).

6) $\models \mathcal{B} \perp \rightarrow \mathcal{B}(\alpha)$

Es decir, que si el agente cree una contradicción entonces deberá creer cualquier cosa. Aquí debemos distinguir varios casos. El primero es que creer una contradicción, como ya había dicho, puede significar creer P y creer no P, o bien, creer P y no P (para una proposición particular P). Lo segundo es que un sistema podría validar 6) para una contradicción particular, pero no para otra. Por ejemplo, en un modelo de algún sistema puede ocurrir que una proposición de la forma "P y no P" carezca de valor de verdad en un mundo determinado, mientras que haya otra fórmula (denotada por $\perp$) que siempre sea falsa. Además cuando decimos que una persona se contradice normalmente lo hacemos fundados en que asevera un par de proposiciones, de una de las cuales puede derivarse (a veces via algunas proposiciones de sentido común) una negación de la otra. Así es que en que la afirmación de que se contradice estamos usando implícitamente una forma de cerradura lógica. Por último, un sistema puede validar 6) de dos maneras distintas. Una es el caso trivial en que el antecedente siempre sea falso por la semántica dada. El sistema no tolerará que al individuo crea una contradicción. En el otro, el sistema podría permitir el antecedente y validar el condicional. Para discernir estas posibilidades agrego la siguiente proposición:

7) $\models \neg\mathcal{B}(\alpha \wedge \neg\alpha)$

La fórmula no es un teorema de $\mathbf{K}$, pero sí de los sistemas normales que son extensión de $\mathbf{KD}$. Al adoptarlo convertiríamos cualquier contradicción del sujeto en una contradicción del sistema.

Ahora bien, aunque buscaré diversas interpretaciones de esas fórmulas en términos de posibles operadores deónticos implícitos. Usualmente oscilaré entre principios puente $\mathfrak{C}\mathcal{O}+$ y $\mathfrak{T}\mathcal{O}+$. Como antes dije, leeré las fórmulas en la versión $\mathfrak{C}\mathcal{O}+$, pero solo como una formulación más natural de $\mathfrak{T}\mathcal{O}+$ porque esta es la que enfrenta la objeción del constreñimiento. Así interpretaré $\models B(\alpha \wedge \neg\alpha) \rightarrow B\alpha$ como "si el agente cree una contradicción, entonces deberá creer cualquier cosa". El antecedente del condicional describe la situación del agente y el consecuente lo que debe hacer (en lo relativo a la creencia). Sin embargo, leeré 7) como "el agente no debe contradecirse" o, mejor dicho, "el agente tiene prohibido creer una proposición y su negación". Es decir, esta vez leo la fórmula como describiendo la situación del agente, sino como imponiéndole una obligación. La razón es que estoy usando la interpretación $\mathfrak{I}$ y, por ello, leo la fórmula $\neg\mathcal{B}(\alpha \wedge \neg\alpha)$ como "el agente debería procurar no creer una proposición y su negación". Desde luego esta prohibición la harán los sistemas para los que 7) es verdadero.

De un sistema que valide los axiomas conocidos como de introspección positiva y negativa respectivamente también son verdaderas las siguientes afirmaciones:
$\models \mathcal{B}\alpha \rightarrow \mathcal{B}\mathcal{B}\alpha$
$\models \neg\mathcal{B}\alpha \rightarrow \mathcal{B}\neg\mathcal{B}\alpha$

La primera simboliza que el sujeto cree que cree algo cuando lo cree, y la segunda, que no cree algo sólo si cree que no lo cree. En general no tomaré demasiado en cuenta estos dos esquemas por dos razones. La primera es que no es obvio si un sistema deba validarlos o no. Nuestras intuiciones no siempre arrojan al respecto un veredicto definitivo. La segunda es que el problema de la creencia de segundo orden es complejo y requiere tal vez un tratamiento aparte. Sin embargo, es interesante ver si un sistema contempla este tipo de extensiones y cómo las trata.

Un criterio de caracter representativo que tomaré en cuenta al considerar los sistemas lógicos es que la forma en que modelan la creencia no debe ser *ad-hoc*, sino que debe arrojar alguna luz en nuestros conceptos epistémicos. En este sentido veremos que los sistemas que solo admiten una presentación sintáctica tienen un desventaja frente a los que ofrecen también una semántica, siempre y cuando esta no sea, a su vez, un artefacto ad-hoc para tener un teorema de completud. Sin embargo, un sistema del primer tipo puede ser un avance parcial en la solución del problema. Para esto, tomaré en cuenta que diferentes concepciones de la creencia pueden requerir diferentes representaciones formales de la misma.

En lo que sigue revisaré algunos sistemas de lógica doxástica que han sido propuestos para tratar con el problema de la omnisciencia lógica.

3.3. La lógica de Levesque

Una tentativa muy interesante de tratar con la omnisciencia lógica fue hecha por Levesque en 1984. Este autor hace la distinción entre creencias explícitas, es decir, creencias que el agente sostiene de manera más o menos explícita; y creencias implícitas, es decir, las consecuencias lógicas de las anteriores. Para representar las primeras aparece de nuevo el problema de la omnisciencia lógica. El sistema de Levesque pretende modelar ambos tipos de creencia. Por supuesto, su innovación consiste en el tratamiento de la creencia explícita. Desde su perspectiva el problema de la omnisciencia lógica se origina en la representación de la creencia (o del conocimiento) a través de mundos posibles. En un mundo posible toda proposición es verdadera o falsa, pero no ambas cosas a la vez. Sin embargo, las creencias de un agente en una situación dada pueden ser modeladas por un asignación de valores de verdad solo a ciertas proposiciones que explícitamente cree en esa situación. Por ejemplo, las creencias de un niño que está jugando pueden incluir que Superman existe, pero no entiende la proposición de que el bronce es una aleación de estaño y cobre, menos aún está dispuesto a afirmarla o a negarla, a defenderla o a atacarla. Asimismo, en principio puede tener creencias incompatibles. Por ello, Levesque recurre al concepto de situación de Barwise y Perry. En una situación determinada una proposición puede ser solamente verdadera, solamente falsa, ambas cosas a la vez o ninguna de ellas. Si admitimos estas cuatro posibilidades y definimos los conectivos lógicos de manera usual, la lógica que obtenemos es la denominada por las siglas FDE. La idea de Levesque es otorgar al agente esa lógica, mientras el sistema conserva la lógica proposicional clásica.

FDE

Una manera de presentar FDE es como una lógica de 4 valores de verdad, B, N, 1, 0. Los dos últimos corresponden a los valores (solamente) verdadero y (solamente) falso, mientras los otros dos, a ser verdadero y falso a la vez, o ni una cosa ni la otra. Los conectivos lógicos son definidos de manera usual, simplemente tomando en cuenta que no ser verdadero no significa ser falso, etc. Por ejemplo, una conjunción es verdadera si y solo si ambos conyuntos son verdaderos; y falsa, si y solo si uno de ellos es falso. Es decir que si P tiene valor B y Q valor 0, entonces $(P \wedge Q)$ es falso, mientras que $(P \vee Q)$ es B. La negación de una proposición con valor B, también tiene valor B. Consideraremos un lenguaje cuyos conectivos lógicos son $\vee$, $\wedge$ y $\neg$. Aunque pude limitarme a los últimos dos, resultará más clara la exposición si utilizo los tres mencionados. Sus significados quedan consignados en las siguientes tablas de verdad:

Tabla 1

$\neg$	Q
1	0
B	B
N	N
0	1

Tabla 2

$\vee$	1	B	N	0
1	1	1	1	1
B	1	B	1	B
N	1	1	N	N
0	1	B	N	0

Tabla 3

$\wedge$	1	B	N	0
1	1	B	N	0
B	B	B	0	0
N	N	0	N	0
0	0	0	0	0

Los demás conectivos serán considerados como abreviaturas a la manera usual (en particular $(\alpha \to \beta) \equiv_{Def} (\neg\alpha \vee \beta))$. De nuevo la consecuencia lógica está dada por la preservación de la verdad, es decir, que si cada premisa de un argumento correcto tiene valor 1 o B, la conclusión también será 1 o B. Si llamamos a 1 y a B valores distinguidos podemos definir así la consecuencia lógica en FDE:

Definición 4. Si $\Sigma \cup \{\alpha\}$ es un conjunto de enunciados, decimos que α es consecuencia lógica de Σ si en cualquier asignación de valores de verdad a las letras proposicionales que otorgue un valor distinguido a cada elemento de Σ, también α recibe un valor distinguido. Como siempre, denotaremos esto como $\Sigma \models_{FDE} \alpha$.

Además del condicional material, introduciremos un condicional de caracter metalingüístico por medio de la siguiente definición:

Definición 5.

$$\alpha \models_{FDE} \beta \equiv_{Def} \models (\alpha \to_{FDE} \beta)$$

Como se puede apreciar, este condicional que expresa una relación de consecuencia lógica no puede figurar más que una vez en una fórmula y siempre como conectivo principal. Estrictamente hablando no es un conectivo. No tiene sentido la expresión $w \models_{FDE} (\alpha \to \beta)$. Cuando haya necesidad de distinguirlos, el símbolo $\to$ representará el condicional material y $\to_{ML}$ el metalingüístico.

Definición 6. Una fórmula es válida (en FDE) si y solo si es verdadera (es decir, tiene valor 1 o B) para cualquier asignación de valores de verdad a sus letras proposicionales.

Advierta que si $\models_{FDE} (\alpha \to \beta)$ y para una asignación α es verdadero, en ella obviamente β también será verdadero. En cambio, no sucede lo mismo si $\to$ es el condicional material. Puede ocurrir que para una interpretación α tenga el valor B y β sea falso. En ella, $(\alpha \to \beta)$ y α serán verdaderas, mientras β es falsa. En particular el *modus ponens* no es una regla válida para el condicional material.

El sistema de Levesque

Vemos ahora cómo define Levesque (1984) la semántica para que el agente tenga como lógica FDE, mientras que la lógica del sistema siga siendo clásica, y cómo relaciona la creencia explícita con la implícita.

El lenguaje del sistema incluye, además de los conectivos usuales de la lógica proposicional, los operadores monádicos $\mathcal{B}$ y $\mathcal{B_I}$ de creencia explícita y de creencia implícita respectivamente. Las fórmulas se obtienen a la manera usual a partir de las letras proposicionales, los conectivos y los operadores de creencia, con la salvedad de que un operador de creencia no puede quedar en el alcance de otro. La semántica del sistema está dada por la siguientes definiciones:

Definición 7. Un modelo CE es una cuaterna $< S, \Omega, \pi_t, \pi_f >$ donde S es un conjunto no vacío cuyos miembros son llamados "situaciones", Ω es un subconjunto de S (que representa las situaciones que podrían ser reales de acuerdo a lo que el agente cree), y donde tanto π_t como π_f son funciones de $S \times LP$ en $\{0, 1\}$, donde LP es el conjunto de letras proposicionales. A los elementos de Ω los llamaremos *situaciones distinguidas*.

Intuitivamente π_t determina qué letras proposicionales son verdaderas en una situación dada y, similarmente, π_f hace lo propio con la falsedad. Es decir, si $\pi_t(w, \rho) = 1$ y $\pi_f(w, \rho) = 1$ entonces podríamos decir que ρ en w tiene el valor B. En cambio, si $\pi_t(w, \rho) = 1$ y $\pi_f(w, \rho) = 0$ diríamos que el valor de ρ en w es 1 (verdadero y no falso).

Definición 8. Dado un modelo CE, $< S, \Omega, \pi_t, \pi_f >$, un elemento w de S es llamado un mundo (posible) o una situación completa si para cada letra proposicional ρ, $\pi_t(w, \rho) = 1$ o $\pi_f(w, \rho) = 1$, pero no ambos. Es decir, cada letra proposicional es o bien verdadera o bien falsa en w.

Definición 9. Dado un modelo CE, $< S, \Omega, \pi_t, \pi_f >$, un elemento w de S es llamado una situación incompleta si hay una letra proposicional ρ, tal que $\pi_t(w, \rho) = 0$ y $\pi_f(w, \rho) = 0$.

Definición 10. Dado un modelo $< S, \Omega, \pi_t, \pi_f >$, una situación $s \in S$ es inconsistente si para alguna letra proposicional ρ, $\pi_t(s, \rho) = \pi_f(s, \rho) = 1$

Definición 11. Un mundo w es compatible con una situación s si las letras proposicionales que son verdaderas en s son verdaderas en w y las letras proposicionales falsas en s son falsas en w.

Por supuesto, ningún mundo es compatible con una situación inconsistente.

Definición 12. Dado un modelo $< S, \Omega, \pi_t, \pi_f >$, y $\Upsilon \subset S$, sea $\varpi(\Upsilon) = \{w \in S | w$ es un mundo y es compatible con algún elemento de $\Upsilon\}$.

Definición 13. Definimos que una situación s soporte la verdad de α y que s soporte la falsedad de α (lo que denotaremos respectivamente $s \models_t \alpha$ y $s \models_f \alpha$) por las siguientes cláusulas recursivas. β y γ representan fórmulas y ρ una letra proposicional:

a) $s \models_t \rho$ si y sólo si $\pi_t(s, \rho) = 1$.

b) $s \models_f \rho$ si y sólo si $\pi_f(s, \rho) = 1$.

c) $s \models_t (\beta \wedge \gamma)$ si y sólo si $s \models_t \beta$ y $s \models_t \gamma$

d) $s \models_f (\beta \wedge \gamma)$ si y sólo si $s \models_f \beta$ o $s \models_f \gamma$

e) $s \models_t (\beta \vee \gamma)$ si y sólo si $s \models_t \beta$ o $s \models_t \gamma$

f) $s \models_f (\beta \vee \gamma)$ si y sólo si $s \models_f \beta$ y $s \models_f \gamma$

g) $s \models_t \neg\beta$ si y sólo si $s \models_f \beta$

h) $s \models_f \neg\beta$ si y sólo si $s \models_t \beta$

i) $s \models_t \mathcal{B}\alpha$ si y sólo si para todo $s \in \Omega$, $s \models_t \alpha$.

j) $s \models_f \mathcal{B}\alpha$ si y sólo si para algún $s \in \Omega$, $s \not\models_t \alpha$.

k) $s \models_t \mathcal{B}_{\mathcal{I}}\alpha$ si y sólo si para todo $s \in \varpi(\Omega)$, $s \models_t \alpha$.

l) $s \models_f \mathcal{B}_{\mathcal{I}}\alpha$ si y sólo si para algún $s \in \varpi(\Omega)$, $s \not\models_t \alpha$.

Note que una situación dada s debe soportar la verdad o la falsedad (pero no ambas cosas) de una fórmula $\mathcal{B}\alpha$. Que el agente no cree (explícitamente) que α puede representarse indistintamente por $s \models_f \mathcal{B}\alpha$ o por $s \not\models_t \mathcal{B}\alpha$. Lo mismo vale para el operador $\mathcal{B}_{\mathcal{I}}$ de creencia implícita.

Definición 14. Dado un mundo w decimos que una fórmula α es verdadera en w si $w \models_t \alpha$. De lo contrario diremos que α es falsa en w.

Aunque estrictamente hablando, las fórmulas sólo tienen valor de verdad en los mundos de un modelo, diré que si $s \models_t \alpha$ entonces α es verdadero en la situación s (y lo propio con la falsedad). Por último,

Definición 15. Una fórmula α es válida (lo que denotamos por $\models \alpha$ o $\models_{Lev} \alpha$) si para todo modelo CE $M =< S, \Omega, \pi_t, \pi_f >$ y mundo w de S, α es verdadera en w.

Definición 16. Un conjunto de fórmulas Σ es satisfacible si hay un modelo CE $M =< S, \Omega, \pi_t, \pi_f >$ y un mundo w de S, tal que si $\alpha \in \Sigma$, entonces α es verdadera en w.

Recuerde que definimos $\alpha \to \beta$ como $\neg\alpha \vee \beta$. Es decir que $s \models_t \alpha \to \beta$ si y sólo si $s \models_f \alpha$ o $s \models_t \beta$; y $s \models_f \alpha \to \beta$ si y sólo si $s \models_t \alpha$ y $s \models_f \beta$.

Como dije antes, las afirmaciones 1) y 2) siguientes no son equivalentes, ni una implica a la otra.

1. $s \models_t \alpha \to \gamma$ y,
2. si $s \models_t \alpha$ entonces $s \models_t \gamma$.

2) no implica 1) pues si α y γ carecen de valor de verdad en s, entonces 2) es verdadero y 1) es falso (porque, en ese caso, $(\alpha \to \gamma)$ carece de valor de verdad en s). Por otro lado, 1) no implica 2), pues si en s α es verdadero y falso, mientras que γ carece de valor de verdad, 1) es verdadero y 2) falso. Sin embargo, en mundos posibles 1) y 2) son equivalentes. Observemos que si α es una fórmula

sin operadores de creencia, y s una situación que no otorga valor de verdad a alguna letra proposicional presente en α, entonces α carece de valor de verdad en s. Por ende, si s es una situación distinguida, $B\alpha$ será falsa (y su negación verdadera en toda situación del modelo). Por ende, ninguna fórmula de esta forma será válida. El sistema no atribuye a priori ninguna creencia al sujeto. En cambio, hace válida toda tautología, pues la validez sólo se evalúa en mundos.

Es claro que si una fórmula α es verdadera en una situación s, entonces también es verdadera en todo mundo w compatible con s. Así que si B(α) es verdadera en un modelo, es decir, si α es verdadera en todas las situaciones de Ω, también lo será en todos los mundos de $\varpi(\Omega)$. Es decir que si $s \models_t \mathcal{B}\alpha$ entonces $s \models_t \mathcal{B}_{\mathcal{I}}\alpha$. La creencia explícita implica la creencia implícita.

Por otro lado, si $\mathcal{B}\alpha$ y $\models (\alpha \rightarrow \beta)$, entonces α es verdadero en todas las situaciones distinguidas y, por lo tanto, es verdadera en todos los elementos de $\varpi(\Omega)$. En en consecuencia, lo mismo sucede con β, es decir, $\mathcal{B}_{\mathcal{I}}\beta$ es verdadera en el modelo. De la misma manera, si Σ es un conjunto de enunciados cada uno de los cuales es creído explícitamente por el agente y β es consecuencia (clásica) de Σ, el agente creerá implícitamente β, lo que concuerda con la caracterización que dió Levesque de la creencia implícita.

Note además que el operador $\mathcal{B}_{\mathcal{I}}$ es cerrado lógicamente, en el sentido en que si $\models_t \alpha$ entonces $\models_t \mathcal{B}_{\mathcal{I}}\alpha$ y $\models_t \mathcal{B}_{\mathcal{I}}(\alpha \rightarrow \beta) \rightarrow (\mathcal{B}_{\mathcal{I}}\alpha \rightarrow \mathcal{B}_{\mathcal{I}}\beta)$.

Levesque ofrece una axiomatización que es sólida y completa con respecto a la semántica provista.

3.3.1. Evaluación

Una vez hechas estas observaciones, veamos cómo la lógica de Levesque afronta el problema de la omnisciencia lógica. Repasemos algunas de las enunciados que enlisté al principio del capítulo y veamos si son verdaderos para la lógica de Levesque:

1. Si $\models \alpha$ entonces $\models \mathcal{B}\alpha$
2. $\models \mathcal{B}(\alpha \rightarrow \beta) \rightarrow (\mathcal{B}(\alpha) \rightarrow \mathcal{B}(\beta))$
3. Si $\models (\alpha \rightarrow \beta)$ entonces $\models (\mathcal{B}(\alpha) \rightarrow \mathcal{B}(\beta))$
4. $\models (\mathcal{B}\alpha \wedge \mathcal{B}\beta) \rightarrow \mathcal{B}(\alpha \wedge \beta)$
5. $\models \mathcal{B}\alpha \rightarrow \mathcal{B}(\alpha \vee \beta)$

6. $\models \mathcal{B}(\beta \wedge \neg\beta) \rightarrow \mathcal{B}\alpha$

7. $\models \neg\mathcal{B}(\alpha \wedge \neg\alpha)$

Por supuesto, interpretaremos el condicional de las fórmulas como material. Nótese que cuando se trata de evaluar la validez de fórmulas cuyo operador principal es un condicional podemos suponer que el antencedente es verdadero en una situación s y verificar si el consecuente es verdadero en s, pues solo nos interesan los mundos. Como dije antes, esto no se cumple en el caso de situaciones en general.

1) es falso pues de que α sea verdadero en todo mundo completo, no se sigue que lo sea en las situaciones distinguidas.

2) es falso. Contrajemplo: si s' el único elemento de Ω y p es verdadera y falsa a la vez en s', mientras que q es falsa o carece de valor de verdad en s' entonces $(\mathcal{B}p \wedge (\mathcal{B}(p \rightarrow q)) \rightarrow \mathcal{B}q)$ no es verdadera en ningún mundo del modelo.

3) es falso pues el antecedente podría ser verdadero en todo mundo, mientras que α podría ser verdadero en la única situación distinguida y β falso o carecer de valor de verdad.

4) es verdadero.

5) es verdadero.

6) es falso.

7) es falso pues en el único mundo distinguido α podría ser verdadero y falso a la vez.

Estos resultados concuerdan con la intención de Levesque de suponer que podemos atribuir a un individuo creencias inconsistentes, sin que, por ello, debamos adjudicarle una creencia en cualquier proposición.

Es interesante observar que para encontrar algunos contraejemplos a las fórmulas anteriores que son inválidas recurrimos a situaciones incompletas entre las situaciones distinguidas, mientras que otras hallaron un contraejemplo solo en modelos con situaciones inconsistentes en Ω.

No hemos previsto ni en el lenguaje ni en la semántica que un operador quede en el alcance de otro, como lo especificamos al inicio de esta sección. Sin embargo, imaginemos que levantamos esta restricción y permitimos fórmulas del tipo $B\alpha$ donde α contiene el conectivo B. ¿Qué ocurre con las fórmulas de introspección positiva y negativa?:

IP) $\mathcal{B}\alpha \rightarrow \mathcal{B}\mathcal{B}\alpha$

IN) $\neg\mathcal{B}\alpha \rightarrow \mathcal{B}\neg\mathcal{B}\alpha$

IP) y IN) son verdaderas. Por ejemplo, si $\mathcal{B}\alpha$ es verdadero en un mundo w, entonces α es verdadero en toda situación distinguida y, por lo tanto, $\mathcal{B}\alpha$ también es verdadero en esas situaciones. Esto parece correcto pues estamos modelando la creencia explícita, pero en realidad estamos yendo contra las intenciones de Levesque pues, por ejemplo, aquí estamos suponiendo que las fórmulas pueden tener un valor de verdad en las situaciones distinguidas, aunque no sean mundos. En la semántica original solo es relevante a la validez de una fórmula en un modelo su valor de verdad en mundos posibles. Aunque la creencia de orden superior no es mi tema, menciono esta posibilidad de extender este sistema porque los sistemas que veré enseguida surgieron, en parte, por el deseo de subsanar la lógica de Levesque en este aspecto.

Siendo la axiomatización propuesta por Levesque completa con respecto a la semántica dada, podría servir también para caracterizar sintácticamente las fórmulas válidas. Sin embargo, me parece que una de las fortalezas de este sistema es su semántica. La idea de reemplazar mundos por situaciones parece adecuada a nuestra noción intuitiva de creencia, al menos hasta cierto punto. El conjunto de mis creencias puede no incluir una proposición ni su negación. Como resultado de la semántica ofrecida, el agente modelado puede creer una contradicción sin creer todas las proposiciones. Desde la lectura normativa que propongo, el sujeto de Levesque no tiene prohibido creer una contradicción, pero si llega a creerla no está obligado a creer cualquier cosa. En esa interpretación, si el agente cree una conjunción debe creer cada conyunto, y si cree una proposición debe creer cada disyunción que contiene a esa proposición como disyunto. Es decir, el operador de creencia es semipenetrante. El siguiente resultado genera una de las principales objeciones que se han hecho a este sistema.

Teorema 17. *Si* $\models_{FDE} \alpha \to_{ML} \beta$*, entonces* $\models_{Lev} (\mathcal{B}\alpha \to \mathcal{B}\beta)$.
Si $\models_{FDE} (\alpha_1 \wedge \cdots \wedge \alpha_n) \to_{ML} \beta$*, entonces* $\models_{Lev} ((\mathcal{B}\alpha_1 \wedge \ldots \wedge \mathcal{B}\alpha_n) \to \mathcal{B}\beta)$.

Es decir FDE es la lógica del agente, mientras la nuestra es clásica. Por ejemplo, de $\mathcal{B}(\alpha \to \beta)$ y $B\alpha$ no podemos deducir $\mathcal{B}\beta$, pues el agente no tiene el *modus ponens* en su lógica, pero de $(\mathcal{B}\alpha \to \mathcal{B}\beta)$ y $\mathcal{B}\alpha$ obviamente podremos deducir $\mathcal{B}\beta$.

Se sigue que si α y β no tienen operadores doxásticos y satisfacen las condiciones sintácticas para que β sea consecuencia de α, entonces si el individuo cree α, debe creer β. El problema es entonces que el individuo modelado es lógicamente omnisciente con respecto a la lógica FDE. Ello, por supuesto, le impide al sistema responder satisfactoriamente a las objeciones de demandas excesivas y trivialidades, aunque al constreñimiento pueda contestar con una lectura $\mathfrak{T}\mathcal{O}+$ de los principios puente generados a partir de la lógica FDE. Aunque la omnisciencia del sujeto con respecto a FDE sea suficiente motivo de insatisfacción, hay otro aspecto que debemos analizar.

Como observé, para refutar 2) se requiere la existencia de situaciones inconsistentes en el conjunto Ω. Hacen notar Fagin y Hapern que la fórmula $\models \mathcal{B}(\alpha \rightarrow \beta) \rightarrow ((\mathcal{B}\alpha \rightarrow \mathcal{B}(\beta \vee (\alpha \wedge \neg\alpha)))$ es válida. Ellos encuentran este resultado poco satisfactorio. Mientras que parece natural que situaciones incompletas (pero no inconsistentes) sean miembros de Ω (que representa las situaciones que podrían ser reales de acuerdo a lo que el agente cree), la presencia en Ω de situaciones inconsistentes no es tan satisfactoria. ¿Es correcto atribuirle a un agente creencias explícitas contradictorias? Yo creo que esa no es la única preocupación de Fagin y Halpern. Otra es que en la semántica de Levesque el sujeto cree una contradicción si todas las posibilidades epistémicas que concibe son inconsistentes. Si ya concebir una situación contradictoria parece difícil, más lo es que todas las concebibles sean inconsistentes. Supongo yo que de allí nace la queja de Halpern y Fagin.

Podría pensarse que es fácil resolver este “problema” haciendo un pequeño cambio en la semántica que consiste en suprimir las situaciones inconsistentes. Esta vez solo consideramos tres valores de verdad, a saber, V, F, N, y definimos los conectivos como lo hicimos antes, por ejemplo, $(\alpha \wedge \beta)$ es verdadero si y sólo si α y β son verdaderos; y es falsa si y solo si α es falsa o β es falsa. Se sobreentiende que en todos los demás casos $(\alpha \wedge \beta)$ tiene el valor N (o, en el sentido clásico, carece de valor de verdad).

Definición 18. Un KL-modelo es una terna $< S, \Omega, \pi_t, \pi_f >$, es que S es un conjunto no vacío (cuyos elementos serán llamados situaciones), $\Omega \subset S$ y π_t, π_f son funciones que a cada letra proposicional y a cada elemento de S asignan 1 o 0, pero si ρ es cualquier letra proposicional y si $\pi_t(\rho) = 1$, entonces $\pi_f(\rho) \neq 1$; y si $\pi_f(\rho) = 1$, entonces $\pi_t(\rho) \neq 1$.

Definición 19. Una fórmula es KL-válida si es verdadera en todo mundo de todo KL-modelo.

Definición 20. Una fórmula α es KL-consecuencia de un conunto de fórmulas Σ si en todo mundo de todo KL-modelo en que son verdaderas todas las fórmulas de Σ, α es verdadera.

Si ahora substituimos la lógica FDE por KL en la lógica de Levesque, es decir, concedemos al razonador la lógica KL, ¿qué resulta? Antes que nada no hay ninguna creencia que el agente deba tener por la simple semántica pues KL carece de verdades lógicas. Por otro lado, ya no tendremos contraejemplo para $\mathcal{B}\alpha$, $\mathcal{B}(\alpha \rightarrow \beta) \models \mathcal{B}\beta$. Esta afirmación ahora es verdadera, es decir, reinstauramos la cerradura lógica. Por ejemplo, $\not\models \mathcal{B}((\mathcal{Q} \wedge \neg\mathcal{Q}) \rightarrow \mathcal{R})$ pero $\models (\mathcal{B}(\mathcal{Q} \wedge \neg\mathcal{Q}) \rightarrow \mathcal{B}\mathcal{R})$, simplemente porque $\models \neg\mathcal{B}(\mathcal{Q} \wedge \neg\mathcal{Q})$.

En resumen, la principal objeción que se ha hecho al sistema de Levesque es que el agente modelado cree todas las consecuencias lógicas relevantes (es decir, FDE) de sus creencias. Podría ser mejor evaluado si lo consideramos como un sistema que muestra las normas con que debe razonar un agente para quien la lógica correcta es FDE. Aún así no toma en cuenta la cuestión de las demandas excesivas, ni la de trivialidades. No contempla la posibilidad de que el razonador tenga alguna creencia de orden superior. Por otro lado, yo considero como una virtud el que las fórmula $\mathcal{B}(\alpha \wedge \neg\alpha) \rightarrow \mathcal{B}\beta$ no sea válida. Es de suponerse que la lógica debe guiar al individuo que descubre que sus creencias son contradictorias, y no recomendarle que crea todo. Sin embargo, no proscribe las creencias contradictorias.

Veremos enseguida variantes al sistema de Levesque propuestos por Fagin y Halpern y por Lakemeyer, en intentos por subsanar lo que estos autores vieron como defectos del sistema original.

3.4. La Lógica de Lakemeyer

En 1987, Lakemeyer propuso un sistema que prentende mejorar el de Levesque en su tratamiento de la omnisciencia lógica. Se trata de un sistema también fundado parcialmente en la lógica FDE. ¿En qué pretende superar los intentos anteriores de enfrentar la omnisciencia lógica? Critica los enfoques sintácticos pues en ellos “los tipos de enunciados creídos pueden ser completamente arbitrarios porque dependen de la forma de los enunciados” (Lakemeyer,1987, p. 402). En el otro extremo se hallan los sistemas en que el sujeto cree con cada proposición todas sus equivalencias lógicas. Su objetivo parece ser de tipo descriptivo: “la principal contribución de este trabajo es que ofrece una semántica plausible para las creencias de un agente que puede mantener meta-creencias y es capaz de sacar de ellas inferencias de manera eficiente...”. En particular, así como no deseamos atribuirle al agente omisciencia lógica, tampoco sería conveniente que fuese incapaz de incrementar sus creencias a través de procesos sencillos de razonamiemnto.

El sistema llamado BLK tomará la semántica de mundos posibles (que es aceptable intuitivamente) con sólo dos variaciones: los modelos estarán formados por situaciones y por dos relaciones de accesibilidad, una para las creencias positivas y la otra para las negativas. Aunque los operadores de creencias podrán aplicarse iteradamente, ningún operador de creencia implícita estará en el alcance de un operador de creencia explícita. Me parece que esta es una decisión acertada. No es fácil interpretar una fórmula en que esta interacción ocurre. Lakemeyer observa con acierto que "vemos las creencias implícitas como caracterizaciones puramente externas de las creencias de un agente y de lo que se sigue de ellas" (p. 404). De nuevo, esto parece referirse a la atribución de creencias (implícitas). No tiene sentido atribuir a un individuo creencias explícitas sobre sus propias creencias implícitas. Sin embargo, sí sería sensato que un agente albergara explícitamente creencias acerca de las creencias implícitas de otro, lo que el sistema no permitirá.

El lenguaje de BLK comprende, además de los conectivos proposicionales usuales, un operador de creencia explícita $\mathcal{B}$ y otro de creencia implícita $\mathcal{B}_{\mathcal{I}}$ con las reglas de formación usuales, con la ya anunciada salvedad de que el operador $\mathcal{B}_{\mathcal{I}}$ no puede ocurrir en el alcance del operador $\mathcal{B}$. Enseguida defino los modelos de BLK:

Definición 21. Un modelo de BLK es una quíntupla $M = \langle S, \pi_t, \pi_f, R, \overline{R} \rangle$ donde:

1) S es un conjunto no vacío (cuyos miembros serán llamados situaciones).

2) π_t y π_f son funciones de LP en $\{0, 1\}$.

3) R y $\overline{R}$ son relaciones binarias de S.

4) $w \in S$ es un mundo si y solo si satisface las dos siguientes condiciones

$$\pi_t(w, \rho) = 1 \leftrightarrow \pi_f(w, \rho) = 0$$

para toda $\rho \in LP$, es decir, para toda letra proposicional y

$$\forall s \in S, wRs \leftrightarrow w\overline{R}s$$

5) Para cualesquiera mundos w y v y cualquier situación s:

$$(wRv \wedge vRs) \rightarrow wRs$$

$$(wRv \wedge wRs) \rightarrow vRs$$

Notación: dado un modelo denotemos por W al subconjunto de S formado por los mundos de S.

Definición 22. Dado un modelo de BLK, $M = \langle S, \pi_t, \pi_f, R, \overline{R} \rangle$, definimos qué significa que una situación s soporte la verdad o la falsedad de un enunciado α (lo que denotaremos respectivamente por $s \models_t \alpha$ y $s \models_f \alpha$) por las siguientes cláusulas recursivas:

a) $s \models_t \rho$ si y sólo si $\pi_t(\rho) = 1$.

b) $s \models_f \rho$ si y sólo si $\pi_f(\rho) = 1$.

c) $s \models_t (\beta \wedge \gamma)$ si y sólo si $s \models_t \beta$ y $s \models_t \gamma$

d) $s \models_f (\beta \wedge \gamma)$ si y sólo si $s \models_f \beta$ o $s \models_f \gamma$

e) $s \models_t (\beta \vee \gamma)$ si y sólo si $s \models_t \beta$ o $s \models_t \gamma$

f) $s \models_f (\beta \vee \gamma)$ si y sólo si $s \models_f \beta$ y $s \models_f \gamma$

g) $s \models_t \neg\beta$ si y sólo si $s \models_f \beta$

h) $s \models_f \neg\beta$ si y sólo si $s \models_t \beta$

i) $s \models_t \mathcal{B}\beta$ si y sólo si para todo $v \in S$ si sRv entonces $v \models_t \beta$.

j) $s \models_f \mathcal{B}\beta$ si y sólo si existe $v \in S$ tal que $s\overline{R}v$ y $v \not\models_t \beta$.

k) $s \models_t \mathcal{B}_{\mathcal{I}}\beta$ si y sólo si para todo $v \in W$ si sRv entonces $v \models_t \beta$

l) $s \models_f \mathcal{B}_{\mathcal{I}}\beta$ si y sólo si $s \not\models_t L\beta$

Adviértase que para una situación dada s de un modelo puede ocurrir que $s \models_t \mathcal{B}_i\alpha$ y $s \models_f \mathcal{B}_i\alpha$, o que $s \not\models_t \mathcal{B}_i\alpha$ y $s \not\models_f \mathcal{B}_i\alpha$, pero no si s es un mundo pues en ese caso R coincide con $\overline{R}$.

Definición 23. Dado un mundo w decimos que una fórmula α es verdadera en w ($w \models \alpha$), si $w \models_t \alpha$. De lo contrario diremos que α es falsa en w.

Definición 24. Un enunciado α es BLK-válido ($\models \alpha$ o $\models_{BLK} \alpha$) si es verdadero en todos los mundos de todos los modelos.

Lema 25. *Dado un mundo w,*

a) $w \not\models_t \mathcal{B}\alpha$ *si y sólo si* $w \models_f \mathcal{B}\alpha$

b) $w \not\models_t \mathcal{B}\alpha$ *si y sólo si* $w \models_t \neg\mathcal{B}\alpha$

Demostración. $w \not\models_t \mathcal{B}\alpha$ si y sólo si $\exists s \in S$ tal que wRs y $s \not\models_t \alpha$, y eso sucede si y solo si $\exists s' \in S$ tal que $w\overline{R}s'$ y $s' \not\models_t \alpha$, lo que es equivalente a $w \models_f \mathcal{B}\alpha$ y también a $w \models_t \neg\mathcal{B}\alpha$. □

Cor Si α no contiene operadores de creencia:

a) $\models_{Lev} \alpha$ si y solo si $\models_{BLK} \alpha$

b) $\models_{Lev} \mathcal{B}_i\alpha$ si y solo si $\models_{BLK} \mathcal{B}_i\alpha$

c) $\models_{Lev} \mathcal{B}_{\mathcal{I}i}\alpha$ si y solo si $\models_{BLK} \mathcal{B}_{\mathcal{I}i}\alpha$

En efecto, podemos comprobar que para fórmulas de este tipo (es decir, que están en el lenguaje de Levesque), sus condiciones de validez son las mismas en ambos sistemas. Por otro lado, es muy claro que $(\mathcal{B}_i\alpha \rightarrow \mathcal{B}_{\mathcal{I}i}\alpha)$ pero el condicional inverso no es válido. Dadas las condiciones impuestas a R y a $\overline{R}$ cuando se aplican a mundos (a saber, la claúsula (5) de la definición 15), es obvio que el operador $\mathcal{B}_{\mathcal{I}i}$ satisface los axiomas de introspección positiva y negativa.

Evaluación

Por lo dicho, BKL enfrenta el problema de la omnisciencia lógica de la misma forma que la lógica de Levesque, cuando no hay en las fórmulas operadores de creencia engarzados. Enseguida recuerdo al lector los enunciados que di al principio y cómo resultan para el sistema de Lakemyer:

1. Si $\models \alpha$ entonces $\models \mathcal{B}\alpha$ Falso.
2. $\models \mathcal{B}(\alpha \rightarrow \beta) \rightarrow (\mathcal{B}\alpha \rightarrow \mathcal{B}\beta)$ Falso.
3. Si $\models (\alpha \rightarrow \beta)$ entonces $\models (B\alpha \rightarrow B\beta)$ Falso.
4. $\models (\mathcal{B}\alpha \wedge \mathcal{B}\beta) \rightarrow \mathcal{B}(\alpha \wedge \beta)$ Verdadero.
5. $\models \mathcal{B}\alpha \rightarrow \mathcal{B}(\alpha \vee \beta)$ Verdadero.

6. $\models \mathcal{B}(\beta \wedge \neg\beta) \rightarrow \mathcal{B}(\alpha)$ Falso.

7. $\models \neg\mathcal{B}(\alpha \wedge \neg\alpha)$ Falso.

Ninguna necesidad tenemos para evaluarlos de la relación $\overline{R}$ (puesto que cuando parte de mundos coincide con R).

¿Qué ocurre con la creencia de orden superior? A pesar de que el modelo es transitivo y euclideano el operador de creencia explícita no satisface ni introspección positiva ni negativa. En efecto, supongamos que w y v son mundos del modelo y que s y s' son situaciones y que wRs, wRv, $w\overline{R}s$, $w\overline{R}v$ y que sRs'. Si $w \models \mathcal{B}\alpha$, entonces, obviamente, $v \models_t \alpha$ y $s \models_t \alpha$, pero nada impide que $s' \not\models_t \alpha$, en cuyo caso $w \not\models_t \mathcal{B}\mathcal{B}\alpha$. Las resticciones impuestas a los modelos solo incluyen el caso en que un mundo se haya relacionado con otro, y este con una situación. En cuanto a la introspección negativa, tenemos un resultado parecido: sean de nuevo w y v mundos y s y s' situaciones y supongamos que wRv, wRs y vRs'. Por transitividad y euclideaneidad vRs, wRs'; y las aserciones anteriores son válidas cuando R es substituida por $\overline{R}$. Si $v \not\models_t \alpha$ y $s' \not\models_t \alpha$, entonces $w \models \neg\mathcal{B}\alpha$ pero $w \not\models \mathcal{B}\neg\mathcal{B}\alpha$ pues w accede a s', pero s es inaccesible desde s'. Esto puede remediarse muy fácilmente si se desea modelar a un agente con introspección positiva o negativa (o normarlo al respecto). En cambio las dos siguientes fórmulas son válidas en BLK:

a) $(\mathcal{B}\alpha \rightarrow \mathcal{B}_{\mathcal{I}}\mathcal{B}\alpha)$

b) $(\neg\mathcal{B}\alpha \rightarrow \mathcal{B}_{\mathcal{I}}\neg\mathcal{B}\alpha)$

Ambas se deben a las condiciones impuestas en la cláusula 5 de la definición de “modelo de BLK”. Pues, por ejemplo, supongamos que $w \models_t \neg B\alpha$ donde w es un mundo del modelo. Entonces $w \models_f \mathcal{B}\alpha$, es decir, hay una situación s tal que wRs y $s \not\models_t \alpha$. Sea u otro mundo del modelo tal que wRu. Dado que wRu y wRs entonces uRs y, por lo tanto, $u \models_f B\alpha$. Con u es cualquier mundo tal que wRu, $w \models_t \mathcal{B}_{\mathcal{I}}\neg\mathcal{B}\alpha$. Este resultado es plausible, dada la forma en que el autor interpreta los dos tipos de creencia. Por ejemplo, consideremos a). Si el agente sostiene sinceramente una opinión α (y tiene, por tanto, la creencia α), es razonable atribuirle un creencia implícita de que tiene esta creencia explícita.

Supongamos que en el sistema de Levesque es posible iterar el operador $\mathcal{B}$ con las mismas condiciones semánticas (incluso podemos imaginar que por cada agente i y cada situación s, existe un conjunto de $\Omega_{i,s}$ de situaciones distinguidas). En cualquier caso, ¿qué distingue a BLK del sistema de Levesque cuando solo está involucrada la creencia explícita? En el caso de Levesque, por definición: $s \not\models_t \mathcal{B}\alpha$ si y solo si $s \models_f \mathcal{B}\alpha$ lo que, a su vez, es equivalente a $s \models_t \neg\mathcal{B}\alpha$. Por lo tanto, en ningún mundo w se da $w \models_t B(B\alpha \wedge \neg B\alpha)$. En ningún caso, el individuo puede creer que cree algo y, a la vez, que no lo cree. En cambio, en BLK las fórmulas $\mathcal{B}\alpha$ y $\neg\mathcal{B}\alpha$ son independientes cuando están evaluadas en una situación (que no es un mundo). Por ejemplo, supongamos una situación s tal que solo tiene acceso por medio de R a la situación en la cual α es verdadera, y que solo tiene acceso a través de $\overline{R}$ a una situación en que α es falsa. Entonces $s \models_t \mathcal{B}\alpha$ y $s \models_f \mathcal{B}\alpha$ (es decir, $s \models_t \neg\mathcal{B}\alpha$). Por tanto, $s \models_t (\mathcal{B}\alpha \wedge \neg\mathcal{B}\alpha)$. Si en el modelo hay un mundo w que sólo tiene acceso a través de R a s, entonces $w \models_t \mathcal{B}(\mathcal{B}\alpha \wedge \neg\mathcal{B}\alpha)$. Es decir, que un individuo puede creer que cree algo y que no lo cree, lo cual parece un defecto de BLK en comparación con el sistema de Levesque. Por otro lado, es satisfacible en BLK $(\mathcal{B}\mathcal{B}\alpha \wedge \mathcal{B}\neg\alpha)$ mientras que son válidas las fórmulas $\mathcal{B}\mathcal{B}(\alpha \wedge \beta) \rightarrow \mathcal{B}\mathcal{B}\alpha$ o $\mathcal{B}(\mathcal{B}\alpha \vee \mathcal{B}\beta) \rightarrow \mathcal{B}\mathcal{B}(\alpha \vee \beta)$. Esto último no es difícil de ver si recordamos que en la evaluación de fórmulas con el operador $\mathcal{B}$ y sin negación solo recurrimos a la relación R.

Surge la cuestión de cuál es exactamente la intención de Lakemmeyer o de cómo debemos interpretar sus fórmulas. Como antes dije, algunos párrafos de su artículo sugieren que tiene en mente una caraterización atributiva de la creencia. Sea de un modo u otro, el principal problema sigue siendo la omnisciencia del sujeto con respecto a la lógica FDE. Es verdad que la lógica FDE que sirve de base a este sistema cuenta con un algoritmo sencillo para la determinación de la consecuencia lógica y, en ese sentido, las demandas que el sistema hace al sujeto son menores, si bien aún excesivas. Sin embargo, esto también ocurría en la lógica de Levesque. La diferencia solo se advierte en el tratamiento de la creencia de orden superior. ¿Hay alguna ventaja intuitiva en usar dos relaciones de accesibilidad? No lo creo, aunque recuerda un poco a la semántica de vecindades de la que luego veremos dos ejemplos.

Los siguientes sistemas también ofrecen una mejora con respecto al de Levesque y se orientan hacia un tratamiento diferente del problema que me ocupa.

3.5. La Lógica de la Conciencia

Fagin y Halpern (1988) propusieron un sistema con un operador de conciencia que mejora en varios aspectos la propuesta de Levesque. Antes de verlo presentaré un sistema, también formulado por ellos, llamado la lógica de la conciencia, que constituye un estadio intermedio entre ambas propuestas. Entre las críticas que dirigen estos autores a la lógica de Levesque se encuentran estas dos: a) que una fórmula válida no es necesariamente verdadera en todas las situaciones y b) que el agente representado puede sostener explícitamente creencias contradictorias. Tal vez eso sería más tolerable si $\mathcal{B}(p \wedge q)$ no fuese equivalente a $(\mathcal{B}p \wedge \mathcal{B}q)$, lo que no es el caso, pues así podríamos suponer que el agente mantuvo alguna vez que p y, en otra ocasión, sostuvo que $\neg p$, pero nunca ambas al mismo tiempo. Sin embargo, creo que no es que a Fagin y Halpern les parezca inverosímil que una persona mantenga simultáneamente creencias obviamente contradictorias. Esto parece correcto en el caso de la atribución de creencias cuando consideramos a la lógica como constitutiva. Por el principio de caridad no atribuimos a una persona creencias de este tipo. Sin embargo, en la concepción representacionista, parece que un individuo podría llegar via una inferencia a la conclusión de que sus creencias implican P y también no P, en cuyo caso la lógica le prescribe que debe revisar su cuerpo de creencias para restaurar la consistencia. Pero, a mi juicio, Fagin y Halpern tienen otra razón más profunda para criticar a Levesque. En el sistema de este, el sujeto solo puede creer P y no P, si cada situación que considera como (epistemológicamente) posible es inconsistente. Es claro que si las situaciones distinguidas representan conjuntos de circunstancias que el sujeto concibe, no serán mundos posibles, en que todas las proposiciones tengan valor de verdad, pero que todas las situaciones que concibe sean contradictorias parece socavar la intuición original. Fagin y Halpern afrontan el reto de evitar estos "defectos" del sistema de Levesque, a la vez que retener sus virtudes en la representación de creencias explícitas e implícitas. Veamos cómo lo consiguen.

La lógica de la conciencia podrá modelar las creencias de varios agentes y la superposición de operadores doxásticos. Su lenguaje L contiene n operadores de creencia implícita: $\mathcal{B}_1$, $\mathcal{B}_2$,... $\mathcal{B}_n$ y otros tantos de creencia explícita: $\mathcal{B}_{\mathcal{I}1}$, $\mathcal{B}_{\mathcal{I}2}$,... $\mathcal{B}_{\mathcal{I}n}$. Además de los símbolos lógicos usuales, L contiene una constante $True$ que representa una fórmula siempre verdadera. Las fórmulas se definen como es usual, con el agregado de que $True$ es una fórmula y de que si α es una fórmula y Δ es un operador de creencia (implícita o explícita), $\Delta\alpha$ también es una fórmula. Es decir, un operador cualquiera de creencia puede quedar en el alcance de otro.

Definición 26. Un modelo para la lógica de la conciencia es una estructura $\langle W, \pi, A_1, ..., A_n, R_1, ..., R_n \rangle$ donde:

1) W es un conjunto no vacío, cuyos elementos serán llamados mundos posibles.

2) π es una función que a cada mundo y a cada letra proposicional le otorga un

valor de verdad (0 o 1). Es decir, $\pi : WXLP \Rightarrow \{0, 1\}$.

3) Cada A_i es un función que asigna a cada mundo del modelo un conjunto de letras proposicionales.

4) Cada R_i es una relación binaria de elementos de W que es transitiva, serial y euclideana, es decir, $R_i \subset WXW$, $\forall u \forall v \forall w((uR_iv \wedge vR_iw) \rightarrow uR_iw)$, $\forall u \exists v uR_iv$ y $\forall u \forall v \forall w((uR_iv \wedge uR_iw) \rightarrow vR_iw)$.

La función A_i determina de qué proposiciones atómicas (y por extensión de qué fórmulas) es consciente el iésimo agente en cada mundo. Ahora todas las "situaciones" serán completas, pero se conseguirá un efecto similar al de Levesque (sin creencias contradictorias) teniendo, además de la función $\models$ que otorgará un valor de verdad a cada fórmula en cada mundo, dos evaluaciones $\models_t^\psi$ y $\models_f^\psi$ relativas a un conjunto de letras proposicionales ψ. Si la fórmula es proposicional o no contiene operadores de creencia explícita, para la determinación de su valor de verdad en un mundo w, entrará en juego únicamente la evalución $\models$ en w aplicada a sus subfórmulas y a las letras proposicionales que en ella y en ψ figuran. En cambio, si la fórmula contiene al menos un operador de creencia explícita también serán relevantes en la determinación de su valor de verdad en un mundo w las evaluaciones relativas en w a un conjunto de letras proposicionales en mundos accesibles a w. Daré primeramente la definición recursiva de estas evaluaciones y más adelanté comentaré algunas de sus cláusulas y consecuencias.

Definición 27. Sea $M = \langle W, \pi, A_1, ..., A_n, R_1, ..., R_n \rangle$ un modelo para la lógica de la conciencia, $s \in W$, σ una letra proposicional, Ψ un conjunto de letras proposicionales y α y β fórmulas de L no atómicas. Definimos $\models$, $\models_t^\psi$ y $\models_f^\psi$ por la siguientes cláusulas recursivas:

1) $M, s \models_t True$.

2) $M, s \not\models_f True$

3) $M, s \models True$

4) $M, s \models_t^\Psi \sigma$ si y solo si $\pi(s, \sigma) = 1$ y $\sigma \in \Psi$.

5) $M, s \models_f^\Psi \sigma$ si y solo si $\pi(s, \sigma) = 0$ y $\sigma \in \Psi$.

6) $M, s \models \sigma$ si y solo si $\pi(s, \sigma) = 1$.

7) $M, s \models_t^\Psi \neg\alpha$ si y solo si $M, s \models_f^\Psi \alpha$.

8) $M, s \models_f^\Psi \neg\alpha$ si y solo si $M, s \models_t^\Psi \alpha$.

9) $M, s \models \neg\alpha$ si y solo si $M, s \not\models \alpha$

10) $M, s \models_t^\Psi (\alpha \wedge \beta)$ si y solo si $M, s \models_t^\Psi \alpha$ y $M, s \models_t^\Psi \beta$.

11) $M, s \models_f^\Psi (\alpha \wedge \beta)$ si y solo si $M, s \models_f^\Psi \alpha$ o $M, s \models_f^\Psi \beta$.

12) $M, s \models (\alpha \wedge \beta)$ si y solo si $M, s \models \alpha$ y $M, s \models \beta$.

13) $M, s \models_t^\Psi \mathcal{B}_i\alpha$ si solo si para cada $u \in W$ tal que sR_iu $M, u \models_t^{\Psi \cap A_i(s)} \alpha$.

14) $M, s \models_f^\Psi \mathcal{B}_i\alpha$ si solo si existe $u \in W$ tal que sR_iu y $M, u \models_f^{\Psi \cap A_i(s)} \alpha$.

15) $M, s \models \mathcal{B}_i\alpha$ si y solo si $M, s \models_t^{LP} B_i\alpha$.

16) $M, s \models_t^\Psi \mathcal{B}_{\mathcal{I}i}\alpha$ si solo si para cada $u \in W$ tal que sR_iu $M, u \models_t^\Psi \alpha$.

17) $M, s \models_f^\Psi \mathcal{B}_{\mathcal{I}i}\alpha$ si solo si existe $u \in W$ tal que sR_iu y $M, u \models_f^\Psi \alpha$.

18) $M, s \models \mathcal{B}_{\mathcal{I}i}\alpha$ si solo si para cada $u \in W$ tal que sR_iu $M, u \models \alpha$.

Definición 28. Una fórmula α es válida en la lógica de la conciencia si para todo modelo M y todo mundo posible s de M, $M, s \models \alpha$

Adviértase que $M, s \not\models \alpha$ es, por definición, equivalente a $M, s \models \neg\alpha$, así es que cada fórmula tiene un sólo valor de verdad en cada mundo. Si, por ejemplo, $\sigma \in \Psi$, entonces, por las cláusulas 4) y 5), $M, s \models_t^\Psi \sigma$ o $M, s \models_f^\Psi \sigma$, pero no ambas y, por lo tanto, $M, s \models_t^\Psi \sigma$ o $M, s \models_t^\Psi \neg\sigma$. Por otro lado, si $\sigma \notin \Psi$, entonces $M, s \not\models_t^\Psi \sigma$ y $M, s \not\models_f^\Psi \sigma$, pero si $\pi(s, \sigma) = 1$ entonces $M, s \models \sigma$; mientras que si $\pi(s, \sigma) = 0$ entonces $M, s \models \neg\sigma$. ¿Qué sucedera entonces con una proposición doxástica que involucre al mundo s y a la letra σ? Supongamos que wRs y que $\sigma \notin A_i(w)$. Por 13) sabemos que

$M, w \models_t^{LP} \mathcal{B}_i\sigma$ si solo si para cada $u \in W$ tal que sR_iu $M, u \models_t^{LP \cap A_i(w)} \sigma$.

Por hipótesis, lo de la derecha es falso, por lo tanto,

$M, w \not\models_t^{LP} \mathcal{B}_i\sigma$ y, por 15), $M, w \not\models \mathcal{B}_i\alpha$.

Finalmente, por 9), $M, s \models \neg\mathcal{B}_i\alpha$. Así es que esa indecisión sobre el valor de una letra en un mundo accesible desde w no causa una indecisión en la creencia en esa proposición.

Evaluación

El problema de la omnisciencia lógica general se evita en algunas de sus manifestaciones. Recuerdo al lector las fórmulas que di al inicio del capítulo:

1) Si $\models \alpha$ entonces $\models \mathcal{B}\alpha$

2) $\models \mathcal{B}(\alpha \to \beta) \to (\mathcal{B}(\alpha) \to \mathcal{B}(\beta))$

3) Si $\models (\alpha \to \beta)$ entonces $\models (\mathcal{B}(\alpha) \to \mathcal{B}(\beta))$

4) $\models (\mathcal{B}(\alpha) \wedge \mathcal{B}(\beta)) \to \mathcal{B}(\alpha \wedge \beta)$

5) $\models \mathcal{B}\alpha \to \mathcal{B}(\alpha \vee \beta)$

6) $\models \mathcal{B}(\beta \wedge \neg\beta) \to \mathcal{B}(\alpha)$

7) $\models \neg\mathcal{B}(\alpha \wedge \neg\alpha)$

(1) es falso pues, por ejemplo, $\mathcal{B}_i(p \vee \neg p)$ puede ser falso en un mundo s tal que $p \notin A_i(s)$. En ese caso si sR_iu $M, u \not\models_t^{A_i(s)\cap LP} p$. Por lo tanto, $M, s \not\models_t^{LP} \mathcal{B}_i(p \vee \neg p)$ y $M, s \not\models \mathcal{B}_i(p \vee \neg p)$.

(2) $(\mathcal{B}_i\alpha \wedge \mathcal{B}_{\rangle}(\alpha \to \beta)) \to \mathcal{B}_i\beta)$ sí es válida porque para que el agente *i* crea $(\alpha \to \beta))$ en una situación *s* las letras que figuran en α o en β deben pertenecer a $A_i(s)$.

(3) es falso, pues $\models (\alpha \to \beta)$ puede ser clásicamente verdadero, mientras que $(\mathcal{B}\alpha \to \mathcal{B}\beta)$ es falso en un situación en que el agente es inconsciente de las letras que aparecen en β. Por ejemplo, la fórmula $\mathcal{B}_iq \wedge \neg\mathcal{B}_i(q \wedge (p \vee \neg p))$ es satisfacible, es decir, $\mathcal{B}_iq \to \mathcal{B}_i(q \wedge (p \vee \neg p))$ no es válida.

(4) es verdadera por las mismas razones que (2). Por otro lado, $\models \mathcal{B}_i(\alpha \wedge \beta) \leftrightarrow (\mathcal{B}_i\alpha \wedge \mathcal{B}_i\beta)$

(5) es verdadera. Es fácil ver que $M, s \models^{LP} \mathcal{B}_i(\alpha \vee \beta)$ si y solo si existe u, tal que sR_iu y $M, s \models^{A_i(s)} \alpha$ o $M, s \models^{A_i(s)} \beta$. Sin embargo, advierta que esto permite adjudicarle al individuo creencias de cuyo contenido no es plenamente consciente.

(6) es verdadera porque el antecedente de la fórmula siempre es falso. De ahí el requisito de que las relaciones R_i sean seriales.

(7) es verdadera, pues para que $M, w \models \mathcal{B}_i(P \wedge \neg P)$ sería necesario y suficiente que para todo $v \in W$, tal que wRv sucediera que: $\pi(v, P) = 1, \pi(v, P) = 0$

y $P \in A_i(w)$ lo que es imposible.

Vemos que el agente no debe creer todas las tautologías ni se le exige que sus creencias sean cerradas bajo consecuencia lógica, pero sí bajo implicación lógica creída. Esto constituye una ventaja porque reduce la objeción de las demandas excesivas. No se resuelve aquí la paradoja de la lotería (4) ni la de trivialidades pues (por 5) el creyente está obligado a ver que crea P solo si cree $(P \vee Q)$, aunque nunca haya pensado en Q. No debe creer contradicciones (tal y como deseaban Fagin y Halpern) y si cree una contradicción entonces debe creer cualquier cosa. Es vacuamente explosivo.

Una de las ventajas de este sistema sobre el de Levesque es que, como el de Lakemeyer, tolera la superposición de operadores epistémicos. No es muy relevante a mi tema, pero es interesante ver cómo los autores solucionan este problema que se presentaba en el sistema de Levesque. Las evaluaciones $\models_f$ desempeñan un papel preponderante en la evaluación de fórmulas en que un operador de creencia explícita precede o sigue al de la negación. Un ejemplo de Fagin y Halpern muestra un interesante contraste con la lógica de Levesque, suponiendo que esta se extiende a operadores de segundo orden de manera natural. En esta, la fórmula $\mathcal{B}(p \vee \neg p)$ es falsa en todo mundo s si hay una situación distinguida en que p carece de valor de verdad. En ese caso, $\neg\mathcal{B}(p \vee \neg p)$ es verdadera en s y lo mismo ocurre con $\mathcal{B}(\neg\mathcal{B}(p \vee \neg p))$, por ello, $\neg\mathcal{B}(\neg\mathcal{B}(p \vee \neg p))$ no es lógicamente válida. En cambio, en la lógica de la conciencia las siguientes afirmaciones son equivalentes:

1. $s \models \mathcal{B}_j \neg\mathcal{B}_i(p \vee \neg p)),$
2. $s \models_t^{LP} \mathcal{B}_j \neg\mathcal{B}_i(p \vee \neg p))$
3. $\forall u(sR_j u \to u \models_t^{A_j(s) \cap LP} \neg\mathcal{B}_i(p \vee \neg p)),$
4. $\forall u(sR_j u \to u \models_f^{A_j(s) \cap LP} \mathcal{B}_i(p \vee \neg p)),$
5. $\forall u(sR_j u \to \exists v(uR_i v \wedge v \models_f^{A_j(s) \cap A_i(u) \cap LP} (p \vee \neg p))).$

Claramente esta última afirmación es siempre falsa aun si $i = j$. Por tanto, $\neg\mathcal{B}_j(\neg\mathcal{B}_i(p \vee \neg p))$, es lógicamente válida. La diferencia crucial con el sistema de Levesque está en que $M, s \models_t^{\Psi} \neg\mathcal{B}_i\alpha$ requiere que α sea evaluada como falsa (y no simplemente como no verdadera) en un mundo accesible a s.

Dado que las relaciones R_i son transitivas y euclideanas, las fórmulas $\mathcal{B}_{\mathcal{I}i}\alpha \to \mathcal{B}_{\mathcal{I}i}\mathcal{B}_{\mathcal{I}i}\alpha$, $\neg\mathcal{B}_{\mathcal{I}i}\alpha \to \mathcal{B}_{\mathcal{I}i}\neg\mathcal{B}_{\mathcal{I}i}\alpha$ son válidas en este sistema. No sucede lo mismo con el operador de creencia explícita: si $u \models \mathcal{B}_i\alpha$ entonces para todo $v \in W$, si uR_iv, entonces $v \models_t^{LP\cap A_i(u)} \alpha$, pero puede suceder que haya un mundo w tal que vR_iw (con lo cual uR_iw), tal que $w \not\models_t^{LP\cap A_i(u)\cap A_i(v)} \alpha$, con lo que $u \not\models \mathcal{B}_i\mathcal{B}_i\alpha$. Es decir, para evaluar una fórmula con varios operadores de creencia explícitos en un mundo u es necesario tomar en cuenta no solo las letras proposicionales de que el agente es conciente en ese mundo, sino también en algunos mundos próximos. Por otro lado la fórmula $(\mathcal{B}_i\mathcal{B}_i\alpha \to \mathcal{B}_\alpha)$ es válida, es decir, que si el agente cree que cree algo realmente lo cree, donde "cree" se refiere a la creencia explícita. Ese parece un resultado que concuerda con la intuición. La razón es que la fórmula $\mathcal{B}_i\mathcal{B}_i\alpha$ es verdadera en un mundo u si para cualesquiera mundos v, w, tales que uR_iv y vR_iw,

$$w \models_t^{A_i(u)\cap A_i(v)\cap LP} \alpha$$

Pero si v es tal que uR_iv entonces $uR_iv \wedge vR_iv$ (porque el modelo es euclideano). Por tanto,

$$v \models_t^{LP\cap A_i(u)\cap A_i(v)} \alpha$$

lo que implica que $v \models_t^{LP\cap A_i(u)} \alpha$ y, en consecuencia, que $u \models \mathcal{B}_i\alpha$. En el penúltimo paso utilicé un caso particular del siguiente lema.

Lema 29. *Si $\Gamma \subset \Theta \subseteq LP$ y $M, w \models_t^\Gamma \alpha$ entonces $M, w \models_t^\Theta \alpha$; por otro lado, si $M, w \models_f^\Gamma \alpha$ entonces $M, w \models_f^\Theta \alpha$.*

Revisando las cláusulas de la definición de verdad en un modelo podemos verificar esta aserción.

En mi opinión las fórmulas en que el operador de creencia implícita se halla bajo el alcance de un operador de creencia explícita son difíciles de interpretar. Por ejemplo, la fórmula $\mathcal{B}_i\alpha \to \mathcal{B}_i\mathcal{B}_{\mathcal{I}i}\mathcal{B}_{\mathcal{I}i}\alpha$ es válida, pero ¿qué significa $\mathcal{B}_i\mathcal{B}_{\mathcal{I}i}\mathcal{B}_{\mathcal{I}i}\alpha$? Asimismo es válida la fórmula $\mathcal{B}_i\neg\mathcal{B}_{\mathcal{I}i}\alpha \to \neg\mathcal{B}_i\alpha$.

Lema 30. *Sea $\Gamma \subseteq LP$, si $M, w \models_t^\Gamma \alpha$ entonces $M, w \models \alpha$; y si $M, w \models_f^\Gamma \alpha$ entonces $M, w \models \neg\alpha$.*

Este lema es fácil consecuencia del anterior. De aquí se deduce el siguiente lema.

Lema 31. $\models \mathcal{B}_i\alpha \to \mathcal{B}_{\mathcal{I}i}\alpha$

Demostración. $u \models \mathcal{B}_i\alpha$ significa que para todo v tal que uR_iv tenemos $v \models_t^{LP\cap A_i(u)} \alpha$ mientras que $u \models \mathcal{B}_{\mathcal{I}i}\alpha$ si y solo si para todo v tal que uR_iv tenemos $v \models \alpha$, así que este lema es consecuencia de los anteriores. $\square$

Cor Si $\models (\alpha \rightarrow \beta)$ y $\models \mathcal{B}_i\alpha$ entonces $\models \mathcal{B}_{\mathcal{I}i}\beta$

Esto concuerda con la idea de que las creencias explícitas son las consecuencias lógicas de las creencia implícitas.

Lema 32. *Si todas las letras proposicionales que aparecen en β figuran en α, $\models (\alpha \rightarrow \beta)$ y $\models \mathcal{B}_i\alpha$ entonces $\models \mathcal{B}_i\beta$*

Esta es un forma restringida de omnisciencia lógica que, a mi juicio, es más tolerable porque toma en cuenta parcialmente las limitaciones cognitivas del agente. Si ya es "consciente" de los átomos que aparecen en α y, por lo tanto en β, parecería más sencillo que advirtiera la dependencia lógica entre ambas y que, por lo tanto, estuviera obligado a creerla.

Fagin y Halpern hacen una observación que sugiere una simplificación y a la vez una extensión de este sistema, a saber, que hay una forma de expresar en el lenguaje la propiedad de ser conciente de una letra proposicional en un mundo. Es claro para cualquier $\rho \in LP$ que $\rho \in A_i(u)$ si y sólo si $u \models \mathcal{B}_i(\rho \vee \neg\rho)$. La lógica de la consciencia general, que veremos enseguida, explota este recurso y lo amplifica. Se basa también la observación de que la semántica ahora presentada genera la objeción de que, dada una fórmula muy complicada, la conciencia de sus letras no generara una conciencia de la respectiva proposición. Esto sugiere agregar a la sintaxis un operador de conciencia que no se limite a letras proposicionales.

Como dije, este sistema puede considerarse una variante del de Levesque con algunas mejoras. Las creencias del agente de Levesque no tienen que ser ser cerradas bajo implicación, ni bajo implicación creída. Su sistema tolera creencias contradictorias. Sin embargo, como dije, no toma en cuenta las limitaciones cognitivas del agente al hacerlo omnisciente con respecto a la lógica FDE. El de Fagin y Halpern no tolera inconsistencias. El conjunto de creencias del agente no es cerrado bajo consecuencia, pero sí bajo implicación creída. Por ello, (2) puede ser interpretado como un principio K, en versión doxástica, es decir, como prescribiendo creencias a un sujeto solo en la medida en que este reconoce las implicaciones correspondientes, aunque sean falsas. En ambos sistemas el operador de creencia explícita es semipenetrante: el individuo debe creer una conjunción, si cree uno de sus conyuntos, y no debe creer[4] una disyunción, si no cree ninguno de sus disyuntos. De nuevo, estoy leyendo las fórmulas con la interpretación deóntica que propuse. Lo que me importa resaltar es que ninguno de los dos sistemas parece afrontar la objeción de las trivialidades, pues por ejemplo, la lectura deóntica

[4]Más precisamente: debe no creer.

$WO+$ de la validez lógica representada por (5), sería: Debes procurar no creer α si no crees $\alpha \vee \beta$. La lógica le impondría al sujeto esta obligación aún si no es consciente de β, lo cual evidentemente es incorrecto en la medida en que estamos tratando de la creencia explícita.

Esta propuesta de Fagin y Halpern es interesante porque ofrece un puente con la lógica de la consciencia general que, a mi juicio, se inspira ya en una idea distinta. Analizaré esta propuesta enseguida,

3.6. La Lógica de la Consciencia General

En el mismo artículo en que se presentó el sistema anterior, Fagin y Halpern introdujeron otro que se recomienda, por lo pronto, por su sencillez. Introduzcamos en el lenguaje, además de n operadores de creencia explícita $\mathcal{B}_i$ y otros tantos de creencia implícita, $\mathcal{B}_{\mathcal{I}\rangle}$, el mismo número de operadores monádicos $\mathcal{A}_\rangle$ ($1 \leq i \leq n$) cada uno de los cuales representa una función A_i^* que en cada situación del modelo elige un conjunto arbitario de fórmulas (aquellas de las que el i-ésimo agente es consciente en esa situación). Las relaciones R_i son euclideanas, transitivas y seriales.

Definición 33. Un modelo para la lógica de la consciencia general es una estructura $\langle W, \pi, A_1^*, ..., A_n^*, R_1, ..., R_n \rangle$ donde:

1) W es un conjunto no vacío, cuyos elementos serán llamados mundos posibles.

2) π es una función que a cada mundo y a cada letra proposicional le otorga un valor de verdad (0 o 1). Es decir, $\pi : W \times LP \Rightarrow \{0, 1\}$.

3) Cada A_i^* es un función que asigna a cada mundo del modelo un conjunto de fórmulas.

4) Cada R_i es una relación binaria de elementos de W que es transitiva, serial y euclideana, es decir, $R_i \subset W \times W$, $\forall u \forall v \forall w((uR_iv \wedge vR_iw) \rightarrow uR_iw)$, $\forall u \exists v uR_iv$ y $\forall u \forall v \forall w((uR_iv \wedge uR_iw) \rightarrow vR_iw)$.

La relación de "verdad en un mundo" ($\models$) para este sistema (LGC) es definida recursivamente de la siguiente manera:

Definición 34. 1) $M, s \models$ True.

2) $M, s \models \rho$ si y sólo si $\pi(s, \rho) = 1$ (donde $\rho \in LP$)

3) $M, s \models \neg\alpha$ si y sólo si $M, s \not\models \alpha$

4) $M, s \models \alpha \wedge \beta$ si y sólo si $M, s \models \alpha$ y $M, s \models \beta$

5) $M, s \models A_i\alpha$ si y sólo si $\alpha \in A_i^*(s)$.

6) $M, s \models \mathcal{B}_i\alpha$ si y sólo $M, s \models \mathcal{A}_i\alpha$ y para todo v tal que sR_iv $M, v \models \alpha$.

7) $M, s \models \mathcal{B}_{\mathcal{I}i}\alpha$ si y sólo para todo v tal que sR_iv $M, v \models \alpha$.

Es decir, para que un agente crea explícitamente una fórmula α en un mundo w se requiere que esa fórmula sea verdadera en todos los mundos accesibles a w (para ese agente) y que en w el agente sea consciente de esa fórmula. En el sistema anterior, las funciones de conciencia (A_i) asociaban a cada mundo un conjunto de letras proposicionales. Con ello Fagin y Halpern recreaban las situciones incompletas de Levesque, utilizando únicamente mundos posibles. Si α y β no tienen letras en común, y α implica β clásicamente entonces la creencia en α no implica la creencia en β porque el agente puede ser consciente de las letras de α y no en las de β. En cambio, si todas las letras que aparecen en β ya estaban en α, entonces $\models \alpha \to \beta$. Las funciones A_i^* evitan este residuo de omnisciencia lógica sin permitir creencias inconsistentes. Aunque sean muy obvios menciono los dos siguientes resultados.

Lema 35. $\models \mathcal{B}_i\alpha \leftrightarrow (\mathcal{B}_{\mathcal{I}i}\alpha \wedge \mathcal{A}_i\alpha)$

Lema 36. $\models \mathcal{B}_i\alpha \to \mathcal{B}_{\mathcal{I}i}\alpha$

Evaluación

Veamos ahora como la lógica de la conciencia general enfrenta el problema de la omnisciencia lógica. Empecemos con las aserciones que utilizamos en la sección anterior:

1) Si $\models \alpha$ entonces $\models \mathcal{B}\alpha$

2) $\models \mathcal{B}(\alpha \to \beta) \to (\mathcal{B}\alpha \to \mathcal{B}\beta)$

3) Si $\models (\alpha \to \beta)$ entonces $\models (\mathcal{B}\alpha \to \mathcal{B}\beta)$

4) $\models (\mathcal{B}\alpha \wedge \mathcal{B}\beta) \to \mathcal{B}(\alpha \wedge \beta)$

5) $\models \mathcal{B}\alpha \rightarrow \mathcal{B}(\alpha \vee \beta)$

6) $\models \mathcal{B}(\beta \wedge \neg\beta) \rightarrow \mathcal{B}(\alpha)$

7) $\models \neg\mathcal{B}(\alpha \wedge \neg\alpha)$

Los primeros cinco incisos son falsos porque podemos excluir de la consciencia del sujeto cualquier fórmula que nos plazca. Por ejemplo, (2) es falsa, pues α y $\alpha \rightarrow \beta$ pueden estar en $A_i^*(s)$ sin que β lo esté. Por otro lado advirtamos que $\neg\mathcal{B}_i(\alpha \wedge \neg\alpha)$ es válida. Para cualquier mundo s, $s \models \mathcal{B}_i(\alpha \wedge \neg\alpha)$ si sólo si $(\alpha \wedge \neg\alpha) \in A_i^*(s)$ y existe un mundo v tal que sR_iv y $v \models (\alpha \wedge \neg\alpha)$. Lo segundo nunca ocurre. Por lo tanto, (7) $\models \neg\mathcal{B}_i(\alpha \wedge \neg\alpha)$ es verdadera, al igual que (6).

¿Qué pasa con la creencia de orden superior? A pesar de que los modelos son transitivos y euclideanos ni $(\mathcal{B}_i\alpha \rightarrow \mathcal{B}_i\mathcal{B}_i\alpha)$ ni $(\neg\mathcal{B}_i\alpha \rightarrow \mathcal{B}_i\neg\mathcal{B}_i\alpha)$ son válidos. Para verlo basta considerar que, por ejemplo, el agente puede ser conciente de α y no de $\mathcal{B}_i\alpha$.

Por otra parte, ni la fórmula $(\mathcal{B}p \wedge \mathcal{B}\neg p)$ ni $\mathcal{B}(p \wedge \neg p)$ son satisfacibles. Para facilitar la lectura en lo que sigue suprimiré el subíndice i cuando no sea ecesario. En general, podemos establecer que:

Lema 37. *Para toda fórmula proposicional α, $\neg\mathcal{B}\alpha$ es satisfacible.*

Demostración. Si α es falsa en algún mundo w de un modelo M, $\neg\mathcal{B}\alpha$ puede hacerse verdadera en un mundo v si lo agregamos al modelo y a R el par (v, w). Si α es una tautología, $\mathcal{B}\alpha$ será falsa en w si $\alpha \notin A_i^*(w)$. □

Lema 38. *Para cualesquiera fórmulas distintas α y β, $\neg(\mathcal{B}\alpha \rightarrow \mathcal{B}\beta)$ es satisfacible, excepto que α sea contradictoria.*

Demostración. La única manera en $\neg(\mathcal{B}\alpha \rightarrow \mathcal{B}\beta)$ sería insatisfacible, sería si α y β son tales que $(\mathcal{B}\alpha \rightarrow \mathcal{B}\beta)$ fuese válida. Podría ocurrir porque α fuese contadictoria, o porque en cada mundo en que $\mathcal{B}\alpha$ fuese verdadero $\mathcal{B}\beta$ también lo es, pero en ese caso podemos falsificar esta fórmula excluyéndola de la conciencia del sujeto. □

Es decir, cada vez que una fórmula cuyo operador principal es $\mathcal{B}$ tendría que ser verdadera (por ejemplo $\mathcal{B}\beta$), dadas ciertas condiciones (pongamos por caso $\mathcal{B}\alpha \wedge (\alpha \rightarrow \beta)$), podemos falsarla excluyéndola de la consciencia. Por ello, para el agente cuyas creencias norma esta lógica no es obligatorio creer el consecuente de un condicional que cree, aún si cree el antecedente, ni una verdad lógica, y le está permitido (no debe) creer una conjunción sin creer ninguno de sus conyuntos. Sin embargo, no es posible hacer que una fórmula que tendría que ser falsa dadas ciertas condiciones se vuelva verdadera. Por ello, si un agente cree una proposición y su negación debe creer todas las proposiciones, pero debe procurar no creer una proposición y también su negación. Recordemos que Fagin y Halpern veían como un defecto que el agente de Levesque creyera explícitamente una contradicción[5].

Esta lógica parece tener algunas ventajas sobre los sistemas anteriores. Es más sencilla que la de Levesque y permite evitar más problemas de omnisciencia lógica. Su diagnóstico respecto a este problema descansa íntegramente en la cuestión de la conciencia. En cierto sentido parece tomar en cuenta la objeción de las trivialidades e incluso la de las demandas excesivas. Recordemos que para superar la primera, tanto Field como Steinberger agregaban a sus principios puente una cláusula de "conciencia":"...y si el agente considera...". Sin embargo, la lógica de la conciencia general no supera completamente las demandas excesivas. Observemos, por ejemplo, que si una fórmula es válida y el agente es consciente de ella en un mundo, la creerá en ese mundo. Es decir, que el individuo estará obligado a (o tendrá razones para) creer una verdad lógica con la sola condición que sea consciente de ella. Es decir:

Si $\models \alpha$ entonces $\models (\mathcal{A}_i\alpha \rightarrow \mathcal{B}_i\alpha)$

Esto parecería tolerable si la consciencia implica un entendimiento suficiente de la fórmula para ver, por ejemplo, que es tautología si tal es el caso. Sin embargo,

Si $\models \alpha \rightarrow \beta$ y $M, w \models (\mathcal{A}_i\alpha \wedge \mathcal{A}_i\beta)$ entonces $w \models (\mathcal{B}_i\alpha \rightarrow \mathcal{B}_i\beta)$

Es decir, si el agente es consciente de dos fórmulas y una implica la otra, entonces deberá ver que si en una situación cree la primera, también crea la segunda. Esto sí impone una demanda excesiva en el agente, a menos que interpretemos el término "consciencia" de una manera inverosímil. Por ejemplo, tendríamos que suponer que quien es "consciente" de los axiomas de una teoría y de alguno de sus teoremas, debería ver que si cree los primeros, también debería creer el segundo. Sin embargo, en el sentido ordinario de las palabra "consciencia", esto no es verdadero ni una norma que el individuo pueda siempre seguir. Así, la objeción de las demandas excesivas queda en pie.

Además de esta obligación impuesta desde la lógica del observador, tendrá el sujeto ciertas restricciones impuestas por sus creencias implícitas bajo ciertas condiciones. Por ejemplo, es verdadero que:

$\models [\mathcal{B}_{\mathcal{I}i}(\alpha \rightarrow \beta) \wedge \mathcal{B}_{\mathcal{I}i}\alpha \wedge \neg\mathcal{A}_i(\alpha \rightarrow \beta) \wedge \neg\mathcal{A}_i\alpha] \rightarrow (\mathcal{A}_i\beta \rightarrow \mathcal{B}_i\beta))$

[5]Ellos aparentemente pensaban en una interpretación descriptiva de las fórmulas.

Es decir, si el sujeto cree implícita, pero no explícitamente un condicional y es consciente del antecedente entonces cree (o deberá creer) explícitamente el consecuente. Si las creencias implícitas son las consecuencias lógicas de las explícitas, este resultado es poco satisfactorio. Aunado a estas objeciones tenemos que la arbitrariedad con que pueden elegirse las funciones A_i^* produce resultados poco satisfactorios (en ciertos contextos) como que una persona pueda creer $(p \wedge q)$ y no $(q \wedge p)$ o que pueda estar consciente de una fórmula muy compleja, pero no de una de sus subfórmulas. Podría, además, ser consciente de que no es consciente de una fórmula, lo cual es muy contra-intuitivo.

El que los modelos sean transitivos y euclideanos no genera para la creencia explícita ni la introspección positiva ni la negativa. En efecto, como hemos visto, el agente no necesariamente cree que cree una proposición que cree, ni necesariamente cree que ignora una proposición que ignora. Desde luego, estos inconvenientes pueden ser vistos como el resultado de una virtud del sistema, a saber, que podemos elegir ciertas restricciones a las funciones A_i^* según los resultados que querramos modelar.

Eso es precisamente lo que Fagin y Halpern investigan enseguida. Mencionan brevemente las consecuencias, ventajas y desventajas que ciertas restricciones podrían tener. A continuación repaso algunas de estas observaciones por ser muy relevantes para mi tema.

Una posibilidad es adoptar la convención de que si $A_i^*(s)$ contiene una conjunción $(\alpha \wedge \beta)$ también contenga $(\beta \wedge \alpha)$ o, por ejemplo, contenga una fórmula si y solo si contiene su negación. En el primer caso, incluso podríamos suponer (posibilidad mencionada por Levesque) que $(\alpha \wedge \beta)$ y $(\beta \wedge \alpha)$ expresan la misma creencia, pero habría que dar otros criterios para identificar creencias. Aunque ambas propuestas parecen razonables, Fagin y Halpern mencionan casos en qué podrían no ser adecuadas. Puede darse el caso de que el agente no reconozca que una proposición es la negación de otra; o que una computadora determine que $(\alpha \wedge \beta)$ es falsa en un tiempo breve, pero no pueda calcular el mismo resultado para $(\beta \wedge \alpha)$, o tarde mucho en hacerlo. Con las restricciones mencionadas tendríamos, por supuesto, que $\models \mathcal{B}_i(\alpha \wedge \beta) \rightarrow \mathcal{B}_i(\beta \wedge \alpha)$ y, definiendo $(\alpha \vee \beta)$ como es usual por medio de la conjunción y la negación, obtendríamos que $\models \mathcal{B}_i(\alpha \vee \beta) \rightarrow \mathcal{B}_i(\beta \vee \alpha)$, pero no mucho más que eso. Es decir, aún tendríamos las limitaciones que son de esperarse en un enfoque meramente sintáctico.

Muy natural parecería que el operador $\mathcal{A}_i(s)$ fuese cerrado bajo subfórmulas. Si el proceso de comprensión de una proposición es composicional, es decir, que el agente la entiende a partir de la comprensión que tiene de sus partes, algo similar debería valer para la consciencia. No que si el individuo comprende las partes, entienda el todo, pero sí seguramente a la inversa. Esta elección tendría entre sus consecuencias que $\mathcal{B}_i(\alpha \rightarrow \beta) \rightarrow (\mathcal{B}_i\alpha \rightarrow \mathcal{B}_i\beta)$ sería válido. Este resultado podría considerarse una mejora porque toma en cuenta las limitaciones cognitivas del agente. Sus creencias serían cerradas, pero no bajo implicación a secas, sino bajo implicación creída por él, sea o no verdadera esta creencia.

Por otro lado, como antes, la fórmula $\mathcal{B}_i\alpha \rightarrow (\mathcal{B}_i\beta \rightarrow \mathcal{B}_i(\alpha \wedge \beta))$ no sería válida. Así es que esta sugerencia, que no es *ad hoc*, sino motivada por buenas razones, aporta una pequeña mejora. De esa forma obtendríamos un sistema a medio camino entre la lógica de la conciencia y el que ahora revisamos. Sin embargo, todo depende de la aplicación que queramos darle al sistema. Por ejemplo, Halpern y Fagin imginan que alguien puede reconocer la validez de $(\alpha \vee \neg\alpha)$ sin ser consciente (en el sentido de "entender") α. Sin embargo, no hago una objeción a esta versión particular, sino a la lógica de la conciencia general.

Hay otras opciones que mencionan los autores, no exactamente por ser recomendables, sino para mostrar la flexibilidad de este enfoque. Por ejemplo, podemos hacer que un agente sea consciente de que es consciente de todo aquello de lo que es consciente, etc.

Una axiomatización completa de esta lógica puede obtenerse observando que $\mathcal{B}_{\mathcal{I}i}$ funciona como un operador modal en el sistema $KD45$ y que $\mathcal{B}_i$ puede ser definido sin pérdida, como ya dije, por $\mathcal{B}_i\alpha \equiv (\mathcal{B}_{\mathcal{I}i}\alpha \wedge \mathcal{A}_i\alpha)$. En tanto que los operadores A_i no tengan ninguna restricción semántica, ningún axioma es necesario para regular su uso. Si queremos, por ejemplo, implementar las dos primeras restricciones mencionadas podemos agregar los esquemas axiomáticos: $\mathcal{A}_i(\alpha \wedge \beta) \leftrightarrow \mathcal{A}_i(\beta \wedge \alpha)$ y $\mathcal{A}_i\alpha \leftrightarrow \mathcal{A}_i\neg\alpha$. Si no ponemos limitaciones a las funciones A_i^*, el sujeto solo tendrá de la lógica principios puentes negativos. Por ejemplo, si cree P y cree $(P \rightarrow Q)$ no está autorizado a creer la negación de Q, pero tampoco está obligado a creer Q. Desde luego también pude interpretarlo en términos de principios $\mathfrak{T}$ o $\mathfrak{B}$ negativos.

Fagin y Halpern anuncian desde el inicio del artículo que los diferentes enfoques que presentan (de los cuales yo solo analicé dos) pueden ser útiles para fines distintos. Su idea es que la falta de omnisciencia se debe a que el agente a veces no sabe las reglas relevantes por no poder enfocarse en muchas materias simutáneamente (pp. 40-41). Una posibilidad para tomar en cuenta las limitaciones cognitivas de un agente, dicen los autores, sería acotar su consciencia a las proposiciones cuya verdad se puede decidir con un algoritmo sencillo en un lapso de tiempo prefijado. Evidentemente habría que explorar más cómo pueden implementarse estrategias de este tipo. Note que esto equivale a proponer una estretagia que limite la normatividad de la lógica a las consecuencias "obvias" de las creencias del sujeto, donde la obviedad se define en términos de la complejidad de un algoritmo.

Sin embargo, creo que para que una opción sea viable, la propiedad de la conciencia debe ser cerrada bajo subfórmulas. Con esta condición, aunque Fagin y Halpern no lo proponen, también podría implementarse una solución similar que concediera al sujeto una lógica diferente a la del sistema.

Enseguida veremos un sistema que aparece como dual de la lógica de la consciencia y, a la vez, como un complemento.

3.7. Principios y Conocimiento Explícito

En 1989, van der Hoek y Meyer propusieron un enfoque para tratar la omnisciencia lógica que puede considerarse el dual del de la lógica de la conciencia general. En lugar de substraer algunas fórmulas de las que el sujeto no es consciente y que, de otro modo, serían verdaderas, en este enfoque nuevas fórmulas pueden ser consideradas como verdaderas por el hecho de que el individuo tiene fe ciega en ellas, o porque expresan principios (tal vez implícitos) de su pensamiento, aun cuando bajo las condiciones semánticas ordinarias serían falsas. El lenguaje tendrá esta vez dos operadores monádicos $\mathcal{P}_i$ y $\mathcal{C}_i$ por cada agente representado. El primero consignará los principios o dogmas que el agente defiende o supone implícitamente como ciertos en una situación dada. El segundo funcionará como en los sistemas de Hintikka. Con ello podremos definir un nuevo predicado $\mathcal{B}_i$ como $\mathcal{B}_i = (\mathcal{C}_i \vee \mathcal{P}_i)$ que representará las creencias del individuo. Por supuesto que esta vez serán los predicados $\mathcal{B}_i$ los que eludirán la omnisciencia lógica[6]. Los modelos contendrán funciones monádicas ϕ_i (una correspondiente a cada operador $\mathcal{P}_i$) que asociarán a cada estado de S un conjunto de fórmulas (aquellas en las que el individuo i cree ciegamente en esa situación). Dado que no hay restricción al conjunto que es imagen de ϕ_i en un mundo, esta propuesta cae dentro del enfoque sintáctico. Llamemos al lenguaje L_m^P. Veamos un poco los detalles.

Definición 39. Un modelo para L_m^P es una estructura de la forma:

[6]Para facilitar la comparación uso el símbolo $\mathcal{B}$ para el operador que soluciona la omnisciencia lógica, pero hay que recordar que este operador representa una ampliación de las creencias implícitas.

$$M =< S, \pi, \phi_1, \phi_2, ...\phi_m, R_1, ...R_m >$$

Donde S es un conjunto no vacío (a cuyos elementos llamaremos estados), π es una función valuación que a cada estado y a cada letra proposicional otorga un valor de verdad, cada R_i es una relación binaria en S, que es serial, transitiva y euclideana; y, por último, ϕ_i es una función que asocia a cada estado de S un conjunto de fórmulas.

Las cláusulas recursivas de la definición de verdad de una fórmula en un mundo s son las ordinarias para los conectivos lógicos; así como para el operador $\mathcal{C}_\rangle$, es decir que:

$M, s \models \mathcal{C}_i\alpha$ si y sólo si para todo $v \in S$ si sR_iv entonces $M, v \models \alpha$

Mientras que para los operadores $\mathcal{P}_i$ y $\mathcal{C}_i$, las cláusulas son:

1) $M, s \models \mathcal{P}_i\alpha$ si y sólo si $\alpha \in \phi_i(s)$

2) $M, s \models \mathcal{C}_i\alpha$ si y sólo si $\alpha \in \phi_i(s)$ o $M, s \models B_i\alpha$

Es decir, $\mathcal{B}_i = \mathcal{P}_i \vee \mathcal{C}_i$.

De nuevo, validez es verdad en todos los estados de todos los modelos.

Evaluación

Tomemos la lista de fórmulas que hemos usado en las propuestas anteriores y veamos cómo la lógica de los principios y del conocimiento implícito las afronta.

1. Si $\models \alpha$ entonces $\models \mathcal{B}_i\alpha$.
2. $\models \mathcal{B}_i(\alpha \to \beta) \to (\mathcal{B}_i\alpha \to \mathcal{B}_i\beta)$.
3. Si $\models (\alpha \to \beta)$ entonces $\models (\mathcal{B}_i\alpha \to \mathcal{B}_i\beta)$.
4. $\models (\mathcal{B}_i\alpha \wedge \mathcal{B}_i\beta) \to \mathcal{B}_i(\alpha \wedge \beta)$.
5. $\models \mathcal{B}_i\alpha \to \mathcal{B}_i(\alpha \vee \beta)$.
6. $\models \mathcal{B}_i(\beta \wedge \neg\beta) \to \mathcal{B}_i(\alpha)$.
7. $\models \neg\mathcal{B}_i(\alpha \wedge \neg\alpha)$.

1) es claramente válida. Si una fórmula α es válida es verdadera en todos los estados accesibles desde s. Por lo tanto, $s \models \mathcal{B}_i\alpha$ y $s \models (\mathcal{P}_i\alpha \vee \mathcal{C}_i\alpha)$.

2) no lo es. Si, por ejemplo, un modelo consta de un sólo estado s tal que sRs y P es verdadera en s, mientras que Q es falsa y, por otro lado, $P_i(s) = \{(P \to Q)\}$, entonces $s \not\models \mathcal{B}_i(P \to Q) \to (\mathcal{B}_iP \to \mathcal{B}_iQ))$.

3) Es falsa. Suponiendo el antecedente verdadero, el consecuente es falso en s, si s tiene acceso a un mundo en que α y β son falsos, mientras que $P_i(s) = \{\alpha\}$.

4) es falsa. $\mathcal{B}_iP \wedge \mathcal{B}_iQ) \to \mathcal{B}_i(P \wedge Q)$ es falsa en el estado s si sRv, sRu y $\pi(v, P) = \pi(u, Q) = 0$, pero $P_i(s) = \{P, Q\}$.

5) es falsa. $(\mathcal{B}_iP \to \mathcal{B}_i(P \vee Q))$ es falsa en el estado s si sRv, y $\pi(v, P) = \pi(v, Q) = 0$, pero $P_i(s) = \{P\}$.

6) es falsa. $\mathcal{B}_i(P \wedge \neg P) \to \mathcal{B}_iQ$ es falsa en s si sRv, y $\pi(v, P) = \pi(v, Q) = 0$, pero $P_i(s) = \{(P \wedge \neg P)\}$.

7) es falsa. $\neg\mathcal{B}_i(P \wedge \neg P)$ es falsa en s si $P_i(s) = \{(P \wedge \neg P)\}$.

Ahora bien, las condiciones impuestas a las relaciones R_i harán que el operador $\mathcal{C}_\rangle$ aplicado a fórmulas que no contienen operadores $\mathcal{B}_j$ satisfaga los principios de introspección positiva y negativa y consistencia (es decir, $\mathcal{C}_i\alpha \to \neg\mathcal{C}_i\neg\alpha$), pero esas propiedades no se transmiten al operador $\mathcal{B}_i$. Por ejemplo, $\mathcal{C}_iQ$ y $\mathcal{C}_i\mathcal{C}_iQ$ pueden ser falsan es s, mientras $P_i = \{B_iQ\}$. En ese caso $(\mathcal{B}_iQ \to \mathcal{B}_i\mathcal{B}_iQ)$ es falsa en s.

La gran libertad que esto deja al creyente puede ejemplificarse por el hecho de que las siguientes fórmulas son satisfacibles en el enfoque de van der Hoek y Meyer:

- $\mathcal{B}_i(p \wedge q) \wedge \mathcal{B}_i\neg p \wedge \mathcal{C}_I\neg q$
- $\neg\mathcal{B}_ip \wedge \neg\mathcal{B}_i\neg p.$
- $\mathcal{B}_i(p \vee \neg p) \wedge \neg\mathcal{B}_ip \wedge \neg\mathcal{B}_i\neg p$
- $(\mathcal{B}_ip \wedge \neg\mathcal{B}_i(p \wedge (q \vee \neg q)))$
- $(\mathcal{B}_ip \wedge \mathcal{B}_i\neg p \wedge \neg\mathcal{B}_iq)$
- $(\mathcal{B}_ip \wedge \mathcal{B}_i\neg p)$
- $\mathcal{B}_i(p \wedge \neg p)$

En cambio, la siguientes fórmula no es satisfacible:

- $\neg\mathcal{B}_i(p \vee \neg p).$

Pues para ser satisfacible deben darse dos condiciones. La segunda es que en un i-sucesor de s no valga $(p \vee \neg p)$. Eso es imposible.

La primera cuestión que surge es si este sistema tolera una lectura normativa. Si aceptamos una interpretación CO $\mathfrak{CO}+$ o $\mathfrak{TO}+$, la lógica impondrá al agente ciertas obligaciones doxásticas algunas fundadas simplemente en sus prejuicios. Sin embargo, una lectura descriptiva o consitutiva (es decir, como normando la atribución de creencias) parece inadecuada. Evidentemente la propuesta surge de una intuición descriptiva: los individuos reales tienen prejuicios que motivan algunas de sus creencias. Para hacer la comparación con otros sistemas, yo olvidaré la motivación que el sistema implementa y leeré los resultados anteriores directamente, considerando a $\mathcal{B}_i$ como un operador de creencia explícita.

Al agente no le está prohibido creer una conjunción cuyos conyuntos no cree, e incluso creyendo la negación de cada uno de ellos. Además puede creer una disyunción sin creer en ninguno de sus disyuntos y no le está prohibido creer una contradicción, y puede creerla sin creer toda proposición. En cambio, debe creer una verdad lógica (o, mejor, debe procurar creerla). A esta lógica podemos hacer una objeción similar a la que hice para la lógica de la conciencia general. Si no hay ninguna restricción a la elección de las funciones π_i nada impide que el agente tenga $(P \wedge Q)$ entre sus principios básicos, pero no $(Q \wedge P)$. Ahora bien, esta parece una cuestión práctica. Es inverosímil encontrar individuos concretos con un conjunto tal de principios.

Dejando de lado los resultados anteriores y volviendo a la intuición básica del sistema, la idea de que un individuo en un punto determinado tenga acceso sólo a mundos en que vale la negación de P y a la vez tenga a P entre sus principios básicos parece contraintuitiva. La motivación para modelar la creencia en términos de mundos de Kripke parece diluirse. No podemos aquí implementar una estrategia similar a la sugerida para el caso de la lógica de la consciencia general. Parecería natural que un individuo que es consciente de un enunciado también lo sea de sus subfórmulas. Ahora está involucrada la verdad. El individuo acepta como verdaderos ciertos principios. ¿Cómo debemos restringir la atribución de principios (es decir, las funciones π_i) para normar o describir verosímilmente las creencias del agente? Lo único que la lógica le exige al individuo es (o lo faculta para) creer las verdades lógicas. Si α implica β, el individuo debe (o tiene razones para) creer $(\alpha \rightarrow \beta)$, pero eso no genera ninguna dependencia (deóntica) entre sus creencias con respecto a estas proposiciones.

Los autores mencionan la posibilidad de construir una lógica que combine operadores $\mathcal{A}_\rangle$ de consciencia con los operadores $\mathcal{P}_\rangle$ de principios. Esto nos permitiría substraer algunas proposiciones al conjunto de creencias Γ de un agente, aunque fuesen consecuencia lógica de otros elementos de Γ y, a la vez, agregar otras, a voluntad. Nótese, sin embargo, que el diagnóstico en cada caso es distinto. En el primero, la eliminación se funda en la inconsciencia del sujeto. En el segundo, la adjunción de creencias obedece a sus prejuicios. En todo caso, no estamos hablando de creencias en el mismo sentido y, por eso, los símbolos que estas lógicas emplean al introducir un operador nuevo son distintos. Un principio "explícito" de un agente puede ser uno del que no está consciente.

Además de las objeciones ya referidas, hay una más general que podría oponerse a otros enfoques similares. Una manera de expresarla es diciendo que en este enfoque la lógica no juega un papel preponderante en la solución de la omnisciencia lógica, sino que deja el problema en manos del usuario. Desde luego, el creyente modelado por este sistema ya no es lógicamente omnisciente pero, como he insistido a lo largo de este trabajo, eso no resuelve el problema de la omnisciencia lógica. En términos del capítulo anterior podríamos decir que esta lógica no supera la objeción del rigor. No impone ninguna restricción a las creencias de un sujeto, excepto que deben incluir los enunciados lógicamente válidos del sistema (es decir, de la lógica del observador). Se objetará que la propuesta es muy flexible y que permite imponer las resticciones a voluntad. A ello puedo responder dos cosas. La primera es que no hay tanta libertad si consideramos que el conjunto de creencias debe tener una estructura mínima (sin lo cual el sujeto no puede ser considerado racional, como lo señaló Cherniak), y que no hay restricciones de sentido común que puedan imponerse a las funciones π_i. La segunda es que, en todo caso, la lógica no determina cuál es esa estructura. Al dejar este punto a la decisión del usuario, no cumple la lógica ninguna función normativa.

3.8. NPL. Una Lógica No-Estándar

En 1990, Fagin, Halpern y Vardi propusieron un sistema de lógica epistémica o doxástica no para resolver, pero sí para mitigar[7] , el problema de la omnisciencia lógica, desarrollando de una manera ligeramente distinta las ideas de Levesque. El sistema de Levesque tenía una lógica clásica, mientras que el agente operaba de acuerdo a la lógica tetravalente (de FDE sin el condicional metalingüístico). Para ello el expediente consistía en formar el modelo con situaciones, pero definir la verdad en mundos (es decir, en situaciones consistentes y completas). Las

[7]Veremos más adelante en qué sentido el problema reaparece en su sistema, si bien de manera menos aguda.

situaciones solo tenían relevancia dentro de Ω, el conjunto de las situaciones distinguidas que determinaba las creencias del agente. La razón que motivaba esta decisión metodológica es que las situaciones solo son alternativas epistémicas en la mente confusa del agente que los toma como posiblemente reales, pero pueden no serlo en realidad. Tres obvias mejoras se sugieren. Una es que el conjunto Ω varíe en cada situación (es decir, reintroducir una relación de accesibilidad). La segunda es permitir varios agentes. Con lo primero las creencias del agente pueden variar de una situación a otra. La primera de estas innovaciones ya aparecía en la lógica de Lakemeyer; la segunda, en la lógica de la consciencia. De tales recursos echan mano Fagin, Halpern y Vardi, pero la mayor novedad de su propuesta es la tercera mejora, a saber, que ahora la verdad de una fórmula en un modelo no se determina solamente en mundos posibles, sino en todas las situaciones. Es decir, en un primer momento, la lógica del sistema coincidirá con la del agente. ¿Cuál es la motivación de los autores para hacer esta propuesta? Como en el sistema de Levesque el agente resultará lógicamente omnisciente, pero con respecto a una lógica más débil que la clásica: "consideramos las implicaciones de basar el conocimiento en una lógica no-estándar. La motivación básica es la observación... de que si debilitamos la "lógica" en "omnisciencia lógica" entonces quizás podamos disminuir la agudeza [de este problema]" (1990: 42).

Otra de sus motivaciones está explicada en el siguiente párrafo:

> Nuestro punto de partida al elegir una lógica no-estándar es la observación de que hay un número de propiedades de la implicación en la lógica estándar que parecen inapropiadas en ciertos contextos. En particular, considere una fórmula tal como $(p \wedge \neg p) \rightarrow q$. En lógica estándar es válida; esto es, de una contradicción uno puede deducir cualquier cosa. Sin embargo, considere una base de conocimiento en la cual los usuarios ingresan datos de tiempo en tiempo... es casi ciertamente el caso de que en una base de conocimiento grande habrá inconsistencias. Uno puede imaginar que en un punto un usuario ingresó el hecho de que el salario De Bob es de 50,000, mientras que en otro momento, quizás un diferente usuario registró que el salario de Bob es de 60,000. Así en lógica estándar cualquier cosa puede ser inferida de una contradicción... (Fagin, Halpern y Vardi, 1990, p.42)

Vemos que los autores imaginan que un agente o una comunidad alberga creencias contradictorias y que la lógica lo autoriza a extender sus creencias en cierta dirección. La lógica clásica es inapropiada en ese papel porque provee principios puente inadecuados. Probablemente dirían lo mismo de todas las implicaciones clásicas no relevantes (es decir, donde no hay información pertinente compartida por las premisas y la conclusión). Sin embargo, para satisfacer esas dos motivaciones los sistemas de Levesque y Lakemayer eran suficientes. ¿Qué cambia con esta nueva propuesta?

Aunque este sistema no se propone como una solución definitiva del problema, lo expongo por dos razones: a) este es uno de los pocos sistemas en que la lógica del sistema es la misma que la del agente, y c) es un sistema que opera con una lógica no clásica. Desde luego, en los sistemas anteriores podríamos haber conservado las intuiciones básicas con las que están construidos y utilizar una lógica no-cl'asica. Ahora veremos un ejemplo concreto.

Como ya dije, los autores toman como punto de partida la lógica de cuatro valores de verdad que sirvió de base a Levesque, pero la presentan de una manera distinta, siguiendo con ello a Routley. En lugar de hablar de situaciones, los modelos estarán formados por mundos en los que cada fórmula atómica tiene uno y sólo un valor de verdad (verdadero o falso). Sin embargo, para cada mundo w en el modelo habrá un mundo $w*$ (no necesariamente distinto a w) de tal manera que para cualquier fórmula α, $\neg\alpha$ será verdadera en w si y sólo si α es falsa en $w*$. Es decir, la negación será definida como un conectivo intensional. El resto de los conectivos se definirá a la manera usual. El valor de $\neg\alpha$ en un mundo será así independiente del valor de α en ese mismo mundo. Además, por definición $(w*)* = w$. Más precisamente:

Definición 40. Una estructura de Routley es una $n + 3$ tupla $\langle S, \pi, R_1, \ldots R_n, *\rangle$, donde:

1. S es un conjunto no-vacío cuyos elementos serán llamados "mundos posibles",

2. π es una función de $S \times LP$ en $\{0, 1\}$,

3. $R_i \subseteq S \times S$,

4. $*$ es una función de S en S tal que para cada $w \in S$, $(w*)* = w$.

Denotaré por NM al conjunto de estructuras de Routley y por $\models_{NM}$ a la relación de soporte caracterizada en la siguiente definición.

Definición 41. Sea $N = \langle S, \pi, R_1, \ldots R_n, *\rangle$ una estructura de Routley y $w \in S$

1) si ρ es una letra proposicional $w \models_{NM} \rho$ si solo si $\pi(w, \rho) = 1$

2) $w \models_{NM} \neg\alpha$ si solo si $w* \not\models_{NM} \alpha$

3) $w \models_{NM} (\alpha \vee \beta)$ si y solo si $w \models_{NM} \alpha$ o $w \models \beta$

4) $w \models_{NM} B_i\alpha$ si y sólo si para cada v tal que wR_iv, $v \models \alpha$.

Abreviamos $(\neg\alpha \vee \beta)$ como $(\alpha \rightarrow \beta)$. A la función $*$ se le conoce como la estrella de Routley.

Hay un cierto sentido en que los tratamientos con situaciones, por un lado, y con la estrella de Routley, por otro, son equivalentes. Claramente dado un modelo M con situaciones (en que las letras pueden tener 4 valores de verdad), podemos tener un modelo de Routley M' paralelo en el siguiente sentido. Si en una situación s una letra proposicional ρ es verdadera y falsa (es decir, $\pi_v(\rho) = 1$ y $\pi_v(\rho) = 0$), en M' tendremos el par de mundos s' y s'^* de tal manera que $M', s' \models \rho$ y $M', s'^* \not\models \rho$. De esta forma, $M', s' \models \neg\rho$. En cambio, si ρ es verdadera (pero no falsa) en s, $M', s' \models \rho$ y $s'^* = s'$. De esta forma, $M', s' \not\models \neg\rho$. De manera similar, si ρ es falsa (pero no verdadera) en s, $M', s' \not\models \rho$ y $s'^* = s'$. De esta forma, $M', s' \models \neg\rho$. Por último, si ρ no es verdadera ni falsa en s, $M', s' \not\models \rho$ y $M', s'^* \not\models \rho$. De esta forma, $M', s' \not\models \neg\rho$.

Llamemos LL al conjunto de todos los modelos de Lakemeyer. Consideremos un modelo LL $M =< W, \pi_t, \pi_f, R, \overline{R} >$. Construiremos ahora un modelo NM "paralelo" (en un senido que precisaré en breve) $M' =< W', R', V' >$. Su dominio será $\{s', s'^* | s \in W\}$, donde V' otorga valores a las letras proposicionales en s' y s'^* según las instrucciones arriba dadas. Definimos R' de tl manera que sRv si y solo si $s'R'v'$; y $s\overline{R}v$ si y solo si $s'^*R'v'$. Por último, $*(s') = s'^*$ y $*(s'^*) = s'$. El siguiente teorema es fácil de demostrar.

Teorema 42. *Dada una fórmula α, un modelo $M \in LL$ y s un mundo de M, entonces en el modelo "paralelo" $M' \in NM$ tenemos que:*

a) $M, s \models_t \alpha \Longleftrightarrow M', s' \models_{NM} \alpha$

b) $M, s \models_f \alpha \Longleftrightarrow M', s' \models_{NM} \neg\alpha$

No es sorprenderse que pueda probarse el condicional opuesto:

Teorema 43. *Dada una fórmula α, modelo $M \in NM$ y s un mundo de M, existen un modelo $M' \in LL$ y un mundo s' de M' tales que:*

a) $M, s \models_{NM} \alpha \Longleftrightarrow M's' \models_t \alpha$

b) $M, s \models_{NM} \neg\alpha \Longleftrightarrow M's' \models_f \alpha$

Los dos anteriores resultados muestran el sentido en que esta semántica es equivalente a la de la creencia explícita de Levesque (cuando ésta es extendida a varios agentes y a formulas con operadores doxásticos encajados) o a la de Lakemeyer. Hasta ahora no hemos hecho más que reformular esta última. La primera diferencia proviene de la siguiente definición.

Definición 44. Una fórmula α es válida (en la lógica no-estándar) ($\models_{NM} \alpha$) si es válida en todo mundo de todo modelo NM.

Recordemos que en las lógicas de Levesque y de Lakemeyer para determinar la validez de una fórmula solo se evaluaba en situaciones completas y consistentes. Por ello las tautologías seguían siendo válidas en esos sistemas. En cambio, no hay ninguna fórmula válida en el sentido de la definición anterior, como podemos fácilmemte constatar. Por ejemplo, si α es una fórmula sin operadores doxásticos, consideremos un modelo M en el que por cada letra proposicional ρ que ocurre en α y cada mundo w de M definimos $\pi(w)(\rho) = 0$ y $\pi(w^*)(\rho) = 1$. Entonces es fácil comprobar que en M, w α no es verdadera. De igual manera podríamos elegir π para hacerla verdadera. Definimos enseguida "consecuencia lógica" no-estándar:

Definición 45. Dado un conjunto $\Sigma \cup \{\alpha\}$, α es consecuencia lógica (no estándar) de Σ (lo que denotaré por $\Sigma \models_{NM} \alpha$ o simplemente $\Sigma \models \alpha$ cuando no haya riesgo de confusión), si para cada modelo NM en que son verdaderas todas las fórmulas de Σ, también α es verdadera.

Ahora bien, $\alpha \models_{NM} \beta$ no es equivalente a $\models_{NM} \alpha \rightarrow \beta$. Ninguna fórmula es válida en NM y, sin embargo, $p \models_{NM} p$ es trivialmente verdadera. Dado que $p \rightarrow p$ es definida como $\neg p \vee p$, esta fórmula que solo dice que si p no es falsa, entonces es verdadera, lo cual no tiene por qué ser el caso en la semántica dada. En cambio, en la lógica de Levesque esta fórmula es válida, porque la validez es verdad en todo mundo (es decir, en toda situación completa y coherente).

Ahora bien, ¿qué sucede en esta lógica con las proposiciones que he usado para expresar omnisciencia lógica? Como no hay ahora verdades lógicas, la primera y la tercera son trivialmente verdaderas (porque sus antecedentes son falsos) y las otras son falsas. Sin embargo, la aseveración 3) puede reemplazarse por: si $\alpha \models_{NM} \beta$ entonces $\mathcal{B}\alpha \models_{NM} \mathcal{B}\beta$,que es verdadera, así es que el sujeto sigue siendo lógicamente omnisciente, pero con respecto a la implicación NM, aunque no hay ninguna fórmula válida que tenga que creer. El problema surge porque hasta ahora la lógica NM no tiene modo de expresar la implicación material en el lenguaje. Es decir, no hay fórmula ϕ tal que $\alpha \models_{MN} \beta$ si y solo si $\models_{MN} \phi$.

Para solventar esta dificultad Vardi, Fagin y Halpern introducen un condicional que será apto para expresar en el lenguaje objeto la relación de implicación:

Definición 46. Sea $\hookrightarrow$ un conectivo binario tal que $M, s \models \alpha \hookrightarrow \beta$ si y solo si $M, s \not\models \alpha$ o $M, s \models \beta$

Note que si α y β carecen de valor de verdad en un mundo w (es decir, $w \not\models \alpha$, $w \not\models \beta$, $w^* \models \alpha$, $w^* \models \beta$), entonces $w \models (\alpha \hookrightarrow \beta)$, pero $w \not\models (\alpha \rightarrow \beta)$. Por otro lado, en FDE definimos un condicional metalingüístico, pero note que es un condicional cuyo valor de verdad no puede variar de un mundo a otro ni puede aplicarse a fórmulas que ya lo contengan y es, por lo tanto, distinto, al representado por el $\hookrightarrow$. Por un lado, $\hookrightarrow$ puede figurar varias veces en una fórmula y, por otro, una fórmula que lo contiene puede ser verdadera en un mundo y falsa en otro del mismo modelo. Sea $L^{\hookrightarrow}$ el lenguaje que se obtiene agregando este conectivo y sea NLP el fragmento proposicional de $L^{\hookrightarrow}$ con su interpetación por estructuras no-estándar. Por la manera en que fue definido $\hookrightarrow$, el siguiente teorema es obvio:

Teorema 47. *Si α y β no contienen el operador $\hookrightarrow$, entonces $\alpha \models_{NM} \beta$ si y solo si $\models_{NLP} \alpha \hookrightarrow \beta$*

Ahora tenemos fórmulas lógicamente válidas, por ejemplo, $(P \wedge Q) \hookrightarrow Q)$ y $(P \hookrightarrow (P \vee Q))$. De hecho, si α y β son fórmulas sin operadores epistémicos y sin los nuevos conectivos, $\models_{FDE} (\alpha \rightarrow_{ML} \beta)$ si y solo si $\models_{NLP} (\alpha \hookrightarrow \beta)$. Sin embargo, esta lógica no es relevante pues $\models_{NLP} (\alpha \hookrightarrow (\beta \hookrightarrow \alpha))$.

Ahora bien, tampoco tenemos una negación, es decir, no hay una fórmula β tal que $M, w \models_{NM} \beta$ si y solo si $M, w \not\models_{NM} \alpha$, pero con $\hookrightarrow$ es posible definir un conectivo que cumpla con esta condición.

Para expresar la negación introducen los autores un símbolo constante, $\perp$, sea como abreviatura de la negación de una fórmula válida elegida arbitariamente, o bien como un símbolo primitivo que es una fórmula y siempre es falsa. La segunda opción es más clara. Ahora introducimos un símbolo de negación (fuerte):

Definición 48. Sea $\sim$ una conectivo unario definido por $\sim \alpha \equiv_{Def} (\alpha \hookrightarrow \perp)$

Intuitivamente, $\sim \alpha$ es verdadero en un mundo w cuando α es falso o carece de valor de verdad en w.

Teorema 49. *Si $M \in NM$ y $s \in M$, $M, s \models_{NM} \sim \alpha$ si solo si $M, s \not\models_{NM} \alpha$*

¿Es NLP paraconsistente? No. Claramente $\models_{NLP} (\perp \hookrightarrow \beta)$.

Evaluación

En la formulación del sistema axiomático usé $\mathbf{K}$ en lugar de $\mathbf{B}$ porque hay algunas características del sistemas que lo aproximan más a ser una representación del conocimiento que de la creencia. Pero el punto más importante era ilustar una estrategia para tratar el problema de la omnisciencia lógica. Dado que, en este sistema el conectivo $\hookrightarrow$ representa la consecuencia lógica (en el sentido en que $\rightarrow$ la representa en la lógica clásica) y $\sim$ la negación (externa), verificamos qué sucede con las siguientes fórmulas:

1. Si $\models \alpha$ entonces $\models \mathcal{B}\alpha$.

2. $\models \mathcal{B}(\alpha \hookrightarrow \beta) \rightarrow (\mathcal{B}(\alpha) \hookrightarrow \mathcal{B}(\beta))$.

3. Si $\models (\alpha \hookrightarrow \beta)$ entonces $\models (\mathcal{B}(\alpha) \hookrightarrow \mathcal{B}(\beta))$.

4. $\models (\mathcal{B}(\alpha) \wedge \mathcal{B}(\beta)) \hookrightarrow \mathcal{B}(\alpha \wedge \beta)$.

5. $\models \mathcal{B}\alpha \hookrightarrow \mathcal{B}(\alpha \vee \beta)$.

6. $\models \mathcal{B}(\beta \wedge \sim \beta) \hookrightarrow \mathcal{B}(\alpha)$.

7. $\models \sim \mathcal{B}(\alpha \wedge \sim \alpha)$.

Sin sorpresa todas son verdaderas. El agente es omnisciente, pero con respecto a la lógica NLP. La idea central del sistema propuesto por Fagin, Halpern y Vardi es preservar la omniscienncia lógica (tal y como esta se manifiesta en el sistema modal K), pero substituir la lógica clásica por otra que, respecto al problema que nos ocupa, es más verosímil. Me parece que esta vía es digna de mayor exploración. Como dije antes, este es también un ejemplo de un sistema con una lógica no-clásica que el sistema y el agente comparten.

3.9. Razonamiento Local

En el mismo artículo de 1988 al que ya hice mención, Fagin y Halpern presentaron un enfoque distinto que, sin necesidad de postular situaciones inconsistentes, permite al agente tener creencias inconsistentes. La idea es que la creencia es contextual: un agente puede albergar diferentes creencias en distintos contextos o conglomerados. Es decir, la inconsistencia surge porque las creencias del individuo provienen de diferentes ámbitos de su mente. Por ejemplo, puede ser alguien que tenga creencias religiosas y a la vez una cosmovisón científica. No es que conciba situaciones contradictorias, sino que sus creencias se almacenan en distintos compartimentos, separados tal vez por la fuente de dónde fueron tomadas.

Con este fin, Fagin y Halpern postulan modelos en los cuales las relaciones de accesibilidad no conectan a un estado (o mundo) con otros, sino con conjuntos (llamados "cúmulos") de mundos. Es decir, tendremos que desde un estado s son accesibles uno o varios cúmulos. Si un enunciado es verdadero en todos los mundos de uno de esos cúmulos, entonces el agente en s creerá en él. Veamos los detalles.

Definición 50. Un modelo cúmulo es una estructura $\langle S, \pi, C_1, C_2, ...C_n \rangle$ donde

1. $S \neq \emptyset$
2. $\pi : (LP \times S) \rightarrow \{1, 0\}$ (donde LP es el conjunto de letras proposicionales).
3. $C_i S \rightarrow \mathcal{P}(\mathcal{P}(\mathcal{S}))$. Para cualquier $s \in S$, $C_i(w) \neq \emptyset$ y si $z \in C_i(w)$, $z \neq \emptyset$.

Como es usual, S es un conjunto no vacío (de elementos llamados "estados"), π una función que asigna a cada letra en cada estado un valor de verdad (0 o 1), mientras que n corresponde al número de agentes y cada C_i es una función que asigna a cada estado un subconjunto de $\mathcal{P}(\mathcal{S})$, más precisamente un conjunto no vacío de subconjuntos no vacíos de S (llamados cúmulos). La idea es que $C_i(s)$ indica los contextos o marcos de referencia que el agente i considera en el estado s.

Ahora tenemos para el agente i dos operadores de creencia, $\mathcal{B}_i$ y $\mathcal{B}_{\mathcal{F}i}$, que Fagin y Halpern denominan respectivamente "de creencia explícita" y "de creencia implícita". Sin embargo, advierten que las nociones que intentan capturar no son las analizadas con los sistemas anteriores. Yo prefiero llamarlas respectivamente "de creencia débil" y "de creencia fuerte" para hacer más visible este contraste. El primero es el que representa las creencias del agente en cada contexto. El segundo las consecuencias lógica anteriores. Por supuesto, es el primero el que soluciona la omnisciencia lógica, pero no representa las creencias explícitas del agente en el sentido en que este sea forzosamente consciente de ellas. Puede tratarse de creencias que regulan sus acciones en un cierto dominio. Las cláusulas que definen las condiciones de verdad para enunciados con operadores booleanos son la usuales, mientras que las que definen el funcionamiento de estos operadores son las siguientes:

$M, s \models \mathcal{B}_i\alpha$ si y sólo si $\exists T \in C_i(s)$ $\forall w \in T$, $M, w \models \alpha$

$M, s \models \mathcal{B}_{\mathcal{F}i}\alpha$ si y sólo si $\forall w \in \bigcap\{T|T \in C_i(s)\}$, $M, w \models \alpha$

La idea es que en cada estado s y para el agente i hay un cúmulo de mundos $C_i(s)$ que son accesibles desde este estado para ese agente. Cada uno de ellos representa lo que el individuo cree en un cierto contexto. De esta manera podemos parafrasear las condiciones anteriores de la siguiente manera: para que un agente i en un estado s crea débilmente una proposición α, es necesario que haya un cúmulo accesible desde s (para i) tal que, en todos los mundos de ese cúmulo, α sea verdadera; y para que un agente i en un estado s crea fuertemente una proposición α, es necesario que para todo mundo que esté en todos los cúmulos accesibles desde s (para i) α sea verdadera. Los autores parecen suponer implícitamente que la intersección de esos cúmulos no es vacía pues si $\bigcap_{T \in C_i(s)} = \emptyset$ entonces para toda α, $s \models \mathcal{B}_{\mathcal{F}\rangle}\alpha$. Por ejemplo, supongamos que $C_i(s) = \{\{w_1, w_3, w_4\}, \{w_2, w_3, w_4, w_5, w_6\}, \{w_3, w_4, w_6\}, \{w_1, w_3, w_4\}\}$. Si α es verdadera en w_1, w_3, y w_4, $s \models B_i\alpha$, porque α es verdadera en todos los mundos del primer cúmulo. En cambio si β es verdadera únicamente en w_1, w_2 y w_4, el agente no creerá (ni débil ni fuertemente) en β en s. Por último si γ es verdadera en w_3 y w_4 entonces i creerá fuertemente en γ porque w_3 y w_4 son los únicos mundos comunes a todos los cúmulos.

Si α es verdadera en todos los mundos de un cúmulo, diremos que es verdadera en ese cúmulo. Como siempre que una fórmula sea válida significa que es verdadera en todo estado de todo modelo.

Ahora bien $\models \mathcal{B}_i\alpha \rightarrow \mathcal{B}_{\mathcal{F}i}\alpha$ pues si $\models \mathcal{B}_i\alpha$ entonces α es verdadera en todos los mundos de un cúmulo T. Entonces es verdadera en todos los mundos que son comunes a todos los cúmulos (mundos que están incluidos en T). Por otro lado, es claro que $s \models (\mathcal{B}_{\mathcal{F}i}\alpha \rightarrow (\mathcal{B}_{\mathcal{F}i}(\alpha \rightarrow \beta) \rightarrow \mathcal{B}_{\mathcal{F}i}\beta))$ y que si $\models (\alpha \rightarrow \beta)$ y $\models \mathcal{B}_{\mathcal{F}i}\alpha$ entonces $\models \mathcal{B}_{\mathcal{F}i}\beta$. Sin restricciones en las funciones C_i el operador $\mathcal{B}_{\mathcal{F}i}$ no satisface ni la introspección negativa ni la positiva.

Evaluación

Antes dije que el malestar de Fagin y Halpern con la teoría de Levesque era no tanto que postulara agentes con creencias contradictorias, sino la forma en que modelaba este fenómeno. En efecto, para que el agente de Levesque crea una contradicción es necesario que solo conciba situaciones en que subsista esta contradicción. Parece extraño que un individuo conciba una situación contradictoria. Fagin y Halpern parten de un diagnóstico diferente y lo plasman en otro tipo de semántica.

> Nuestra observación clave es que una razón por la cual la gente tiene creencias inconsistentes es que las creencias tienden a venir en conglomerados independientes. Casi podemos ver a un agente como una sociedad de mentes, cada una con su propio conjunto (o conglomerado) de creencias, las cuales pueden contradecirse unas con otras. Este fenómeno puede ocurrrir incluso en la ciencia. El físico Eugen Wigner... notó que las dos grandes teorías con las cuales los físicos razonan son la teoría de los fenómenos cuánticos y la teoría de la relatividad. Sin embargo,... Wigner pensó que las dos teorías podrían muy bien ser incompatibles. (Fagin, Halpern, 1988, p. 58)

De nuevo evaluemos esta lógica en relación a la omnisciencia lógica considerando la validez y satisfacibilidad de ciertas fórmulas. Empecemos con la lista que utilizamos en las secciones anteriores:

1. Si $\models \alpha$ entonces $\models \mathcal{B}_i\alpha$
2. $\models \mathcal{B}_i(\alpha \to \beta) \to (\mathcal{B}_i\alpha \to \mathcal{B}_i\beta)$
3. Si $\models (\alpha \to \beta)$ entonces $\models (B_i\alpha \to B_i\beta)$
4. $\models (\mathcal{B}_i\alpha \wedge \mathcal{B}_i\beta) \to \mathcal{B}_i(\alpha \wedge \beta)$
5. $\models \mathcal{B}_i\alpha \to \mathcal{B}_i(\alpha \vee \beta)$
6. $\models \mathcal{B}_i(\beta \wedge \neg\beta) \to \mathcal{B}_i(\alpha)$
7. $\models \neg\mathcal{B}_i(\alpha \wedge \neg\alpha)$

(1) es verdadera.

(2) es falsa. $\mathcal{B}_i(\alpha \rightarrow \beta) \rightarrow (\mathcal{B}_i\alpha \rightarrow \mathcal{B}_i\beta)$, no es válida pues α puede ser verdadera en un cúmulo y $(\alpha \rightarrow \beta)$ en otro. En la intersección de ambos cúmulos β será verdadera, pero la intersección no necesariamente formará un cúmulo.

(3) es claramente verdadera.

(4) es falsa. De nuevo, α puede ser verdadera en un cúmulo y β en otro, sin que $(\alpha \wedge \beta)$ sea verdadera en todos lo mundos de un cúmulo.

(5) claramente es verdadera.

(6) es verdadera pues su antecedente es falso (no hay mundo en que $(\alpha \wedge \neg\alpha)$ sea verdadera). Por lo mismo,

(7) es verdadera.

En cambio, $((\mathcal{B}\alpha \wedge \mathcal{B}\neg\alpha) \rightarrow \mathcal{B}(\gamma))$ puede ser falsa, si hay un cúmulo en que α es verdadera y otro en que es falsa en todo mundo del cúmulo, mientras que γ no es verdadera en ningún cúmulo.

Por otro lado, ni ($\mathcal{B}_i\alpha \rightarrow \mathcal{B}_i\mathcal{B}_i\alpha$), ni $(\neg\mathcal{B}_i\alpha \rightarrow \mathcal{B}_i\neg\mathcal{B}_i\alpha)$ son válidos.

Otra manera de formular los resultados anteriores es observando que las fórmulas (o conjuntos de fórmulas) siguientes que son satisfacibles.

- $\mathcal{B}_i(p \wedge q) \wedge \mathcal{B}_i\neg p \wedge \mathcal{B}_i\neg q$
- $\neg\mathcal{B}_i p \wedge \neg\mathcal{B}_i\neg p.$
- $\mathcal{B}_i(p \vee \neg p) \wedge \neg\mathcal{B}_i p \wedge \neg\mathcal{B}_i\neg p$
- $\{\mathcal{B}_i p, \mathcal{B}_i(p \rightarrow q), \neg\mathcal{B}_i q\}$
- $\{\mathcal{B}_i p, \mathcal{B}_i\neg p, \neg\mathcal{B}_i q\}$
- $\{\mathcal{B}_i p \wedge \mathcal{B}_i\neg p)\}$

mientras que los siguientes no son satisfacibles:

- $\neg\mathcal{B}_i(p \vee \neg p).$
- $\{\mathcal{B}_i p, \neg\mathcal{B}_i(p \wedge (q \vee \neg q))\}.\ B_i(p \wedge \neg p).$

Es decir, el agente deberá creer todas las verdades lógicas y las consecuencias lógicas de todo lo que cree. En cambio, no se requiere que crea las que él cree que son consecuencias lógicas de sus creencias. Por otro lado, no le está prohibido creer una conjunción creyendo la negación de cada uno de sus conyuntos. Asimismo puede creer una disyunción sin creer en ninguno de sus disyuntos. No debe creer una contradicción explícita $(\alpha \wedge \neg\alpha)$, mientras que debe creer una verdad lógica. Esto es de esperarse porque las verdades lógicas son comunes a todos los cúmulos. Por otro lado, puede creer una proposición y también creer su negación, pero no por ello debería (o tendría razones para) creer todas las proposiciones.

Bajo la condición:

$$\text{para todo } w, v \in S, \text{ si } v \in T \in C_i(w) \text{ entonces } T \in C_i(v)$$

tenemos que $\models \mathcal{B}_i\alpha \to \mathcal{B}\mathcal{B}\alpha$. Pues si $\models_w \mathcal{B}_i\alpha$ entonces existe un bloque T en $C_i(w)$ tal que para todos los $v \in T$ tenemos que $v \models \alpha$ pero también $v \models \mathcal{B}_i\alpha$ pues $T \in C_i(v)$.

Por otro lado, bajo la condición de que:

$$\forall T \in C_i(s) \forall t \in T C_i(t) \subseteq C_i(s)$$

tenemos que $\models \neg\mathcal{B}_i\alpha \to \mathcal{B}_i\neg\mathcal{B}_i\alpha$.
pues, si $s \models \neg\mathcal{B}_i\alpha$ entonces para cada $T \in C_i(s)$ existe $v \in T$ tal que $v \not\models \alpha$. Ahora debemos probar que $s \models \mathcal{B}_i\neg\mathcal{B}_i\alpha$, es decir, que hay un $T \in C_i(s)$ tal que para todo $v \in T$ $v \models \neg\mathcal{B}_i\alpha$. Tomemos un $T \in C_i(s)$ y $r \in T$. Por la condición de arriba, $C_i(r) \subseteq C_i(s)$. Por lo tanto, para cada $T \in C_i(r)$ existe $v \in T$ tal que $v \not\models \alpha$ y, por lo tanto, $r \not\models \mathcal{B}_i\alpha$, de lo que se concluye que en cada $r \in T$, $r \models \neg\mathcal{B}_i(\alpha)$. Por lo tanto, $s \models \mathcal{B}_i\neg\mathcal{B}_i\alpha$.

Este sistema está basado en la intuición de que un agente puede creer algo en un determinado contexto, mientras que, en otro, no creerlo o incluso creer su negación. Como hemos visto permite descarticular ciertas fórmulas que modelan a un agente omnisciente, si éstas contienen varios operadores de creencia, pero no si sólo contienen uno. A mi juicio, no representa una solución satisfactoria al problema de la omnisciencia lógica pues mientras que exige al agente que sus creencias sean cerradas bajo implicación creída, pero no bajo consecuencia lógica. Es decir, rechaza los principios K (o, más bien, B) pero no formas más agudas de omnisciencia lógica. Es decir, que esta lógica no toma en cuenta las limitaciones cognitivas del agente. Sin embargo, en lo que se refiere a las creencias inconsistentes aventaja a otros sistemas porque si bien el agente debe evitar creer una contradicción explícita, no le recomienda creer cualquier cosa si cree una proposción y además cree su negación. El recurso a conglomerados o cúmulos fue novedoso en lógica epistémica y, como veremos enseguida, dio pie a otros sistemas.

3.10. Modelos Fusión

Un enfoque similar al de Fagin y Halpern que acabamos de ver fue propuesto por Jaspars en 1991 y 1993. Las relaciones R_i pondrán en contacto a algunos mundos con un conjunto no vacío de conjuntos de mundos. Las condiciones de evaluación de los operadores B_i serán diferentes.

¿Cuál es la motivación central de Jaspars? Es similar a la de Fagin y Halpern y también se basa en una forma diferente de diagnosticar y representar el problema de las creencias inconsistentes:

> En este artículo nos enfocaremos en el caracter imposible que los mundos pueden tener como alternativas doxásticas. Aquí nos concentraremos únicamente en representar la creencia inconsistente y sus consecuencias con respecto a las capacidades lógicamente omniscientes de un agente cognitivo. De acuerdo a nosotros el cuarto valor de verdad (Ambos=ambos, verdadero y falso) no es necesario para representar estados de creencia inconsistentes. Proponemos un análisis alternativo por sobredeterminación, usando fusión de mundos clásicos... En la configuración presentada en este artículo un agente puede confundir sus mundos accesibles. (Jaspars, 1993, p. 130)

En este enfoque los modelos son estructuras de la forma:

$$M = \langle S, \pi, R_1, ...R_m \rangle$$

donde, como es usual, S es un conjunto de estados (o mundos) y π una valuación de letras proposicionales en cada mundo. Esta vez, sin embargo, para cada i, $R_i \subseteq S \times \mathcal{P}(\mathcal{P}(\mathcal{S}) - \emptyset)$. Es decir, a algunos estados (posiblemente no a todos) R_i asocia un conjunto de subconjuntos no vacíos de S. Si sR_iT diremos que T es un s-cúmulo o un s-bloque (cuando la referencia a i no importe). Ahora la cláusula decisiva en la definición de verdad es:

$M, s \models \mathcal{B}_i\alpha$ si y solo si $\forall T \subseteq S$ si $R_i(s, T)$ entonces $\exists t \in T$ tal que $M \models_t \alpha$.

Es decir, que el agente creerá α en el estado s si en cada s-bloque hay un mundo en que vale α. Adviértase que si s no está en el dominio de R_i toda fórmula $B_i(\alpha)$ será verdadera en s.

Evaluación

Con esto en mente, evaluemos conforme a esta lógica la verdad de los siguientes enunciados:

1. Si $\models \alpha$ entonces $\models B_i\alpha$.
2. $\models \mathcal{B}_i(\alpha \to \beta) \to (\mathcal{B}_i\alpha \to \mathcal{B}_i\beta)$.
3. Si $\models (\alpha \to \beta)$ entonces $\models (\mathcal{B}_i\alpha \to \mathcal{B}_i\beta)$.
4. $\models (\mathcal{B}_i\alpha \wedge \mathcal{B}_i\beta) \to \mathcal{B}_i(\alpha \wedge \beta)$.
5. $\models \mathcal{B}_i\alpha \to \mathcal{B}_i(\alpha \vee \beta)$.
6. $\models \mathcal{B}_i(\beta \wedge \neg\beta) \to \mathcal{B}_i(\alpha)$.
7. $\models \neg\mathcal{B}_i(\alpha \wedge \neg\alpha)$.

Claramente 1) es verdadero pues en cada cada s-bloque habrá un mundo en el que vale α.

2) Es falso pues en cada s-bloque puede haber un mundo en que α sea verdadero y otro en que $\alpha \to \beta$ sea verdadero, mientras que β puede ser falso en todos los mundos de un s-bloque.

3) Es verdadera pues si en cada s-bloque hay un mundo en el que α es verdadero, en ese mundo β será verdadero (pues $(\alpha \to \beta)$ es verdadero).

4) Es falsa. Puede ser que α sea verdadera en algún mundo de cada bloque y β sea verdadera en otro mundo de cada bloque, mientras que $\alpha \wedge \beta$ sea falsa en todo un bloque.

5) Es verdadera. Es un caso particular de 3).

6) Es verdadera. También es una ejemplificación de 3). El antecedente siempre es falso en s, excepto que s no esté relacionado a ningún bloque por la relación R_i, en cuyo caso tanto el antecedente como el consecuente son verdaderos.

7) Es falsa. Claramente $(\alpha \wedge \neg\alpha)$ es falsa en un bloque completo. Por lo tanto $B_i(\alpha \wedge \neg\alpha)$ es falsa en cada s que esté en la relación R_i con algún bloque. Sin embargo, $\mathcal{B}_i(\alpha \wedge \neg\alpha)$ será falsa en un mundo que no esté relacionado a través de R_i a ningún bloque. Ese sería un caso en el que el individuo creería todo.

El agente puede creer P y también no-P sin que eso genere una obligación de (o razón para) creer todas las proposiciones. En cambio la creencia en $(P \wedge \neg P)$ obligaría al sujeto a creer todo. Este es un rasgo interesante de este sistema. El individuo puede creer contradicciones, la "lógica" no se lo prohibe. Dicho en otras palabras (6) es verdadero, pero no porque el antecedente sea contradictorio. Note que $(\mathcal{B}_i\alpha \wedge \mathcal{B}_i\neg\alpha) \rightarrow \mathcal{B}_i(\alpha \wedge \neg\alpha)$ no es válida.

Obtenemos prácticamente los mismos resultados que en el sistema anterior. De nuevo, no conseguimos evitar las formas más elementales de la omnisciencia lógica. El individuo cree (o debe creer) todas las verdades de la lógica y todas las consecuencias lógicas de sus creencias. En cambio, ninguna obligación le imponen sus creencias lógicas, excepto que sean verdaderas). Es decir, sus creencias no son (o no deben ser) cerradas bajo las implicaciones en que cree. El sistema no toma en cuenta las limitaciones cognitivas del agente ni la objeción de trivialidades. Sin embargo, se basa en un diagnóstico distinto, aunque menos claro, del origen de la contradicción de creencias.

3.11. Modelos de Rantala

Este es un enfoque que evita todos los problemas de omnisciencia lógica. Rantala (1982) define un nuevo tipo de modelos de Kripke, en la cual hay dos tipos de mundos. El primero lo forman los mundos posibles que se comportan de manera estándar. Las proposiciones que son verdaderas en cada uno de ellos forman conjuntos máximamente consistentes. En el segundo tipo hay mundos "imposibles" en los que la evalución de las fórmulas es completamente arbitraria, es decir, la verdad de una fórmula en ellos no se define recursivamente a partir del valor de verdad de sus subfórmulas. La idea es que, aunque esos mundos son metafísica o lógicamente imposibles, constituyen alternativas epistémicas en la mente de los agentes. Más específicamente:

Definición 51. Un modelo de Rantala es una estructura $\langle W, W^i, \pi, \sigma, R, \rangle$, tal que:

a) W es un conjunto no vacío.

b) W^i es un conjunto posiblemente vacío de S, cuyos elementos son llamados mundos imposibles. $W \cap W^i = \emptyset$.

c) π es una función que a cada elemento de W y letra proposicional asocia un valor de verdad; y que a cada elemento de W^i y fórmula del lenguaje asocia un valor de verdad.

d) $R \subseteq (W \cup W^i) \times (W \cup W^i)$ es una relación (de accesibilidad).

Llamaremos “mundos posibles” a los elementos de W. Dado un modelo M y un mundo posible w de M una fórmula es evaluada de acuerdo a la definición recursiva usual. Es decir, si $w \in W$,

- $w \models \rho$ si y solo si $\pi(w, \rho) = 1$ donde ρ es una letra proposicional.
- $w \models \neg\alpha$ si y solo si $w \not\models \alpha$
- $w \models (\alpha \wedge \beta)$ si y solo si $w \models \alpha$ y $w \models \beta$ En cambio, si $w \in W^i$ y α no comienza con el operador $\mathcal{B}$:
- $w \models \alpha$ si y solo si $\pi(w, \alpha) = 1$ Por último, si $w \in W$,
- $w \models \mathcal{B}_i\alpha$ si y solo si para todo $v \in (W \cup W^i)$ tal que wRv, $v \models \alpha$.

Definición 52. Una fórmula α es válida ($\models \alpha$ o $\models_{RAN} \alpha$) si sólo si es verdadera en cada mundo posible de cada modelo (de Rantala).

De nuevo la validez es evaluada solo en mundos lo que resulta en que la lógica del sistema es clásica. En cambio, podemos atribuir a un agente la lógica que queramos.

Evaluación

Es obvio que con esta lógica se evita toda forma de omnisciencia lógica. Todas las afirmaciones siguientes son falsas.
1) Si $\models \alpha$ entonces $\models \mathcal{B}_i\alpha$
2) $\models \mathcal{B}_i(\alpha \to \beta) \to (\mathcal{B}_i\alpha \to \mathcal{B}_i\beta)$
3) Si $\models (\alpha \to \beta)$ entonces $\models (B_i\alpha \to B_i\beta)$
4) $\models (\mathcal{B}_i\alpha \wedge \mathcal{B}_i\beta) \to \mathcal{B}_i(\alpha \wedge \beta)$
5) $\models \mathcal{B}_i\alpha \to \mathcal{B}_i(\alpha \vee \beta)$
6) $\models \mathcal{B}_i(\beta \wedge \neg\beta) \to \mathcal{B}_i(\alpha)$
7) $\models \neg\mathcal{B}_i(\alpha \wedge \neg\alpha)$
pues si w tiene acceso a un mundo imposible v, $\mathcal{B}_i\alpha$ puede ser falsa o verdadera, independientemente del valor que en v tengan las subfórmulas de α.

Aquí retomo la crítica que hice a las lógicas de los principios. La idea central de la propuesta de Rantala es que los mundos que el agente concibe como posibles no tienen por qué estar sometidos a las restricciones racionales o lógicas del sistema y que, por tanto, debemos permitir todas las variaciones posibles. De nuevo las creencias pueden formar un conjunto deforme, sin ninguna coherencia ni estructura interna. Asimismo la lógica deja el problema de la omnisciencia en manos del usuario y nada nos dice sobre cómo el agente forma o modifica sus creencias. La ventaja del sistema es su gran flexibilidad. De hecho, con algunas modificaciones puede subsumir algunas de las propuestas anteriores, pero eso no le brinda ninguna ventaja. No es difícil imaginar cómo, por ejemplo, la lógica de la conciencia general podría ser presentada como un caso particular del sistema de Rantala. Todo el peso recaerá sobre las restricciones que impongamos a las evaluaciones de fórmulas en mundos imposibles. Es decir, ahí caben todas las decisiones. Por sí solo, el sistema no impone ninguna obligación al agente. La lógica no juega su papel normativo. Considerado desde el punto de vista de la atribución de creencias en que la lógica juega un papel constitutivo, el conjunto de creencias del sujeto debe tener una estructura lógica mínima, como lo señalaron Davidson, Cherniak y Dennett.

3.12. Las Lógicas de la Justificación

En la literatura reciente se ha ofrecido como solución al problema de la omnisciencia lógica la llamada "lógica de la justificación". Bajo este rubro se agrupan varios sistemas lógicos todos ellos gobernados por la idea central de representar no sólo proposiciones, sino también sus justificaciones. Una de las maneras de motivar la introducción de la lógica de la justificación consiste en advertir que de los conocidos componentes de la definición clásica del conocimiento, a saber, creencia, verdad y justificación, sólo los dos primeros han sido considerados en lógica epistémica. Por ejemplo, desde la creación de la lógica epistémica por Hintikka ha sido usual incluir el axioma $(\Box p \rightarrow p)$ si el símbolo $\Box$ representa el conocimiento de algún agente; y excluirlo si $\Box$ solo representa su creencia. Esto da cuenta del segundo de los componentes enumerados del conocimiento. Sin embargo la justificación ha estado ausente en las formalizaciones. No debe pensarse que la lógica de la justificación tiene como única motivación apuntalar la lógica epistémica. La justificación misma es su primer tema de estudio, aunque puede ser empleada para construir una lógica de la creencia justificada o del conocimiento. En consonancia con lo que hasta ahora he desarrollado, yo consideraré la justificación como ligada a la creencia, aunque casos hay en que la justificación parece suficiente para garantizar la verdad de la creencia correspondiente y, quizás, para convertir a esta en conocimiento. La idea central de la lógica de la justificación es que, además de las fórmulas de la lógica clásica, tengamos un nuevo tipo de fórmulas divididas sintácticamente en dos partes por el símbolo " : ". En la parte izquierda aparecerán lo que llamaremos polinomios de prueba, que simbolizarán justificaciones estructuradas, mientras que a la derecha habrá, en una primera etapa, fórmulas de la lógica clásica. Conforme se reitera el proceso de construcción, después del símbolo ":" podrán figurar fórmulas formadas que contengan, a su vez, el símbolo ":"). La fórmula resultante representará el hecho de que el polinomio de prueba que aparece a la izquierda del principal símbolo ":" es una justificación de la fórmula subsecuente o bien, en otra posible lectura, que el polinomio es considerado por el agente como una justificación de dicha fórmula. Incluso se ha sugerido que el polinomio podría representar alguna condición no justificacional sino externista. Así como las fórmulas ordinarias se componen de átomos y operaciones proposicionales, los polinomios de prueba estarán constituidos de átomos (justificaciones no analizadas) engarzados por operadores propios a las justificaciones. De estos, dos aparecerán en todos los sistemas. El primero, llamado "aplicación", es una especie de concatenación ligada a la regla del *modus ponens*. Por ejemplo si r y s son justificaciones de las fórmulas de la forma α y $\alpha \to \beta$ respectivamente $(r \cdot s)$ será una justificación de β. Equivale a lo que en un sistema formal con *modus ponens* sería una prueba del siguiente estilo:

$\vdots$

n $(\alpha \to \beta)$ por r

$\vdots$

m α por s

m+1 β por *modus ponens* de n y m (por $r \cdot s$)

Así '$r \cdot s$' representa una concatenación del la prueba r y de la prueba s más una aplicación del *modus ponens*. Artemov (2008, p. 3) lo introduce así: "La operación aplicación realiza una acción epistémica, un paso deductivo de acuerdo con la regla del modus ponens".

La otra operación, denominada "suma" y que se representa ordinariamente con '+', equivale a una reunión de dos justificaciones. Cuando aparece en la lógica de las pruebas (pariente cercana de la lógica de la justificación), Artemov dice que $s + t$ puede interpretarse como la concatenación de las pruebas s y t (Artemov, 2008, p. 482). En relación a la justificación habla de "la evidencia combinada $s + t$" (Ibid.). La regla que gobernará este operador representará la monotonicidad de la justificación, es decir, que si a una justificación de una proposición P anexamos otra justificación cualquiera, obtendremos algo que seguirá siendo una justificación de P. Aunque algunos axiomas que rigen estos operadores aparecían ya en la lógica de las pruebas, Artemov los ha motivado en (2008) atendiendo a ejemplos y argumentos clásicos de epistemología. Así el axioma (de aplicación):

$$s : (F \rightarrow G) \rightarrow (t : F \rightarrow (s \cdot t)G) \qquad (*)$$

que representa un especie de principio de cerradura para la justificación puede ser motivada por el hecho de que en argumentos clásicos, como el de Gettier, es necesaria. Recordemos que Smith está justificado en su creencia de que Jones obtendrá el empleo, como también lo está en su creencia de que Jones tiene 10 monedas en el bolsillo. De allí concluye que el hombre que obtendrá el empleo (que en realidad es él) tiene 10 monedas en el bolsillo, lo que es verdadero. Al obtener esta conclusión a través de la deducción a partir de dos premisas justificadas, podemos decir que esta conclusión también está justificada y es aquí que aparece el principio de cerradura (aplicado a la justificación) representado por $(*)$. Sin embargo, esta "cerradura" tiene características inusuales porque no tiene la forma de un clásico axioma de distribución de un operador modal sobre un condicional.

De la misma forma, los axiomas (llamados de suma)

$$s : F \rightarrow (s + t) : F$$

$$t : F \rightarrow (s + t) : F$$

pueden ser motivados por la condición que Lehrer y Paxon (1969) agregaron a las tres clásicas en la definición de conocimiento, a saber:

> No hay una verdad ulterior la cual, si el sujeto la hubiera conocido, habría anulado (*defeated*) la justificación actual para la creencia.

Artemov sabe que la cláusula anterior se refiere más bien a una operación de revisión (o actualización) de creencias por la adquisición de nueva información, pero su enfoque no incluye operaciones tan complejas y, por ello, se limita a esta lectura en términos de polinomios de prueba que no generan incompatibilidad. Dos posibles axiomas provienen o pueden ser motivados por las condiciones suplementarias que el fiabilismo de Goldman impone al conocimiento a saber:

> Una creencia de un agente está justificada sólo si la verdad de esa creencia ha causado que el sujeto la tenga (en el modo apropiado); y para que una creencia verdadera y justificada cuente como conocimiento, el agente debe ser capaz de reconstruir correctamente (en su mente) la cadena causal (1967).

Estas dos condiciones pueden representarse respectivamente por el axioma:

$$t : F \rightarrow F$$

y la regla:

si $\vdash F$ entonces uno puede construir una justificación p tal que $\vdash p : F$

Al primero la llama "Factualidad"; a la segunda, "Principio de Internalización". Por supuesto que, respecto a la segunda, la cláusula "uno puede construir una justificación", se traduce en "hay una justificación" y, en última instancia, en esa regla de inferencia. Artemov insiste en que tal formalización no recoje sino parcialmente las condiciones que los mencionados autores agregan a la definición clásica del conocimiento. En cualquier caso, no es conveniente que estas dos condiciones aparezcan, una como axioma y otra como regla, en todos los sistemas. Es interesante estudiar sistemas más generales en que la justificación no es factual, es decir, no necesariamente está ligada a creencias verdaderas. El principio de internalización aparece en muchos sistemas como regla derivada.

Empezaré por exponer los sistemas más sencillos de la lógica de la justificación, lo que me permitirá mostrar uno de los dos sentidos en que puede considerarse que estas lógicas afrontan el problema de la omnisciencia lógica. más adelante mostraré cómo uno de estos sistemas puede combinarse con otro de lógica doxástica y las consecuencias que ello tiene.

El lenguaje de las lógicas de la justificación más sencillas tiene el siguiente vocabulario: variables proposicionales: $p, q, r \ldots$, símbolos para los conectivos lógicos usuales ($\neg$, $\vee$, $\wedge$, $\rightarrow$, $\leftrightarrow$), variables de justificación: $x, y, z, \ldots$, constantes de justificación: $a, b, c, \ldots$, símbolos para las operaciones sobre justificaciones: aplicación ($\cdot$) y suma ($+$). Las reglas de formación de fórmulas son las siguientes:

Definición 53. Son polinomios de prueba:

1. Las variables y las constantes de justificación,

2. $(\rho \cdot \sigma)$ y $(\rho + \sigma)$, si ρ y σ son polinomios de prueba.

Definición 54. Son fórmulas:

1. Las letras proposicionales,

2. Las expresiones $\tau : \alpha$ donde τ es un polinomio de prueba y α es una fórmula.

3. Las expresiones $\neg\alpha$, $(\alpha \wedge \beta)$, $(\alpha \to \beta)$, $(\alpha \vee \beta)$ y $(\alpha \leftrightarrow \beta)$, si α y β son fórmulas.

La lógica de la justificación básica, J_0, tiene como axiomas:

1. Todas las instancias de tautologías clásicas,

2. Las fórmulas de la forma $\rho : (\alpha \to \beta) \to (\sigma : \alpha \to (\rho \cdot \sigma) : \beta)$ (axiomas de aplicación).

3. Las fórmulas de la forma $\rho : \alpha \to (\rho + \sigma) : \alpha$; así como las de la forma $\rho : \alpha \to (\sigma + \rho) : \alpha$ (axiomas de suma).

Su única regla de inferencia es el *modus ponens*.

J_0 no tiene teoremas de la forma $\rho : \alpha$, aunque permite el establecimiento de ciertas relaciones deductivas. Para obtener un sistemas más interesante necesitamos agregar como axiomas las que Artemov llama "especificaciones constantes", lo que corresponde a la idea de que "un agente acepta los axiomas lógicos (incluidos los relativos a justificaciones) como justificados" (Artemov 2008, p.483). Así por cada axioma α podemos postular que el agente tiene una justificación e_1, es decir, $e_1 : \alpha$ y esto, a su vez, puede estar justificado por e_2, es decir, $e_2 : (e_1 : \alpha)$ (lo que podemos abreviar por $e_2 : e_1 : \alpha$). Más precisamente tenemos la siguiente definición:

Definición 55. Una especificación constante CS para una lógica es un conjunto de fórmulas de la forma:

$$e_n : e_{n-1} : \ldots : e_1 : \alpha$$

donde α es un axioma y tal que si $e_i : e_{i-1} : \ldots : e_1 : \alpha \in CS$ entonces $e_{i-1} : \ldots : e_1 : \alpha \in CS$.

Las especificaciones constantes pueden ser de varios tipos. Especialmente relevantes son:

1. Las especificaciones axiomáticamente apropiadas, es decir, aquellas que, por cada axioma α, hay una justificación constante e_1 tal que $e_1 : \alpha$ es miembro de las especificaciones constantes, y tal que si $e_{i-1} : \ldots : e_1 : \alpha$ es miembro de las especificaciones constantes, existe e_i tal que $e_i : e_{i-1} : \ldots : e_1 : \alpha$ también es miembro de este conjunto.

2. La especificación total, CST que contiene por cada axioma α y cualesquiera justificaciones $e_1, e_2, \ldots e_n$, la fórmula $e_n : e_{n-1} : \ldots : e_1 : \alpha$.

Introducimos ahora dos sistemas:

Dado un conjunto CS de especificaciones constantes, sea J_{CS} el sistema que se obtiene de J_0 agregando como axiomas las fórmulas de CS. Por otro lado, el sistema J es J_0 con la regla de internalización, a saber:

Si α es un axioma y $e_i, e_{i-1} \ldots, e_1$ son constantes de justificación, entonces puede inferirse

$$e_i : e_{i-1} : \ldots : e_1 : \alpha$$

Equivalentemente $J = J_{CST}$

El siguiente es un ejemplo de cómo funciona el sistema J

Teorema 56. *Si β es consecuencia tautológica de $\{\alpha_1, \alpha_2, \ldots \alpha_n\}$ entonces $\vdash_J$ $(x_1 : \alpha_1 \wedge x_2 : \alpha_2 \wedge \ldots \wedge x_n : \alpha_n) \rightarrow (e \cdot x_1 \cdot x_2 \ldots \cdot x_n) : \beta$ para cualquier constante de justificación e.*

Demostración. Por hipótesis $(\alpha_1 \rightarrow (\alpha_2 \rightarrow \ldots (\alpha_n \rightarrow \beta)))$ es una tautología y, por tanto, un axioma de J. Por la regla de internalización podemos derivar $e : (\alpha_1 \rightarrow (\alpha_2 \rightarrow \ldots (\alpha_n \rightarrow \beta)))$. Utilizando el axioma de aplicación y modus ponens, derivamos $(x_1 : \alpha_1 \rightarrow (e \cdot x_1) : (\alpha_2 \rightarrow \ldots (\alpha_n \rightarrow \beta)))$. Después de n pasos llegamos al teorema. $\square$

Teorema 57. *Si $\vdash_J \alpha$ entonces existe un polinomio de justificación τ tal que $\vdash_J \tau : \alpha$.*

Demostración. Si $\vdash_J \alpha$ entonces hay una prueba para α. Para cada paso de la prueba, demostraremos el teorema. Si el n-ésimo paso es un axioma, el teorema debe valer por la regla de internalización (es decir, $\vdash_J e : \alpha$, si α es un axioma). Si no, proviene de pasos anteriores por modus ponens o por internalización. En el primer caso, si viene de β y $\beta \rightarrow \alpha$ por modus ponens, entoces, por hipótesis de inducción, hay polinomios de prueba ρ y σ tales que $\vdash_J \rho : \beta \rightarrow \alpha$ y $\vdash_J \sigma : \beta$. Ahora aplicamos modus ponens al axioma $\vdash_J (\rho : (\beta \rightarrow \alpha)) \rightarrow (\sigma : \beta \rightarrow ((\rho \cdot \sigma) : \alpha)$, con lo que obtenemos el resultado. Si el nésimo paso de la prueba es $c : \alpha$ y se obtuvo de α por internalización, entonces, de nuevo por internalización podemos derivar $d : c : \alpha$ para alguna constante de justificación d. $\square$

Note que el teorema extiende la regla de internalización a todo teorema.

Antes de exponer otros sistemas que son extensiones de los ya mencionados, resumiré la semántica que suele darse para J_0.

Definición 58. Un J_0-modelo es una cuaterna $M = \langle W, R, V, \xi\rangle$ tal que $\langle W, R, V\rangle$ es un modelo de Kripke y ξ una función que a cada par (r, α), donde r es un polinomio de prueba y α una fórmula, asocia un subconjunto de W, y tal que cumple además las siguientes dos condiciones:
1) $\xi(r, \alpha \to \beta) \cap \xi(s, \alpha) \subseteq \xi(r \cdot s, \beta)$
2) $\xi(r, \alpha) \cup \xi(s, \alpha) \subseteq \xi(r + s, \alpha)$

$\xi(r, \alpha)$ es el conjunto de mundos en que r es una evidencia admisible para α y las condiciones 1) y 2) asegurarán que los axiomas de aplicación y suma sean válidos.

La definición de verdad en un mundo (de un modelo) para un enunciado de J_0 cuyo operador lógico principal es un conectivo clásico es la usual. A ella agregamos la cláusula:

$$u \models r : \alpha \text{ si y solo si } u \in \xi(r, \alpha) \text{ y } \forall v(uRv \to v \models \alpha)$$

Esta cláusula se asemeja a la que la lógica de la consciencia general establece para enunciados de creencia explícita, pero ahora tenemos un polinomio de prueba estructurado.

Por supuesto que con el fin de obtener un teorema de completud para J_{CS} (dado un conjunto CS específico) es necesario agregar una condición extra a la semántica.

Definición 59. Dado un conjunto CS de especificaciones constantes, un J_0-modelo $M = \langle W, R, V, \xi\rangle$, y $w \in W$, diremos que w respeta CS si para cada $r : \alpha \in CS$, $w \in \xi(r, \alpha)$; y que M respeta CS si para cada $w \in W$, w respeta CS.

Teorema 60. *Dado un conjunto CS, $\vdash_{JCS} \alpha$ si y solo si α es verdadero en todo mundo de todo modelo que respete CS.*

Es posible dar otro tipo de semántica más sencilla para las lógicas de la justificación, pero Artemov reconoce que la anterior es la más intuitiva. Por ejemplo, si $j : \alpha$ se interpreta como "el sujeto cree α por la razón j", entonces esta semántica recoge la idea de que un agente en una situación cree una proposición α por una justificación dada, si α es verdadera en todas las variantes que concibe y dicha justificación es adecuada a α en tal situación. Como dije antes, a pesar de las semejanzas con la lógica de la consciencia general, una estructura extra liga las condiciones de verdad de fórmulas en un mundo, además de las asociadas a los conectivos lógicos. Por ejemplo, en el caso de una semántica CS axiomáticamente adecuada, $w \models r : \alpha$ si y solo si existe un término de justificación s tal que $w \models s : r : \alpha$.

Otro sistema puede obtenerse agregando a J el ya mencionado (esquema de) axioma de factualidad, a saber:

$$(t : \alpha) \rightarrow \alpha \qquad\qquad A4$$

y podemos formar los sistema J_0T y JT agregando $A4$ a J_0 y J respectivamente. En ambos casos, para obtener teoremas de solidez y completud debemos agregar a las semánticas respectivas la condición de que la relación R sea reflexiva. Desde luego, este axioma no resulta apropiado para representar la creencia. La semejanza que aquí comienza a aparecer con los sistemas clásicos de lógica modal no es casual. ¿Qué sucede con las introspecciones positiva y negativa y el axioma D?

Respecto al axioma D y la propiedad de no tener creencias (justificadamente) contradictorias, sabemos que $\vdash_{J_0} (r : \alpha \wedge s : \neg\alpha) \rightarrow (e \cdot r \cdot s :\perp)$, para alguna constante e. Por lo tanto, para tener esta propiedad basta con agregar como axioma el esquema $\neg t :\perp$ (donde t representa un polinomio de justificación) y la condición de que los modelos sean seriales.

Para la introspección positiva, Artemov agrega al lenguaje un nuevo operador unario de justificaciones, $!t$, que introduce de la siguiente manera:

> Este principio [$KF \rightarrow KKF$] tiene una adecuada contraparte explícita: el hecho de que el agente acepta t como una evidencia suficiente para F sirve como una evidencia suficiente para $t : F$. Frecuentemente tal metaevidencia tiene una forma física, por ejemplo, el reporte de un árbitro verificando que una prueba de un artículo es correcta... (Artemov, 2008, p. 494).

El axioma correspondiente es:

$$t : \alpha \rightarrow !t : (t : \alpha)$$

De nuevo, es posible demostrar corrección y completud de los sistemas que se obtienen de los ya mencionados agregando, por un lado, este esquema axiomático y, por otro, la condición de que la relación de accesibilidad sea transitiva. La introspección negativa, en cambio, plantea una dificultad mayor, como el propio Artemov reconoce, y no la voy a considerar por no ser esencial a mi tema.

Habría varias maneras de interpretar los sistemas anteriores. La más natural sería de corte externalista. Es decir, las justificaciones de un enunciado son objetivas, las reconozca o no el sujeto. Sin embargo, es sencillo integrar la lógica de la justificación a una lógica doxástica. En lo que sigue mostraré cómo puede construirse un sistema, que llamaré $JKD4$, de justificación doxástica.

El lenguaje de $JKD4$ contiene variables x, y, z... y constantes a, b, c... de justificación, símbolos para operaciones sobre justificaciones $\cdot$, $+$ y !, letras proposicionales, los conectivos lógicos usuales y varios operadores de creencia $\mathcal{B}_1, \mathcal{B}_2, \ldots \mathcal{B}_m$. Para simplificar la exposición, supondré que $m = 1$. Los polinomios de justificación se forman como ya fue dicho. Las fórmulas se construyen de acuerdo a la siguiente definición:

Definición 61. Son fórmulas:

1. Las letras proposicionales,
2. Las expresiones $\tau : \alpha$ donde τ es un polinomio de prueba y α es una fórmula.
3. Si α y β son fórmulas, también lo son: $\neg\alpha$, $(\alpha \wedge \beta)$, $(\alpha \rightarrow \beta)$, $(\alpha \vee \beta)$, $(\alpha \leftrightarrow \beta)$ y $\mathcal{B}\alpha$.
4. Si $\tau : \alpha$ es una fórmula, también lo es $!\tau : (\tau : \alpha)$

Definición 62. Si α, β y γ, son fórmulas y s, $s_1, s_2 \ldots s_n$ y t, polinomios de prueba, son axiomas:

CP1 Las instancias de tautologías.

J1 $s : (\alpha \rightarrow \beta) \rightarrow (t : \alpha \rightarrow (s \cdot t) : \beta)$.

J2 $(s : \alpha \vee t : \alpha) \rightarrow (s + t) : \alpha)$.

J3 $t : \alpha \rightarrow !t : (t : \alpha)$.

J4 $\neg t : \perp$.

K1 $\mathcal{B}(\alpha \rightarrow \beta) \rightarrow (\mathcal{B}\alpha \rightarrow \mathcal{B}\beta)$.

K2 $(\mathcal{B}\alpha \rightarrow \mathcal{B}\mathcal{B}\alpha)$.

K3 $\neg\mathcal{B}\perp$.

KJ $(t : \alpha \rightarrow \mathcal{B}\alpha)$. Y son reglas de inferencia:

R1 Modus ponens.

R2 De $\vdash s_{n-1} : \ldots s_1 : \alpha$ donde α es un axioma, podemos deducir:

$\vdash s_n : s_{n-1} : \ldots : s_1 : \alpha$.

R3 de $\vdash \alpha$ podemos deducir $\vdash \mathcal{B}\alpha$.

Lema 63. $\vdash t : \alpha \rightarrow \mathcal{B}t : \alpha$

Demostración. La siguiente es una derivación en $JKD4$:

1. $t : \alpha \rightarrow !t : (t : \alpha)$
2. $!t : (t : \alpha) \rightarrow \mathcal{B}(t : \alpha)$.
3. $t : \alpha \rightarrow \mathcal{B}(t : \alpha)$

□

Lema 64. *Para cualquier fórmula α, hay un polinomio de prueba $Up_\alpha(x)$ tal que $\vdash x : \alpha \rightarrow Up_\alpha(x) : B\alpha$*

Demostración. La siguiente es una derivación en $JKD4$:

1. $(x : \alpha \rightarrow \mathcal{B}\alpha)$ por KJ.
2. $a : (x : \alpha \rightarrow \mathcal{B}\alpha)$ por R2.
3. $(!x : x : \alpha \rightarrow (a \cdot !x) : B\alpha)$ por aplicación y *modus ponens*.
4. $x : \alpha \rightarrow !x : x : \alpha)$ por J3.
5. $x : \alpha \rightarrow (a \cdot !x) : \mathcal{B}\alpha)$ por transitividad.

Por lo tanto, $Up_\alpha(x) = (a \cdot !x)$. □

Teorema 65. *Si $\vdash_{JKD4} \alpha$ entonces hay un polinomio de prueba s tal que $\vdash_{JKD4} s : \alpha$*

Demostración. Sea $PR_1, PR_2, \ldots PR_n$ una prueba en $JKD4$ donde $PR_n = \alpha$, vamos a probar que para cada i $(1 \leq i \leq n)$ hay un polinomio de prueba s_i tal que $\vdash_{JK4} s_i : PR_i$
Caso 1. PR_1 es un axioma. Por la regla $R2$, $\vdash_{JKD4} a : PR_1$
Caso 2 De PR_m y $PR_K = PR_m \to \beta$ concluimos β. Por hipótesis de inducción $\vdash_{JKD4} s_m : PR_m$ y $\vdash_{JKD4} s_k : PR_m \to \beta$. Por lo tanto, son teoremas de $JKD4$:
$(s_k : PR_m \to \beta) \to (s_m : PR_m \to (s_k \cdot s_m) : \beta))$ axioma J1
$(s_k \cdot s_m) : \beta$ por dos aplicaciones de R1
Caso 3. Si $s_n : \ldots s_1 : \alpha$ fue derivado de $s_{n-1} : \ldots s_1 : \alpha$ por la regla $R2$, entonces, aplicando de nuevo $R2$, tenemos que $\vdash_{JKD4} s_{n+1} : s_n : \ldots s_1 : \alpha$
Caso 4. Si $\mathcal{B}\alpha$ fue obtenido de α por R3, entonces por hipótesis de inducción existe s tal que $\vdash_{JKD4} s : \alpha$. Por lema anterior, hay un polinomio de prueba p tal que $\vdash_{JKD4} p : \mathcal{B}\alpha$ $\square$

Veremos ahora una semántica para el sistema $KJD4$

Definición 66. Un marco de evidencia es una estructura $FR = \langle W, R, R^e \rangle$ tal que

$$W \neq \emptyset$$

$$R \subseteq W \times W$$

$$R^e \subseteq W \times W$$

$$R \text{ y } R^e \text{ son transitivas y seriales, y } R \subseteq R^e$$

Definición 67. Dado un marco de evidencia $FR = \langle W, R, R^e \rangle$ una función de evidencia posible ϵ para FR es una función de mundos y términos de justificación a conjuntos de fórmulas tal que se cumplen las siguientes condiciones siguientes:

1. Monotonicidad. Si uR^ev implica $\epsilon(u,t) \subseteq \epsilon(v,t)$
2. Aplicación. $(\alpha \to \beta) \in \epsilon(w,s)$ y $\alpha \in \epsilon(w,t)$ entonces $\beta \in \epsilon(w, s \cdot t)$
3. Suma $\epsilon(u,s) \cup \epsilon(u,t) \subseteq \epsilon(u, s+t)$
4. Inspección si $\alpha \in \sigma(u,t)$ entonces $t : \alpha \in \sigma(u, !t)$
5. Si $\alpha \in \epsilon(u,t)$ entonces para cualquier constante de justificación s, $t : \alpha \in \epsilon(u,s)$.

Como antes, que $\alpha \in \epsilon(w,t)$ (con w un mundo posible del marco y t un polinomio de justificación) significa que t es una buena justificación para α en w.

Definición 68. Un modelo de evidencia es una estructura $M = \langle W, R, R^e, V, \epsilon\rangle$ donde $FR = \langle W, R, R^e\rangle$ es un marco, V una valuación de las letras proposicionales en cada mundo (o, equivalentemente V es una función de LP en $\mathcal{P}(\mathcal{W})$). y ϵ es una función de evidencia posible para FR.

Definición 69. Dado un modelo $M = \langle W, R, R^e, \epsilon, V\rangle$ y un mundo $w \in W$, definimos qué significa que una fórmula α sea verdadera en M, w (lo que denotaré $M, w \models \alpha$) por las siguientes cláusulas recursivas:
1) si $\rho \in LP$, $M, w \models \rho$ si y solo si $w \in V(\rho)$.
2) $M, w \models \neg\alpha$ si y solo si $M, w \not\models \alpha$
3) $M, w \models \alpha \wedge \beta$ si y solo si $M, w \models \alpha$ y $M, w \models \beta$.
4) $M, w \models \mathcal{B}\alpha$ si y solo si para todo $v \in W$ tal que wRv, se tiene que $v \models \alpha$.
5) $M, w \models t : \alpha$ si y solo si $\alpha \in \epsilon(w, t)$ y para todo $v \in W$ tal que $wR^e v$, se tiene que $v \models \alpha$.

Lema 70. *Si $u \models t : \alpha$ y $uR^e v$ entonces $v \models t : \alpha$*

Demostración. Hipótesis $u \models t : \alpha$, es decir, $\alpha \in \epsilon(u, t)$ y si $uR^e v$ entonces $v \models \alpha$. Sean v y z tales que $uR^e vR^e z$. PD: a) $\alpha \in \epsilon(v, t)$ y $z \models \alpha$.
a) es por monotonía. Dado que $uR^e v$ y $\alpha \in \epsilon(u, t)$ entonces $\alpha \in \epsilon(v, t)$.
b) Por transitividad. Dado que $uR^e vR^e z$ entonces $uR^e z$ y, como $u :\models t : \alpha$, $z \models \alpha$. □

Teorema 71. *Si $\vdash_{JKD4} \alpha$ entonces $M \models \alpha$ para todo modelo de evidencia.*

Demostración. Veamos que los axiomas son válidos (verdaderos en todos los mundos de todos los modelos de evidencia) y que las reglas de inferencia preservan esa propiedad. Esto es obviamente verdadero de todas las instancias de tautologías y del *modus ponens*.
Los axiomas K1 K2 y K3 (de distribución, iteración del operador doxástico B) y de consistencia son obviamente válidos porque R es transitiva y serial.
El axioma J1, $s : (\alpha \to \beta) \to (t : \alpha \to (s \cdot t) : \beta)$ es válido porque ϵ es una función de evidencia, es decir, satisface $(\alpha \to \beta) \in \epsilon(w, s)$ y $\alpha \in \epsilon(w, t)$ entonces $\beta \in \epsilon(w, s \cdot t)$
Similarmente J2 (el axioma de suma) es válido por las condiciones impuesta a ϵ
En cuanto a J3, $t : \alpha \to !t : (t : \alpha)$. Sea u tal que $u \models t : \alpha$. Debemos demostar que $u \models !t : t : \alpha$, es decir que si $uR^e v$ entonces $v \models t : \alpha$ y que si $\alpha \in \epsilon(u, t)$ entonces $t : \alpha \in \epsilon(u, !t)$. Lo primero ya fue probado y lo segundo es porque ϵ es una función de evidencia.
El axioma J4 ($\neg t : \perp$) es válido por la serialidad de R^e.

El axioma KJ, $(t : \alpha \rightarrow \mathcal{B}\alpha)$. Si $u \models t : \alpha$ entonces $\alpha \in \epsilon(u, t)$ y para todo v, si $uR^e v$, entonces $v \models \alpha$, pero dado que $R \subseteq R^e$, para todo z si uRz entonces $z \models \alpha$. Por lo tanto, $v \models \mathcal{B}\alpha$.
Es claro que las reglas de inferencia preservan validez. □

También es posible probar la completud del sistema con respecto a esa semántica.

Desde luego que pudimos combinar otra lógica de la justificación con otra lógica epistémica o doxástica.

Evaluación 1

Una vez que hemos visto las principales características de las lógicas de la jutificación, veamos qué nos ofrecen como solución al problema de la omnisciencia lógica bajo la perspectiva de este trabajo. Empecemos por un sistema de este tipo sin operadores doxásticos, del tipoJ_{CS} para algún conjunto dado de especificaciones constantes. La lectura más natural de las verdades de este sistema parecería ser en términos de principios puente de la modalidad $\mathcal{R}$ y, más específicamente, de la forma $\mathfrak{B}\mathcal{R}+$. Es decir, los términos de justificación serían interpretados como razones y no agregaríamos ningún otro operador doxástico. Sin embargo, algo se perdería si simplemente leemos un condicional del tipo $J : \alpha \rightarrow t : \beta$ como "si el agente tiene razones para α entonces tiene razones para β", porque precisamente no habríamos mantenido un registro de la razones que lo pueden conducir a esas creencias. Por otro lado, esas razones no tendrían por qué estar al alcance del sujeto ni referirse a él. Que un argumento genere una buena razón para creer una tesis es algo que no depende de ningún sujeto. Es mejor volver a la cuestión de la omnisciencia lógica en el caso recién visto en que la lógica de la justificación se combina con una lógica doxástica. Podríamos interpretar las fórmulas de la forma $j : \alpha$ y $\mathcal{B}\alpha$ como representando la creencia explícita y la implícita respectivamente. Invitan a esta lectura los teoremas $t : \alpha \rightarrow \mathcal{B}\alpha$, $t : \alpha \rightarrow \mathcal{B}t : \alpha$ y $\mathcal{B}(\alpha \rightarrow \beta) \rightarrow (\mathcal{B}\alpha \rightarrow \beta)$, entre otros. Tal es la interpretación que proponen Artemov y Nogina (2005). La cuestión de la omnisciencia lógica recae entonces sobre las fórmulas del tipo $t : \alpha$. ¿Cómo debemos interpretarlas? Una primera posibilidad nos concierne poco aquí. Se trata de leer $j : \alpha$ o bien como "j es una justificación para α" o como "j es una justificación para que el agente crea α", pues j podría en ambos casos estar más allá de las posibilidades del agente o ser completamente ignorada por él. Otra opción, mencionada por Artemov (2008), es "el agente acepta α por la justificación j". Parece una lectura más adecuada.

Para ejemplificar el caso, veamos el análisis que hace Artemov usando lógicas de la justificación de un argumento paradigmático en la literatura epistemológica y en que juega un papel importante un principio de cerradura que otorga al agente justificación para creer consecuencias lógicas de algunas de sus creencias justificadas. Es el ejemplo del granero rojo que Kripke diseñó como contraejemplo a la caracterización del conocimiento que dio Nozick. Supongamos que nuestro agente conduce su auto por una comarca en que hay graneros, unos reales y otros postizos. El individuo puede creer que tiene enfrente un granero cuando percibe uno. Sin embargo, no podemos decir que sabe que allí hay un granero, si sólo se basa en dicha percepción. Sin embargo, no hay ningún granero postizo de color rojo. Por ello, podemos decir que el agente cree que hay un granero rojo cuando ve uno y, según la definición de Nozick, también lo sabe. En este caso, la justificación que tiene es adecuada. El problema es que aparentemente de la proposición de que hay un granero rojo (que sabe) puede concluir lógicamente que allí hay un granero (lo que, en realidad, ignora). Veamos ahora el análisis de este caso que ofrece Artemov empleando la lógica de la justificación.

Representemos por G y $(G \wedge R)$ respectivamente las proposiciones "hay allí un granero" y "hay allí un granero rojo". Sean v y r las respectivas justificaciones que tiene el agente para creerlas. r es una justificación factual; la otra, no. Si interpretamos $j : \alpha$ como el agente cree α por la justificación j el análisis de la derivación que supuestamente lleva del conocimiento a la ignorancia es la siguiente:

$v : G$

$r : (G \wedge R)$

$(G \wedge R) \to G$

$e : ((G \wedge R) \to G)$ por internalización.

$(r : (G \wedge R)) \to (e \cdot r) : G$ por aplicación y *modus ponens*

$(e \cdot r) : G$

De aquí podríamos concluir que tiene una justificación adecuada $r \cdot e$ para sostener que hay allí un granero (y, por ello, tal vez podemos decir que lo sabe) y una justificación inadecuada v para aseverar que lo que ve es un granero.

Recuerde que consideramos a la lógica de la justificacíon combinada con una lógica epistémica, a B como un operador de conocimiento implícito y a $j : \alpha$ como afirmando que el agente sabe explícitamente α, tal y como lo hacen Artemov y Nogina en (2005). En esta interpretación $t : \alpha$ representa que el agente cree α en virtud de la razón t. Por supuesto que no sería una lectura correcta desde una perspectiva completamente descriptivista, pues de que el sujeto cree $\alpha \to \beta$ por la razón s y α por la justificación t no se sigue que el sujeto crea β en virtud de $s \cdot t$, pues bien pudiera ser que no creyera β o que creyera esta proposición por otras

razones. En cambio, esta segunda interpretación daría pie a una lectura normativa del tipo $Co+$ o $Bo+$. Por ejemplo, el axioma de aplicación podría ser leído de estas dos formas:
1) si el agente cree $(\alpha \rightarrow \beta)$ por la razón s y α por la razón t, entonces debería creer β por la razón $(s \cdot t)$.
2) si el agente debe creer $(\alpha \rightarrow \beta)$ por la razón s y α por la razón t, entonces debería creer β por la razón $(s \cdot t)$.
De nuevo, el principio 2) es demasiado poco normativo. Solo se aplica a creencias que el agente ya deba tener. Aquí vale la objeción de MacFarlane: la lógica solo sería normativa para creencias que ya son correctas en algún sentido. Sin embargo, esta interpretación puede servir en el caso de ciertas formas de conocimiento deductivo, como diré más adelante. Queda la opción 1, aunque tomando en cuenta la objeción del constreñimiento, podríamos leerla más como un principio Wo+.
En cambio, el axioma de suma parecería más apto para un principio puente negativo, por ejemplo:
Si el sujeto cree α en virtud de la razón s, no debería dejar de creer α por las razones $s + t$.
Esta lectura es sugerida por otro análisis del granero rojo propuesto por Artemov en el marco de una lógica epistémica de la justificación, en la que la fórmula $t : \alpha$ es interpretada como "el agente sabe α por la razón t". Yo la interpreto como "el agente debe creer α por la razón t". La derivación susodicha podría representarse así:
$\neg v : G$
$r : (G \wedge R)$
$(G \wedge R) \rightarrow G$
$e : ((G \wedge R) \rightarrow G)$ por internalización.
$(r : (G \wedge R)) \rightarrow (r \cdot e) : G$ por aplicación y *modus ponens*
$(r \cdot e) : G$
Ahora el primer renglón establece que el sujeto no debe creer G solo por v, mientras que el último declara que el individuo debe creer G por la razón $(r \cdot e)$.

Con alguna de esas interpretaciones en mente podrían leerse las fórmulas de la lista que hemos empleado reiteradamente y que en el tamiz de la lógica $JKD4$ toman las siguientes formas:
1) Si $\models \alpha$ entonces existe τ tal que $\models \tau : \alpha$.
2) $\models \rho : (\alpha \rightarrow \beta) \rightarrow (\sigma : \alpha \rightarrow (\rho \cdot \sigma) : \beta)$.
3) Si $\models (\alpha \rightarrow \beta)$ entonces existe ρ tal que $\models_J (\tau : \alpha \rightarrow (\rho \cdot \tau) : \beta)$.
4) Existe ϵ tal que $\models (\tau : \alpha \wedge \rho : \beta) \rightarrow (\epsilon \cdot \tau \cdot \rho) : (\alpha \wedge \beta)$.

5) Existe ρ tal que $\models \tau : \alpha \to (\rho \cdot \tau) : (\alpha \vee \beta)$.
6) Existe ρ tal que $\models \tau : (\beta \wedge \neg\beta) \to (\rho \cdot \tau) : (\alpha)$.
7) Existe ρ tal que $\models \rho : \neg(\alpha \wedge \neg\alpha)$,
en donde τ, ρ y σ son polinomios de justificación.

Todas son verdaderas en la semántica correspondiente a $JKD4$. El primero deberá leerse como: si α es válido, hay una justificación en virtud de la cual el agente debe creerla (explícitamente). Sin embargo, si solo decimos "hay una justificación"se pierde un elemento importante, pues en realidad para cada fórmula hay una justificación concreta, dada la cual podríamos considerar a 1 como un esquema que da lugar a principios puente del tipo: si α es válido el agente debería creerla en virtud de tal justificación concreta. El segundo caso ya fue tratado más arriba y algo similar puede decirse del tercero. Los últimos cuatro incisos son casos particulares del segundo y admiten la lectura normativa que he propuesto. El último diría que hay una justificación en virtud de la cual el individuo no deberá creer una contradicción. De hecho, sabemos que $\not\models \bot$, así es podemos decir que no hay ninguna justificación para que un individuo acepte una contradicción. Aparentemente no estamos tomando en cuenta las demandas excesivas, pero hay otro recurso de la lógica de la justificación para hacer frente a la omnisciencia lógica y que veremos en la siguiente sección.

Un criterio cuantitativo de omnisciencia lógica

Enseguida veremos cómo la lógica de la justificación pretende resolver el problema de la omnisciencia lógica a través de la medición de la complejidad computacional.

Este tema fue tratado por Artemov y Kuznets en 2006 y, de nuevo, en 2009. Afrontan el problema de la omnisciencia lógica desde la perspectivas de las demandas excesivas. El agente tiene capacidades limitadas y le toma tiempo, energía y espacio hacer inferencias. A este respecto señalan: "...un modelo adecuado del conocimiento debería reflejar este hecho en algún grado de generalidad; la teoría de la complejidad provee una plataforma razonable para tal enfoque" (2009, p. 14). Ellos proveen dos definiciones rigurosas de omnisciencia lógica que permiten separar con precisión a los sistemas de lógica epistémica que adolecen de este defecto. Esbozaré a grandes rasgos en qué consisten estos criterios.

Su exposición es bastante general. Consideremos una lógica **L**, con un lenguaje $\mathcal{L}$. Podemos identificar a **L** con el conjunto de enunciados válidos de $\mathcal{L}$ según

L.

Definición 72. Un sistema de prueba para $\mathcal{L}$ es una función sobreyectiva $Pr(\Sigma) \Rightarrow$ **L** calculable en tiempo polinomial, donde Σ es un conjunto de sucesiones finitas de expresiones de **L** llamadas pruebas.

Es decir, PR es una función que a cada prueba π le asocia la fórmula que π demuestra. Por supuesto, solo algunas sucesiones de expresiones de **L** pertenecen a Σ, y cada fórmula válida tiene al menos una prueba.

Sea μ una función que asocia a cada prueba una medida de su tamaño y $\| \ldots \|$ una función que asocia a cada fórmula su tamaño. Tales tamaños pueden ser evaluados en términos del número de símbolos lógicos que figuran en la(s) expresion(es) o bien del número de bits que expresan. El tamaño de pruebas podría usarse para determinar el de las fórmulas, pero no siempre es conveniente (no lo es, por ejemplo, si el tamaño de una prueba es el número de fórmulas que la componen).

Definición 73. Una lógica **L** es llamada un sistema epistémico si hay un subconjunto $K\mathcal{L}$ aserciones de conocimiento y una función $OC : K\mathcal{L} \Rightarrow \mathcal{L}$ que satisface las siguientes propiedades:

a) $OC(K\alpha) = \alpha$

b) OC es calculable en tiempo polinomial de $\|K\alpha\|$ y

c) si $(K\alpha)$ es válida según **L** entonces α también lo es.

Por supuesto que $K\alpha$ representa la aserción de que α es conocido y OC "extrae" el objeto de conocimiento de la aserción respectiva.

Definición 74. Sea **L** sistema epistémico. El fragmento reflejado, FR(**L**), de L es $K\mathcal{L} \cap \mathbf{L}$.

Es decir, FR(**L**) es el conjunto de aserciones de conocimiento válidas de **L**. Si $CO(FR(\mathbf{L})) = \mathbf{L}$ el fragmento reflejado se llama completo. Esto significa que para todo enunciado válido α hay una aserción de conocimiento $K\alpha$ válida. Con estos elementos podemos ya introducir los criterios de Artemov y Kuznetz para determinar que un sistema padece de omnisciencia lógica.

Definición 75. Sea E un sistema de prueba para un sistema epistémico **L**. E no es lógicamente omnisciente (o pasa LOT), si existe un polinomio P tal que para cualquier aserción de conocimiento $K\alpha$ válida, hay una prueba de α en E con el tamaño acotado por $P(\|K\alpha\|)$.

Definición 76. Sea E un sistema de prueba para un sistema epistémico **L**. E es fuertemente no lógicamente omnisciente (o pasa SLOT), si hay un algoritmo determinístico polinomial en ($\|K\alpha\|$) tal que, para cualquier aserción de conocimiento $K\alpha$ válida es capaz de restaurar una prueba de α.

En el artículo de 2006, la definición es dada para **L**, pero relativa al sistema de prueba. Además ambos criterios dependen de la forma en que es medido el tamaño de las fórmulas (y LOT también del tamaño de las pruebas).

Los autores demuestran inmediatamente que sistemas como K, K4, D y D4 no pasan estos test. La prueba es sencilla porque en ellos, una aserción de conocimiento es demostrable si y solo si su objeto de conocimiento lo es. Se puede probar algo similar para otros sistemas normales.

Con respecto a las lógicas de la justificación surge la cuestión de qué debe considerarse como una aserción de conocimiento (o de creencia). Parecería natural considerar a fórmulas del tipo $j : \alpha$ como expresando que j es una justifiación suficiente para que α sea conocida. De ser así la justificación debe ser factual, es decir, el sistema contendrá un esquema de axioma del tipo $t : F \to F$. Con la creencia esta condición no es necesaria. $j : \alpha$ puede leerse como j es una justificación para que α sea creída, o bien, j es la justificaci'on que S tiene para creer α. En todo caso, hay que tomar en cuenta que en $j : \alpha$ pueden aparecer constantes cuyo siginificado solo es dado a través de los axiomas de especificaciones constantes. Por ello, la aserción de conocimiento será del tipo:

$$\bigwedge CS \to j : \alpha$$

donde CS es un conjunto finito que especifica las constantes que aparecen en $j : \alpha$. Es con respecto a estas aseveraciones de conocimiento que Artemov y Kuznets demuestran que ciertos sistemas de lógica de la justificación pasan LOT y SLOT es decir son (fuertemente) no omniscientes. No entraré aquí en los detalles formales.

Kuznetz y Artemov consideran que el defecto de los sistemas que, antes que el suyo, afrontaron la omnisciencia lógica era su enfoque cualitativo. Las aserciones de conocimiento (o creencia) no registraban la información sobre las fuentes de ese conocimiento. El trabajo o tiempo que lleva al agente adquirirlo no aparecía. Además, algunos de ellos no dan cuenta de la competencia lógica del sujeto, a saber, los que determinan arbitrariamente el conjunto de fórmulas o enunciados que representan las creencias del individuo en una situación dada. Un enfoque cualitativo da pie a que podamos atribuir a un sujeto un conocimiento que no es factible que tenga. Como lo explican los autores en términos llanos: "Atribuimos el efecto de omnisciencia lógica a una situación donde para alguna aserción de conocimiento "corta" [del tipo] F es conocido, es imposible, encontrar factiblemente pruebas de F en E". (2009, p. 14). En cambio, en las lógicas de la justificación las atribuciones de conocimiento, gracias a que incorporan un término de justificación, permiten hallar la prueba de su objeto de conocimiento y estimar la verosimilitud de que el agente lo adquiera.

Aunque algunos de los autores de los sistemas que antes examiné sopesan sus propuestas en términos de complejidad computacional para mostrar sus ventajas, ninguno de ellos hace consistir en esto la solución al problema de la omnisciencia lógica. Artemov y Kuznetz son los primeros que proponen tests en términos precisos para determinar si un sistema de lógica epistémica o doxástica adolece de esta propiedad. Por otro lado, muestran que la lógica de la justificación no es lógicamente omnisciente, de acuerdo a estos criterios. Es decir, que toma en cuenta las limitaciones cognitivas de los agentes al ofrecer una medida del costo necesario para tener una creencia con una determinada justificación.

Evaluación 2

¿Qué podemos decir en términos de los retos de Harman? Antes que nada, la lógica de la justificación, en la interpretación que he propuesto, representa un refinamiento considerable. Como vimos en algunos ejemplos, no siempre una proposición debe o no ser integrada al cuerpo de creencias de un sujeto. Puede ocurrir que deba ser creída en virtud de una razón y no deba serlo si la razón en su favor es otra. Esto parece una importante ventaja de este sistema. Esto da un dimensión extra a la normatividad de la lógica. El que una proposición deba ser aceptada por un sujeto depende de las creencias que tiene y de las razones por las que las adoptó. La segunda ventaja, muy grande, es que podemos medir el esfuerzo cognitivo que toma a un agente realizar una inferencia para tener una creencia justificada. Podemos preguntarnos si una vez que tiene justificaciones para creer ciertas proposiciones qué tanto esfuerzo le lleva sacar alguna consecuencia lógica concreta de ellas. Por lo tanto, podemos medir qué tan verosímil es que adquiera esta creencia justificada. Aquí podríamos determinar que ciertas conclusiones son inmediatas, en términos precisos, y seguir con ello las propuestas de H. Field y F. Steinberger. Otras objeciones podrían tal vez afrontarse agregando otros recursos. Vimos que la objeción del constreñimiento puede ser evitada leyendo los principios puente en su versión W en lugar de C. La objeción de las trivialidades puede ser resuelta con un operador de consciencia o simplemente agregando al consecuente de un principio puente la frase “si... (la conclusión en cuestión) es traída a la atención del agente”. En cuanto a las creencias inconsistentes, recordemos que, por ejemplo, en LKDJ4 hemos introducido el axioma $\neg r :\perp$, pero tal vez sea posible diseñar un sistema de este tipo sin esta restricción.

Los más conocidos sistemas de lógica de la justificación tienen como base la lógica clásica. Artemov aclara que podrían desarrollarse estos sistemas usando otra lógica, pero ahora me atengo a los ejemplos que antes expuse. En ellos la lógica del agente coincide con la del sistema, lo que apunta a una evaluación externa que, sin embargo, afronta las demandas excesivas de una forma muy apropiada.

Sin embargo hay un aspecto en que la lógica de la justificación tiene un campo de acción restringido y no resuelve las objeciones de Harman. Los principios puente que obtendremos de ella serán del tipo: “si el agente cree X por la razón j, entonces debe creer Y por la razón s”. Sin embargo, esos principios solo puede aplicarlos el agente que conoce las razones por lo que fue llevado a creer el antecedente. Desde luego que en algunos contextos, por ejemplo, en algunas disciplinas sobre todo científicas, el individuo tiene un registro de estas razones. Sin embargo, en el contexto del razonamiento cotidiano, el sujeto no suele recordar por qué tiene cierta creencia.

Harman distinguió dos tipos de cambio de creencia según la estrategia que siguen. Uno es de corte fundacionista, en el que el agente necesita llevar un registro de las justificaciones que tiene para poder modificar racionalmente sus creencias. El coherentista lo niega. El punto de vista fundacionista sostiene que algunas de nuestras creencias dependen, para su justificación, de otras y estas de otras y así sucesivamente hasta llegar a ciertas creencias básicas con una justificación intrínseca. La revisión de creencias consistirá en descartar las que no tienen una justificación satisfactoria (o la pierden), y en agregar otras que están basadas en creencias ya adquiridas. En cambio, en el enfoque coherentista, el individuo puede adherirse a una creencia para lo cual no tiene justificación, excepto que no choca con el resto de sus creencias. Después de caracterizar ambas estrategias, Harman cita estudios empíricos y da ejemplos que muestran que solemos cambiar de creencias basados en un estilo coherentista. Por supuesto que esto no tiene que ver con cómo deberíamos proceder sino en cómo, de hecho, lo hacemos. Solo lo traigo a colación para agregar que la lógica de la justificación se amolda más a una estrategia fundacionista en la medida en que, leída normativamente, ofrece directrices para adquirir una creencia bajo cierta justificación cuando se tienen otras fundadas en ciertas razones. Para un sujeto que no recuerda la justificación última de sus creencias estas directrices son inúltiles. Sin embargo, esto solo acota la solución que vía la lógica de la justificación puede ofrecerse de la omnisciencia lógica.

Capítulo 4

Sistemas sin Omnisciencia lógica II

4.1. Revisión de creencias

Para introducir la lógica epistémica de tipo dinámico e ilustrar de qué manera en ella se plantea y se trata el problema de la omnisciencia lógica, comenzaré por exponer brevemente una estrategia para modelar el cambio de creencias que ha tenido un desarrollo reciente muy importante para mi tema. Se trata de las teorías de revisión de creencias que tiene sus orígenes en un artículo seminal de Alchurrón, G$\ddot{a}$nderfords y Mackinson de 1985. Clasifico a estas teorías como dinámicas simplemente porque de manera deliberada tratan el cambio de creencias.

Es cierto que los formalismos que veremos en esta sección no son de las misma naturaleza que el resto de los sistemas que he revisado. No son "lógicas" en el sentido en que no ofrecen o representan un aparato deductivo que norme las inferencias de un razonador "ideal" (aunque no omnisciente). Sin embargo, tengo varias razones para considerar su estudio relevante a mi tema. Una es que finalmente se presentan como un sistema normativo para la adquisión o eliminación de creencias. Son un marco para atribución de racionalidad y, en ese sentido, también pueden ser consideradas como estableciendo condiciones constitutivas de la creencia (o de la agencia). En segundo lugar, la teoría original de los mencionados autores "padecía de omnisciencia lógica" porque suponía, como veremos, que en todo momento las creencias de un individuo eran cerradas bajo consecuencia lógica. Las versiones más modernas de la teoría subsanan este defecto y es interesante ver con qué recursos lo hacen. En tercer lugar, veremos que con estos recursos responden a las objeciones de Harman y rescatan algunas de las intuiciones de MacFarlane, Field y Steinberger.

4.1.1. La teoría AGM de revisión de creencias

Esta teoría fue propuesta por los tres mencionados autores (a lo que debe el acrónimo con el que se le conoce) y pretende caracterizar, primeramente de forma axiomática, cómo se modifica (o, más bien, cómo debería transformarse) el conjunto de creencias de un individuo cuando se entera de una información y la agrega a su conjunto de creencias o, al contario, cuando quiere abandonar una de sus creencias. Dije “enterarse”, pero la teoría no depende de que dicha información sea verídica, sino de que el sujeto la tome como tal. La teoría distingue tres posibles operaciones. El caso más sencillo es el de la “expansión”. Se trata de agregar una proposición al conjunto de creencias del sujeto, sin importar si esto produce o no una inconsistencia en el conjunto resultante. El segundo caso, “contracción", es aquel en que el individuo quiere descartar una proposición de su conjunto de creencias. En el tercer caso, “revisión”, el individuo quiere incorporar consistentemente una proposición a sus creencias. Como veremos, la teoría se basa en dos principios guía: uno es que el sujeto modelado es (extremadamente) racional. El segundo es que las últimas dos operaciones están inspiradas por un principio de mutilación mínima, es decir, se trata de conservar en lo posible el conjunto de creencias que el sujeto tenía antes hacer el cambio.

Las tres operaciones mencionadas se delimitan por una serie de postulados. Más tarde unos resultados muestran cómo pueden ser caracaterizadas en términos conjuntistas y de consecuencia lógica. En las presentaciones de la teoría suele dejarse indeterminada la relación de consecuencia lógica de que se trate, lo que permite ampliar el rango de aplicación de la teoría. Normalmente se emplea un operador $\mathfrak{Cn}$ que aplicado a un conjunto general el conjunto de todas sus consecuencias lógicas. Diversas propiedades de $\mathfrak{Cn}$ se requieren en diferentes teoremas. Enuncio a continuación las más comunes. $\Gamma \models_{CL} \beta$ denotará que la fórmula β es consecuencia en la lógica clásica del conjunto Γ. Sea Σ un conjunto de enunciados.

1) $\Sigma \subset \mathfrak{Cn}(\mathfrak{t})$.

2) $\mathfrak{Cn}(\mathfrak{Cn}(\mathfrak{d})) = \mathfrak{Cn}(\mathfrak{d})$.

3) Si $\Sigma \subset \Gamma$ entonces $\mathfrak{Cn}(\mathfrak{t}) \subset \mathfrak{Cn}(\mathfrak{d})$.

4) Si $\Gamma \models_{CL} \beta$ entonces $\beta \in \mathfrak{Cn}(\mathfrak{d})$.

5) Si $\beta \in \mathfrak{Cn}(\mathfrak{d})$ entonces existe $\Sigma \subseteq \Gamma$, tal que Σ es finito y $\beta \in \mathfrak{Cn}(\mathfrak{t})$.

Es interesante observar en cada prueba a qué propiedades de este operador se recurre. Empecemos con la expansión. Denotaré por “E” al conjunto de enunciados del lenguaje.

Definición 77. Dado un operador $\oplus$: $\wp(E) \times (E) \rightarrow \wp(E)$, será llamado de expansión si satisface las siguientes propiedades: En esta sección representaré con letras mayúsculas griegas conjuntos de fórmulas y con minúsculas fórmulas.

($\oplus$1) $\Gamma \oplus \alpha$ es un conjunto creencia.

($\oplus$2) $\alpha \in \Gamma \oplus \alpha$

($\oplus$3) $\Gamma \subseteq \Gamma \oplus \alpha$

($\oplus$4) $\Gamma \oplus \alpha$ es el menor conjunto que satisface los ítems anteriores.

Aunque está caracterizada a través de postulados, es más fácil describir la expansión en términos de la relación de consecuencia.

Definición 78. Llamaremos conjunto-creencia o teoría a un conjunto de proposiciones cerrado bajo consecuencia lógica. Es decir, Γ es un conjunto creencia si y solo si $\mathfrak{Cn}(\mathfrak{d}) = \mathfrak{d}$.

La expansión puede ser caracterizada de la siguiente forma:

Teorema 79. $\Gamma \oplus \beta = \mathfrak{Cn}(\Gamma \cup \{\beta\})$

Las siguientes dos operaciones, llamadas respectivamente “contracción” y “revisión”, igualmente generan un conjunto de enunciados a partir de otro (las creencias del sujeto), y un enunciado más. Son caracterizadas a través de lo que los autores llaman “postulados de racionalidad”. La primera es la contracción (o retracción), donde una proposición β, que tal vez había sido aceptada como parte de un conjunto Γ de creencias, es ahora rechazada, junto con las proposiciones que la implican, y de tal manera que el conjunto resultante sea un conjunto-creencia y no contenga la proposición en cuestión. Está caracterizada por los postulados siguientes. Recuerde que E el conjunto de enunciados del lenguaje en cuestión.

Definición 80. Un operador $\ominus$: $\wp(E) \times (E) \rightarrow \wp(E)$, será llamado de contracción si cumple las siguientes propiedades:

Sea Γ un conjunto creencia y α una proposición.

($\ominus$1) $\Gamma \ominus \alpha$ es un conjunto creencia (Cerradura).

($\ominus$2) $\Gamma \ominus \alpha \subseteq \Gamma$ (Inclusión).

($\ominus$3) si $\alpha \notin \Gamma$ entonces $\Gamma \ominus \alpha = \Gamma$ (Vacuidad).

($\ominus$4) Si $\alpha \notin Con(\emptyset)$ entonces $\alpha \notin \Gamma \ominus \alpha$ (Éxito).

($\ominus$5) Si $\mathfrak{Cn}(\alpha) = \mathfrak{Cn}(\beta)$ entonces $\Gamma \ominus \alpha = \Gamma \ominus \beta$ (Extesionalidad).

($\ominus$6) $\Gamma \subseteq ((\Gamma \ominus \alpha) \oplus \alpha)$ (Recovery).

Cerradura garantiza que el sujeto transite siempre entre conjuntos lógicamente cerrados. Inclusión obedece al imperativo de máxima economía. Al eliminar una creencia el agente no deberá agregar ninguna otra. Vacuidad implica que el sujeto nada debe hacer si la proposición a eliminar no es una creencia suya. Éxito prescribe que salvo en el caso de que la proposición a descartar sea una tautología, no deberá aparecer en el conjunto resultante. Extensionalidad indica que la teoría es indiferente a la forma sintáctica con la que se expresa una proposición. El último axioma es el más controvertido. Si el agente se retracta de una creencia y posteriormente la suscribe de nuevo, no debe perder ninguna de las creencias que tenía antes de esta doble operación. De nuevo la justificación radica en el principio de máxima economía. La información es muy valiosa y el agente, al deshacerse de una proposición, no debe echar por la borda más de lo necesario. Por ello, cuando decide aceptar de nuevo esa proposición, debe recuperar lo que antes tenía.

Otra manera de presentar esta operación es a través de una construcción conjuntista. Los autores prueban unos teoremas de representación que muestran la equivalencia entre ambos enfoques. Daré un brevísimo esbozo de esta construcción.[1] Supongamos que tenemos un conjunto creencia Γ y queremos hacer una retracción de Γ con una proposición β. Si β es consistente con Γ, es decir, si β no es elemento de Γ, el resultado es simplemente Γ. Dejando de lado este caso trivial, veamos cómo puede hacerse la retracción. De cualquier subconjunto Ω de Γ diremos que es un subconjunto máximo de Γ que no implica β, si efectivamente es un subconjunto de Γ, no tiene a β como consecuencia y cualquier superconjunto de Ω ya implica a β. Denotaré con $\Gamma \perp \beta$ al conjunto de todos los subconjuntos máximos de Γ que no implican β. Formalmente:

Definición 81. $\Omega \in \Gamma \perp \beta$ si y solo si se cumplen las tres condiciones siguientes:
1) $\Omega \subseteq \Gamma$
2) $\beta \notin \mathfrak{Cn}(\Omega)$
3) Si $\Omega \subsetneq \Theta \subseteq \Gamma$ entonces $\beta \in \mathfrak{Cn}(\Theta)$.

[1] En parte lo hago porque construcciones similares aparecerán en otro tipo de revisiones.

Enlisto enseguida algunas de las propiedades de $\Omega \in \Gamma \perp \beta$:

Lema 82. *(lema 1)*
Si $\Theta \subset \Gamma$ *y* $\alpha \notin Con(\Theta)$ *entonces existe* Σ *tal que* $\Theta \subseteq \Sigma \subseteq \Gamma$ *y* $\Sigma \in \Gamma \perp \alpha$*.*

Lema 83. *(lema 2)*
$\Gamma \perp \emptyset = \emptyset$ *si y sólo si* $\alpha \in \mathfrak{Cn}(\emptyset)$*.*

Lema 84. *(lema 3)*
Si Γ *es un conjunto creencia y* $\Pi \in \Gamma \perp \alpha$ *entonces* $\Pi = \mathfrak{Cn}(\Pi)$*.*

Lema 85. *(lema 4)*
$\Gamma \perp \alpha = \Gamma \perp \beta$ *si y sólo si* $\forall x (x \subset \Gamma)$*,* $\alpha \in \mathfrak{Cn}(x) \leftrightarrow \beta \in \mathfrak{Cn}(x)$

Lema 86. *(lema 5)*
Sea Γ *un conjunto creencia y* α*,* β *enunciados, entonces:*

1) Si $\mathfrak{Cn}(\alpha) = \mathfrak{Cn}(\beta)$ *entonces* $\Gamma \perp \alpha = \Gamma \perp \beta$*.*

2) Si $\alpha \in \Gamma$ *y* $\Gamma \perp \alpha = \Gamma \perp \beta$ *entonces* $\mathfrak{Cn}(\alpha) = \mathfrak{Cn}(\beta)$*.*

Lema 87. *(lema 6)*
Sea Γ *una teoría y* α *una proposición. Si* $x \in \Gamma \perp \alpha$ *entonces para todo* $\gamma \in \Gamma$ *tal que* $\gamma \notin Con(x)$*,* $x \in \Gamma \perp \gamma$*.*

Ahora tomemos una función ϵ que selecciona algunos de los conjuntos máximos de Γ que no implican β. Si la intersección de los conjuntos así obtenidos no es vacía, apliquemos a esta intersección el operador $\mathfrak{Cn}$ que nos da todas sus consecuencias lógicas. El producto final resultará una contracción de Γ por β, a la que denotaré por $\Gamma \ominus_\epsilon \beta$. Formalmente:

Definición 88. ϵ es una función de elección para Γ si
Para todo $\alpha \in E$, $\emptyset \neq \epsilon(\Gamma \perp \alpha) \subseteq \Gamma \perp \alpha$ si $\Gamma \perp \alpha \neq \emptyset$ y
$\epsilon(\Gamma \perp \alpha) = \{\Gamma\}$ en otro caso.

Definición 89. Dada una función de elección ϵ para un conjunto de enunciados Γ, definimos la contracción parcial (*parcial meet contraction*) con ϵ de Γ por α, denotada $\Gamma \ominus_\epsilon \alpha$, por la siguiente identidad

$$\Gamma \ominus_\epsilon \alpha = \bigcap \epsilon(\Gamma \perp \alpha).$$

Es fácil demostrar que esta operación es una contracción. Viceversa, cada contracción es de este tipo, como se demuestra en el siguiente teorema.

Teorema 90. *Cualquier operador de retracción es una contracción meet parcial, es decir, para cualquier operador* $-$ *que satisface los postulados AGM de contracción existe una función* γ *tal que* $\Gamma - \alpha = \bigcap \gamma(\Gamma \bot \alpha)$.

Pasemos ahora a la revisión. Consiste en agregar una proposición a un conjunto de creencias de tal manera que el resultado sea cerrado bajo consecuencia lógica, consistente (si la entrada (*input*) no es contradictoria) y el cambio sea mínimo, tanto en lo que se refiere a las creencias que se eliminan, como a las que se añaden.

Definición 91. Dado un operador $\circledast$: $\wp(E) \times E \to \wp(E)$, será llamado de revisión si satisface las siguientes propiedades. Sea Γ un conjunto creencia y α una proposición.

($\circledast$1) $\Gamma \circledast \alpha$ es un conjunto creencia. (Cerradura)

($\circledast$2) $\alpha \in \Gamma \circledast \alpha$. (Éxito)

($\circledast$3) $\Gamma \circledast \alpha \subseteq \Gamma \oplus \alpha$. (Inclusión)

($\circledast$4) si $\neg\alpha \notin \Gamma$ entonces $\Gamma \oplus \alpha \subseteq \Gamma \circledast \alpha$. (Vacuidad)

($\circledast$5) $\neg\alpha \in \mathfrak{Cn}(\emptyset)$ (es decir, $\models \neg\alpha$) si y sólo si $\Gamma \circledast \alpha = \mathfrak{Cn}(\bot)$. (Consistencia)

($\circledast$6) Si $\mathfrak{Cn}(\alpha) = \mathfrak{Cn}(\beta)$ entonces $\Gamma \circledast \alpha = \Gamma \circledast \beta$. (Extensionalidad)

El primer postulado determina que después de la revisión el conjunto de creencias sea cerrado bajo consecuencia lógica. El segundo indica que el resultado de la revisión deberá incluir la entrada. Inclusión (o "cota superior") postula que la revisión no agregue nada más allá de la expansión. Vacuidad prescribe que si la entrada es consistente con el conjunto de creencias, entonces la revisión procede como la expansión (dado ($\circledast$3)). Consistencia asevera que el resultado de la revisión será consistente si la entrada lo es. Extensionalidad expresa que la teoría es indiferente a la sintaxis, no distingue entre revisiones de un mismo conjunto con proposiones equivalentes.

Para caracterizar de manera no axiomática esta operación, basta con recurrir a algunas conocidas funciones que transforman contracciones en revisiones o viceversa. Supongamos que tenemos una contracción $\bigodot$ y que defininos una operación $\bigotimes$ por la llamada ecuación de Levy: $\Gamma \bigotimes \beta \equiv_{def} (\Gamma \bigodot \neg\beta) \oplus \beta$. Se puede demostrar que $\bigotimes$ es una operación de revisión. Por otro lado la identidad de Harper tansforma una revisión en una contracción. Si $*$ es una revisión, la operación

$-$ definida por la identidad de Harper: $\Gamma - \beta \equiv_{def} (\Gamma * \neg\beta) \cap \Gamma$ es una contracción. Se puede demostrar que estas operaciones son mutuamente inversas, es decir que si partimos de una contracción $\odot$ y, usando la identidad de Levy, obtenemos una revisión, entonces si de esta definimos una contracción resulta la contracción original:

$$\Gamma \odot \alpha = ((\Gamma \odot \alpha) \oplus \neg\alpha) \cap \Gamma)$$

y si $\Gamma * \alpha$es una revisión:

$$\Gamma * \alpha = ((\Gamma * \alpha) \cap \Gamma) \oplus \alpha)$$

Dado que por cada función ϵ hay una retracción y una revisión, tenemos una gran cantidad de estas operaciones para normar las creencias del agente. Algunas generan resultados inverosímiles. Como ejemplo, considremos dos casos particulares: El primero está dado por la siguiente definición:

Definición 92. Llamaremos *contracción intersección completa (full meet contraction)* y la denotaremos por $\Gamma \sim \alpha$ a la operación definida por la siguientes cláusulas:

$$\Gamma \sim \alpha = \begin{cases} \Gamma & \text{si } (\Gamma \perp \alpha) = \emptyset, \\ \cap(\Gamma \perp \alpha) & \text{en caso contrario.} \end{cases}$$

Es decir, esta operación está determinada por una función ϵ que selecciona todos los elementos de $\Gamma \perp \alpha$. Sin embargo, si tomamos la identidad de levy para definir una revisión, $\Gamma \boxdot \alpha =_{Def} (\Gamma \sim \neg\alpha) \oplus \alpha$, resulta que $\Gamma \boxdot \alpha = \mathfrak{Cn}(\alpha)$. Evidentemente este es un resultado poco satisfactorio. Otros ejemplos similares lo ofrecen las llamadas contracciones *maxi–choice*. Consiste cada una de ellas en elegir un elemento del $\Gamma \perp \alpha$. Más precisamente:

Definición 93. Sea f una función tal que dado un conjunto $\Gamma \perp \alpha$ selecciona uno de sus elementos. Definimos la contracción maxi–choice de Γ por α determinada por f (lo que denotaremos por $\Gamma \div_f \alpha$) como

$$\Gamma \div_f \alpha = \begin{cases} f(\Gamma \perp \alpha) & \text{si } (\Gamma \perp \alpha) \neq \emptyset, \\ \cap(\Gamma \perp \alpha) & \text{en caso contrario.} \end{cases}$$

Definición 94. Llamemos revisión maxichoice a la revisión obtenida de una contracción maxichoice por la identidad de Levy. La denotaremos $\Gamma \dot{+} \alpha$, es decir, $\Gamma \dot{+}_f \alpha$=$((\Gamma \div_f \neg\alpha) \oplus \alpha\}$

El siguiente resultado muestra la inadecuación de las funciones maxichoice.

Lema 95. *Para cualquier teoría Γ y cualquier proposición α tal que $\neg\alpha \in \Gamma$, entonces $\Gamma \dot{+}_f \alpha$ es completa, es decir, para cualquier enunciado ω, $omega \in \Gamma \dot{+}_f \alpha$ o $\neg\omega \in \Gamma_f \dot{+} \alpha$.*

Algunas funciones selección no generan retracciones-revisiones verosímiles desde el punto de vista normativo. Con la noción de "atrincheramiento" es posible mejorar la teoría en este sentido. Intuitivamente el atrincheramiento de un individuo S a una creencia α es el grado en que S no está dispuesto a renunciar a α. ¿Cómo atrapar formalente ests noción? Para ilustrar una forma en que puede hacerse y mejorar así las propuestas anteriores, enlisto enseguida dos axiomas sugeridos por Gärdenfords que estipulan el comportamiento de una revisión con una fórmula conjuntiva.

$\circledast 7$ $\Gamma \circledast (\alpha \wedge \beta) \subset Con((\Gamma \circledast \alpha) \cup \{\beta\})$ para cualquier conjunto creencia Γ.

$\circledast 8$ $Con((\Gamma * \alpha) \cup \{\beta\} \subset \Gamma \circledast (\alpha \wedge \beta)$ para cualquier conjunto creencia Γ si $\neg\beta \notin \Gamma \circledast \alpha$.

Evidentemente estos postulados son una extensión de $\circledast 3$ y $\circledast 4$ y están motivados por el desideratum de que el cambio operado en Γ al agregar $\alpha \wedge \beta$ sea mínimo. Hay otros dos postulados de la retracción que, en un sentido que diré enseguida, corresponden a los dos anteriores y son:

$\ominus 7$ $(\Gamma \ominus \alpha) \cap (\Gamma \ominus \beta) \subset \Gamma \ominus (\alpha \wedge \beta)$ para cualquier conjunto creencia Γ.

$\ominus 8$ $\Gamma \ominus (\alpha \wedge \beta) \subset \Gamma \ominus \beta$ para cualquier conjunto creencia Γ si $\alpha \notin \Gamma \ominus (\alpha \wedge \beta)$.

La relación entre estos pares de axiomas está consignada en los dos siguientes teoremas.

Teorema 96. *Sea $\ominus$ un operador de retracción y $\circledast$ el correspondiente operador de revisión definido por la identidad de Levy, entonces $\ominus$ satisface $\ominus 7$ si y sólo si $\circledast$ satisface $\circledast 7$.*

Teorema 97. *Dado un operador de retracción $\ominus$ y el correspondiente operador de revisión $\circledast$ definido del primero por la identidad de Levy, $\ominus$ satisface $\ominus 8$ si y sólo si $\circledast$ satisface $\circledast 8$.*

Evidentemente los dos nuevos postulados para la revisión representan un refinamiento que hace al agente correspondiente más racional. Daré un breve ejemplo de cómo con la noción de atrincheramiento puede lograrse una mejora similar.

Supongamos dada una relación $\leq$ entre elementos de $\Gamma\bot\alpha$. La relación $\leq$ modela el grado en que el agente está dispuesto a renunciar a un conjunto maximal de creencias de Γ que no implican α. Dados dos de estos conjuntos x e y, $x \leq y$ indica que el agente no renunciaría a y sin antes o simultáneamente renunciar a x. Se puede postular que la relación $\leq$ satisfaga ciertas propiedades. El siguiente es un ejemplo de cómo estas condiciones sobre la relación $\leq$ afectan las operaciones de retracción y revisión.

Definición 98. Dada una relación binaria $\leq$ entre los enunciados del lenguaje, una función selección f es marcadora para $\leq$ si para cualquier conjunto creencia x y enunciado α

$$f(K\bot\alpha) = \{x \in (K\bot\alpha) | y \leq x \text{ para todo } y \in (K\bot\alpha)\}$$

es decir, f selecciona los conjuntos a los que el agente estaría menos dispuesto a renunciar.

Una ventaja de este concepto es que supone que la disposición de un individuo a abandonar creencias es holística, no recae sobre proposiciones aisladas.

El siguiente teorema muestra que una retracción determinada por una función selección que es marcadora de una relación binaria entre elementos de $K\bot\alpha$ modela mejor el comportamiento de un agente racional que una función selección arbitraria.

Teorema 99. *Sea $\ominus$ una contracción tal que*

$$K \ominus \alpha = \bigcap f(K\bot\alpha) \quad \textit{para toda } \alpha$$

donde f es una función selección que marca una relación binaria $\leq$ entre elementos de $K\bot\alpha$, entonces $\ominus$ satisface $\ominus_7$.

A diferencia de otros sistemas hasta ahora vistos, la teoría AGM pretende explícitamente resolver el problema planteado por Harman, a saber, establecer normas para el cambio racional de creencias en contextos informales. Sin embargo, en la medida en que los conjuntos de que parten y a que llevan las operaciones de esta teoría son cerrados bajo consecuencia lógica el problema de la omnisciencia lógica seguirá presente. En ese sentido no ofrecen un buen modelo, aunque sus postulados se llamen de racionalidad. Supone que el agente puede detectar si una proposición es consistente con el conjunto de todas sus creencias y que, en caso, contrario, es capaz de aceptar y generar un nuevo sistema de creencias lógicamente cerrado. Se ha tomado este rasgo como parte de una idealización aceptable. Sin embargo Harman, Cherniak y Goldman habían señalado no solo la imposibilidad de que un homínido real tuviera la capacidad para hacer tales operaciones sino, en muchos casos, la inconveniencia de hacerlas. Por otro lado, la teoría tiene la ventaja de que la consistencia a la que acabo de hacer mención no se refiere a la consecuencia clásica, sino una que es propia de la lógica del agente y de la que la teoría AGM no supone nada, excepto las propiedades del operador $\mathfrak{Cn}$ que fueron utilizadas en las demostraciones. En principio, podemos suponer que el sujeto modelado tiene otra lógica con respecto a la cual será omnisciente.

Presentaré enseguida una teoría de revisión de creencias que no genera conjuntos lógicamente cerrados.

4.2. Variantes de la teoría AGM sin omnisciencia lógica

La llamada teoría base de revisión de creencias es un refinamiento de la teoría AGM. Su principal característica es que no opera (solamente) con conjuntos cerrados bajo consecuencia lógica, sino con conjuntos arbitrarios de enunciados (que llamaré conjuntos base). Es decir, no predetermina la estuctura lógica del conjunto de creencias inicial del agente, pero sí impone algunos constreñimientos para el cambio de creencias y, estos, una vez más, se guían por un principio de minimalidad. Dado un conjunto-creencia Γ, una entrada α podía estar en tres situaciones con respecto a Γ. Podía ser un miembro de Γ, o su negación podía pertenecer a Γ, o bien ser independiente de este conjunto (ser consistente con él sin ser elemento suyo). Ahora que Γ no tiene por qué ser cerrado la primera posibilidad da lugar a dos opciones: pertencer α a Γ o ser consecuencia de Γ, pero no miembro suyo. En ambos casos diremos que el agente acepta α; en el primer caso, explícitamente; en el segundo, de manera implícita. De nuevo tendremos las tres operaciones de la teoría AGM. La expansión es fácilmente definida.

Definición 100. Si Γ es un conjunto arbitario de enunciados la expansión-base de Γ con β se define como $\Gamma \cup \{\beta\}$

Usaré la misma notación que antes y denotaré a este conjunto como $\Gamma \oplus \beta$.

La contracción-base

Pasemos ahora a la contracción. La idea es que si el agente cree α y hace una contracción con esta entrada no tenga después opinión respecto a α y, como antes, que haga la modificación más económica y racional posible. ¿Qué postulados de la retracción pueden conservarse? Cerradura, por supuesto que no. Éxito, Inclusión y Vacuidad, en cambio, no presentan problema en las siguientes versiones:

1) Éxito: si $\beta \notin \mathfrak{Cn}(\emptyset)$, entonces $\beta \notin \mathfrak{Cn}(\Gamma \ominus \beta)$.

2) Inclusión: $\Gamma \ominus \beta \subseteq \Gamma$

3) Vacuidad: si $\beta \notin \mathfrak{Cn}(\Gamma)$ entonces $\Gamma \ominus \beta = \Gamma$.

Note que, según Éxito, si β no es una tautología, no debe solamente estar excluida del conjunto resultante, sino que no debe ser consecuencia lógica de este. La idea es que después de la contracción el sujeto ya no tiene opinión sobre β (salvo si β es tautología). β desaparece incluso de sus creencias implícitas. En cuanto al axioma:

- Extensionalidad: Si $\mathfrak{Cn}(\alpha) = \mathfrak{Cn}(\beta)$ entonces $\Gamma \ominus \alpha = \Gamma \ominus \beta$

debemos incuirlo si nos queremos que la teoría tenga como objetos a los enunciados (y no a las proposiciones). Sin embargo, es necesario introducir un postulado un poco más fuerte del cual el axioma de Extensionalidad es un corolario, a saber

4) Uniformidad. Si para cada $\Sigma \subseteq \Gamma$, $\alpha \in \mathfrak{Cn}(\Sigma)$ si y solo si $\beta \in \mathfrak{Cn}(\Sigma)$, entonces $\Gamma \ominus \alpha = \Gamma \ominus \beta$.

A pesar de su mayor fuerza parece más intuitivo pues postula que las contracciones con α y con β coincidirán si estas proposiciones son equivalentes desde el punto de vista del sujeto. Más adelante volveré sobre este punto.

¿Qué debemos hacer con *Recovery*? Hansson (1991) muestra mediante el siguiente ejemplo que no podemos adoptar este postulado, en presencia de Inclusión y Éxito.

Ejemplo 101. Sea $-$ un operador que cumple inclusión y éxito. Sea $\Gamma = \{(P \wedge Q)\}$, Consideremos $(\Gamma - P) \oplus P$. Por inclusión $(\Gamma - P) \subseteq \Gamma$. Por éxito y dado que P no es tautología, $P \notin \mathfrak{Cn}(\Gamma - P)$. Por lo tanto, $(\Gamma - P) = \emptyset$ mientras que $(\Gamma - P) \oplus P = \{P\}$. No se cumple *Recovery* $(\Gamma \subseteq ((\Gamma - P) \oplus P))$.

Para reemplazar a este postulado y rescatar la idea de minimalidad con la que fue introducido, Hanson propuso tomar uno de dos posibles postulados. Ahora solo consideraré el primero:

5) Relevancia. Si $\beta \in \Gamma$ y $\beta \notin \Gamma \ominus \alpha$, entonces hay un conjunto Σ tal que $(\Gamma \ominus \alpha) \subseteq \Sigma \subseteq \Gamma$ y $\alpha \notin \mathfrak{Cn}(\Sigma)$, pero $\alpha \in \mathfrak{Cn}(\Sigma \cup \{\beta\})$

Es decir, al contraer Γ con α solo son susceptibles de removerse los enunciados β que, de alguna manera, contribuyen a implicar α. De nuevo, esto supone que el agente tiene buenas razones para eliminar las creencias que son abandonadas durante la retracción, pero también que no elimina más de la cuenta. Ahora veremos que Inclusión y Relevancia implican Vacuidad.

Teorema 102. *Si $\ominus$ es un operador que satisface Inclusión y Relevancia, entonces satisface también Vacuidad.*

Demostración. Supongamos que $\alpha \notin \mathfrak{Cn}(\Gamma)$, pero $\Gamma \ominus \alpha \neq \Gamma$. Debemos llegar a una contradicción. Puesto que $\Sigma \ominus \alpha \subsetneq \Gamma$, existe $\beta \in \Gamma$ y $\beta \notin \Gamma \ominus \alpha$. Por Relevancia, existe Σ, $\Gamma \ominus \alpha \subseteq \Sigma \subseteq \Gamma$ y es tal que $\alpha \notin \mathfrak{Cn}(\Sigma)$, pero $\alpha \in \mathfrak{Cn}(\Sigma \cup \{\beta\})$. Por otro lado, tenemos que $\beta \in \Gamma$, $\Sigma \subseteq \Gamma$ y $\alpha \notin \mathfrak{Cn}(\Gamma)$, por lo que $\alpha \notin \mathfrak{Cn}(\Sigma \cup \{\beta\})$, lo que es una contradicción. $\square$

Podemos entonces adoptar como postulados de la racionalidad para definir la contracción sobre conjuntos base a Éxito, Inclusión, Uniformidad y Relevancia. Para evitar riesgo de confusión llamaré a la operación definida por estos axiomas "contracción-base". Más formalmente:

Definición 103. Dado un operador $-\colon \wp(E) \times E \to \wp(E)$, será llamado de contracción-base si satisface los siguientes postulados. Sea Γ un conjunto creencia y α una proposición.

1) $\alpha \notin \Gamma - \alpha$. (Éxito)

2) $\Gamma - \alpha \subseteq \Gamma$. (Inclusión)

3) Si para cada $\Sigma \subseteq \Gamma$, $\alpha \in \mathfrak{Cn}(\Sigma)$ si y solo si $\beta \in \mathfrak{Cn}(\Sigma)$, entonces $\Gamma \ominus \alpha = \Gamma \ominus \beta$. (Uniformidad)

4) Si $\beta \in \Gamma$ y $\beta \notin \Gamma \ominus \alpha$, entonces hay un conjunto Σ tal que $(\Gamma \ominus \alpha) \subseteq \Sigma \subseteq \Gamma$ y $\alpha \notin \mathfrak{Cn}(\Sigma)$, pero $\alpha \in \mathfrak{Cn}(\Sigma \cup \{\beta\})$ (Relevancia)

Veamos ahora de qué manera puede ser caracterizada esta operación a través de una construcción. $\Gamma\perp\alpha$ se define como antes.

Definición 104. $\Theta \in \Gamma\perp\alpha$ si y solo si Θ cumple tres condiciones:
a) $\Theta \subset \Gamma$,
b) $\alpha \notin \mathfrak{Cn}(\Theta)$ y
c) si $\Theta \subset \Sigma \subseteq \Gamma$, entonces $\alpha \in \mathfrak{Cn}(\Sigma)$

$\Gamma\perp\alpha$ contiene a todos los subconjuntos máximos de Γ que no implican α. Obviamente si $\alpha \in \mathfrak{Cn}(\emptyset)$ (es decir, si α es una tautología), $\Gamma\perp\alpha = \emptyset$. Para probar el inverso se requiere el siguiente lema.

Lema 105. *Para cualquier conjunto* Γ*, si* $\Sigma \subseteq \Gamma$ *y* $\alpha \notin \mathfrak{Cn}(\Sigma)$*, existe* $\Theta \in \Gamma\perp\alpha$ *tal que* $\Sigma \subseteq \Theta$*.*

Los siguientes teoremas son análogos a los que di en la sección anterior, excepto que ahora Γ representa cualquier conjunto de enunciados.

Lema 106. *Si* $\Gamma\perp\alpha = \emptyset$ *entonces* $\alpha \in \mathfrak{Cn}(\emptyset)$*.*

Por otro lado, es claro que si $\alpha \notin \mathfrak{Cn}(\Gamma)$, entonces $\Gamma\perp\alpha = \{\Gamma\}$ (y viceversa).

Lema 107. $\Gamma\perp\alpha = \Gamma\perp\beta$ *si solo si para todo* Σ *tal que* $\Sigma \subseteq \Gamma$*, tenemos que* $\alpha \in \mathfrak{Cn}(\Sigma)$ *si y solo si* $\beta \in \mathfrak{Cn}(\Sigma)$*.*

Estos lemas harán menos sorprendente el teorema de caracterización que veremos enseguida. La construcción que define la contracción en el caso de la teoría AGM es idéntica a la que ahora presentaré solo que esta se aplica a conjuntos-base.

Definición 108. ϵ es una función de elección para Γ si
Para todo α, $\epsilon(\Gamma\perp\alpha) \subset \Gamma\perp\alpha$ si $\Gamma\perp\alpha \neq \emptyset$ y
$\epsilon(\Gamma\perp\alpha) = \{\Gamma\}$ en otro caso.

Definición 109. Dada una función de elección ϵ para un conjunto de enunciados Γ, definimos la contracción parcial con ϵ de Γ por α, denotada $\Gamma \ominus_\epsilon \alpha$, por la siguiente identidad

$\Gamma \ominus_\epsilon \alpha = \bigcap \epsilon(\Gamma\perp\alpha)$.

Lema 110. $\Gamma \ominus_\epsilon \alpha$ *es una contracción-base.*

De nuevo, se puede demostrar que toda contracción-base es una contracción parcial de este tipo:

Teorema 111. *Si* $\Gamma - \alpha$ *es una contracción-base y*

$$\epsilon(\Gamma\bot\alpha) == \begin{cases} \{\Sigma \in \Gamma\bot\alpha | \Gamma - \alpha \subset \Sigma\} & \textit{si } (\Gamma\bot\alpha) \neq \emptyset, \\ \epsilon(\Gamma\bot\alpha) = \{\Gamma\} & \textit{en caso contrario.} \end{cases}$$

entonces $\Gamma - \alpha = \Gamma \ominus_\epsilon \alpha$

Con ello obtenemos una caracterización de la contracción para conjuntos base. Ahora veremos una contracción similar.

La contracción Kernel

En 1994 Hansson propuso otra contracción que es, en algunos aspectos, el dual de la contracción parcial-meet que acabamos de ver. La idea es considerar, en lugar de los conjuntos máximos que no implican α, los conjuntos mínimos de los que α es consecuencia lógica. Como en la sección previa, trataré siempre de conjuntos-base, es decir, que no tienen que ser lógicamente cerrados.

Definición 112. Una operación $- : (\wp(E)) \times E \Rightarrow (\wp(E))$ es una contracción Kernel si y solo si satisface:
1) Inclusión: $\Gamma - \alpha \subseteq \Gamma$
2) Éxito: Si $\alpha \notin \mathfrak{Cn}(\emptyset)$, entonces $\alpha \notin \mathfrak{Cn}(\Gamma - \alpha)$
3) Uniformidad: Si para todo $\Sigma \subset \Gamma$, $\alpha \in \mathfrak{Cn}(\Sigma)$ si y solo si $\beta \in \mathfrak{Cn}(\Sigma)$, entonces $\Gamma - \alpha = \Gamma - \beta$.
4) Retención del núcleo (core-retainment). Si $\beta \in \Gamma$ y $\beta \notin \Gamma - \alpha$, entonces existe $\Theta \subset \Gamma$ tal que $\alpha \notin \mathfrak{Cn}(\Theta)$, pero $\alpha \in \mathfrak{Cn}(\Theta \cup \{\beta\})$.

De nuevo, hay para la contracción kernel un teorema de caracterización.

Definición 113. Sea Γ un conjunto base y α un enunciado, definimos $\Gamma \amalg \alpha$ por la siguiente condición: $\Sigma \in \Gamma \amalg \alpha$ si y solo si:
1) $\Sigma \subseteq \Gamma$
2) $\alpha \in \mathfrak{Cn}(\Sigma)$
3) Si $\Theta \subsetneq \Sigma$ entonces $\alpha \notin \mathfrak{Cn}(\Theta)$.

Es decir, $\Gamma \amalg \alpha$ contiene a los subconjuntos mínimos de Γ que implican α. De nuevo, enuncio brevemente algunas propiedades de este operador que facilitan la comprensión de los teoremas de caracterización.

Lema 114. *Si* $\Theta \subseteq \Gamma$, $\gamma \in \Gamma$ *y* $\alpha \in \mathfrak{Cn}(\Theta \cup \{\gamma\})$, *pero* $\alpha \notin \mathfrak{Cn}(\Theta)$, *entonces existe* $\Sigma \subseteq \Theta \cup \{\gamma\})$ *tal que* $\gamma \in \Sigma$ *y* $\Sigma \in \Gamma \amalg \alpha$.

Lema 115. *Si* $\Theta \subseteq \Gamma$, *y* $\alpha \in \mathfrak{Cn}(\Theta)$, *pero* $\alpha \notin \mathfrak{Cn}(\emptyset)$, *entonces existe* $\Sigma \subseteq \Theta$ *tal que* $\Sigma \in \Gamma \amalg \alpha$.

Lema 116. $\Gamma \amalg \alpha = \Gamma \amalg \beta$ *si y solo si para cualquier* $\Sigma \subseteq \Gamma$ $\alpha \in \mathfrak{Cn}(\Sigma)$ *si y solo si* $\beta \in \mathfrak{Cn}(\Sigma)$.

Definición 117. Una función incisión para un conjunto base Γ es una función ϕ tal que para cada α:
1) $\phi(\Gamma \amalg \alpha) \subset \bigcup(\Gamma \amalg \alpha)$
2) si $\emptyset \neq \Sigma \in (\Gamma \amalg \alpha)$ entonces $\Sigma \cap \phi(\Gamma \amalg \alpha) \neq \emptyset$.

La función incisión jurga un papel similar al de la función selección en la contracción parcial meet. Toma por lo menos un elemento en cada conjunto que pertenece a $\Gamma \amalg \alpha$. Los elementos seleccionados representan las creencias a las que el sujeto está menos apegado. Como en lo que sigue requeriré hablar de la diferencia entre dos conjuntos, a fin de evitar confusión, denotaré a esta operación con el símbolo $\setminus$. Es decir, $A \setminus B = A \cap B^c$

Definición 118. Si ϕ es una función incisión para un conjunto base Γ, llamaremos la ϕ-contracción kernel (denotada con el símbolo $\div_\phi$) a la operación definida por:

$$\Gamma \div_\phi \alpha = \Gamma \setminus \phi(\Gamma \amalg \alpha)$$

Teorema 119. *Si* ϕ *es una función incisión para un conjunto base* Γ, *entonces* $\Gamma \div_\phi \alpha$ *es una retracción kernel.*

El siguiente teorema puede considerarse el inverso del anterior pues demuestra que cada retracción Kernel es una ϕ-contracción para alguna función incisión ϕ.

Teorema 120. *Sea* $\dot{-}$ *una operación sobre* Γ *que satisface Éxito, Inclusión, Uniformidad y Retención del Núcleo. Sea* f *una función definida como:*

$$f(\Gamma \amalg \alpha) = \Gamma \setminus (\Gamma \dot{-} \alpha)$$

Entonces f *es una función incisión y* $\Gamma \div_f \alpha = \Gamma \dot{-} \alpha$.

Demostración. Observemos primeramente que f está bien definida. El conjunto $\Gamma \amalg \alpha$ podría ser determinado por referencia a β, es decir como $\Gamma \amalg \beta$ pero, en ese caso, como ya vimos, para cualquier $\Sigma \subseteq \Gamma$, $\alpha \in \mathfrak{Cn}(\Sigma)$ si y solo si $\beta \in \mathfrak{Cn}(\Sigma)$.

Entonces, por uniformidad, $\Gamma \dot{-} \alpha = \Gamma \dot{-} \beta$.
Veamos ahora que $\Gamma \div_f \alpha = \Gamma \dot{-} \alpha$. Por inclusión, $\Gamma \dot{-} \alpha \subset \Gamma$. De allí que:
$\Gamma \div_f \alpha = \Gamma \setminus f(\Gamma \perp\!\!\!\perp \alpha) = \Gamma \setminus (\Gamma \setminus (\Gamma \dot{-} \alpha)) = (\Gamma \dot{-} \alpha)$.
Mostraré ahora que f es una función incisión. Supongamos que $\beta \in f(\Gamma \perp\!\!\!\perp \alpha)$, es decir, $\beta \in \Gamma \setminus (\Gamma \dot{-} \alpha)$. Por Retención del Núcleo existe $\Sigma \subseteq \Gamma$ tal que $\alpha \notin \mathfrak{Cn}(\Sigma)$ y $\alpha \in \mathfrak{Cn}(\Sigma \cup \{\beta\})$. Por el lema XXX, existe $\Theta \in \Gamma \perp\!\!\!\perp \alpha$ tal que $\Theta \subseteq \Sigma \cup \{\beta\}$ y $\beta \in \Theta$. En consecuencia $\beta \in \bigcup \Gamma \perp\!\!\!\perp \alpha$. Con ello queda demostrado que $f(\Gamma \perp\!\!\!\perp \alpha) \subseteq \bigcup \Gamma \perp\!\!\!\perp \alpha$.
Sea $\emptyset \neq \Theta \in \Gamma \perp\!\!\!\perp \alpha$ y suponga que $\Theta \cap f(\Gamma \perp\!\!\!\perp \alpha) = \emptyset$. Claramente $\alpha \notin \mathfrak{Cn}(\emptyset)$. Ya probamos que $\Gamma \dot{-} \alpha = \Gamma \setminus f(\Gamma \perp\!\!\!\perp \alpha)$. Se sigue que $\Theta \subseteq \Gamma \dot{-} \alpha$. Por lo tanto, $\mathfrak{Cn}(\Theta) \subseteq \mathfrak{Cn}(\Gamma \dot{-} \alpha)$ y como $\alpha \in \mathfrak{Cn}(\Theta)$ entonces $\alpha \in \mathfrak{Cn}(\Gamma \dot{-} \alpha)$ lo que contradice Éxito. Concluimos que si $\emptyset \neq \Theta \in \Gamma \perp\!\!\!\perp \alpha$ entonces $\Theta \cap f(\Gamma \perp\!\!\!\perp \alpha) \neq \emptyset$. Por lo tanto f es una función incisión. $\square$

Ahora que tenemos caracterizadas tanto axiomática como lógicamente a la contracciones base y Kernel para conjuntos base, veamos cómo generar con ellas revisiones.

Revisión de conjuntos base

Se utilizan ahora dos identidades para generar revisiones a partir de las contracciones que acabamos de ver. La primera es la ya conocida identidad de Levy y la segunda es su "inverso". Si $-$ representa una contracción base o Kernel, las dos identidades en cuestión son:

$$\Gamma \circledast \alpha = (\Gamma - \neg\alpha) \oplus \alpha$$

$$\Gamma \circledast \alpha = (\Gamma \oplus \alpha) - \neg\alpha$$

Usando dichas contracciones se generan cuatro tipos de revisión. Como en los casos anteriores estas operaciones pueden ser definidas axiomáticamente. En lo que sigue doy un muy breve resumen de las cuatro opciones. A las revisiones que parten de la identidad de Levy suele llamárseles "internas"; a las que se generan por el inverso de esa identidad, externas. Antes que nada enuncio los axiomas que, en cierta combinación, servirán para definirlas.

- Éxito. $\alpha \in \Gamma \circledast \alpha$.
- Inclusión. $\Gamma \circledast \alpha \subseteq \Gamma \oplus \alpha$.
- No-contravención. Si $\neg\alpha \notin \mathfrak{Cn}(\emptyset)$ entonces $\neg\alpha \notin \mathfrak{Cn}(\Gamma \circledast \alpha)$.

- Retención del Núcleo. Si $\beta \in \Gamma$ y $\beta \notin \Gamma \circledast \alpha$, entonces existe Σ tal que $\Sigma \subset \Gamma \cup \{\alpha\}$, $\neg\alpha \notin \mathfrak{Cn}(\Sigma)$ y $\neg\alpha \in \mathfrak{Cn}(\Sigma \cup \{\beta\})$.

- Relevancia. Si $\beta \in \Gamma$ y $\beta \notin \Gamma \circledast \alpha$, entonces existe Σ tal que $\Sigma - \alpha \subset \Sigma \subset \Gamma \cup \{\alpha\}$, $\neg\alpha \notin \mathfrak{Cn}(\Sigma)$ y $\neg\alpha \in \mathfrak{Cn}(\Sigma \cup \{\beta\})$.

- Uniformidad. Si para cada $\Sigma \subset \Gamma$, $\neg\alpha \in \mathfrak{Cn}(\Sigma)$ si y solo si $\neg\beta \in \mathfrak{Cn}(\Sigma)$, entonces $\Gamma \cap (\Gamma \circledast \alpha) = \Gamma \cap (\Gamma \circledast \beta)$

- Uniformidad Débil. Si $\alpha \in \Gamma$ y $\beta \in \Gamma$ y para cada $\Sigma \subset \Gamma$, $\neg\alpha \in \mathfrak{Cn}(\Sigma)$ si y solo si $\neg\beta \in \mathfrak{Cn}(\Sigma)$, entonces $\Gamma \cap (\Gamma \circledast \alpha) = \Gamma \cap (\Gamma \circledast \beta)$

- Pre-expansión. $(\Gamma \oplus \alpha) \circledast \alpha = \Gamma \circledast \alpha$.

Ahora veremos las caracterizaciones de las cuatro revisiones mencionadas. Sean Γ un conjunto base y α y β enunciados. Revisión base interna: Si ϵ es una función elección, la operación

$$\Gamma \circledast_{\epsilon} \alpha = \bigcap \epsilon(\Gamma \bot \neg\alpha) + \alpha$$

Es la revisión parcial meet interna de Γ por α (determinada por ϵ) y se caracteriza por satisfacer Éxito, Inclusión, no-contravención, Relevancia y Uniformidad (es decir, cualquier operación que satisfaga estos postulados es una revisión parcial meet interna para una ϵ adecuada). Revisión base externa: si ϵ es una función elección, la operación

$$\Gamma \circledast_{\epsilon} \alpha = \bigcap \epsilon((\Gamma + \alpha) \bot \neg\alpha)$$

Es la revisión parcial meet externa de Γ por α (determinada por ϵ) y se caracteriza por satisfacer Éxito, Inclusión, No-contravención, Relevancia, Pre-expansión y Uniformidad Débil (es decir, cualquier operación que satisfaga estos postulados es una revisión parcial meet externa para una ϵ adecuada). Revisión kernel interna: si ϕ es una función incisión, la operación

$$\Gamma \circledast_{\phi} \alpha = (\Gamma \setminus \phi(\Gamma \perp\!\!\!\perp \neg\alpha)) + \alpha$$

es la revisión Kernel interna de Γ por α (determinada por ϕ) y se caracteriza por satisfacer Éxito, Inclusión, No-contravención, retención del Núcleo y Uniformidad (es decir, cualquier operación que satisfaga estos postulados es una revisión kernel interna para una ϕ adecuada). Revisión Kernel externa: si ϕ es una función incisión, la operación

$$\Gamma \circledast_{\phi} \alpha = (\Gamma + \alpha) \setminus \phi((\Gamma + \alpha) \amalg \neg\alpha)$$

es la revisión Kernel externa de Γ por α (determinada por ϕ) y se caracteriza por satisfacer Éxito, Inclusión, No-contravención, retención del Núcleo, Preexpansión y Uniformidad Débil (es decir, cualquier operación que satisfaga estos postulados es una revisión Kernel externa para una ϕ adecuada).

4.2.1. Semirevisiones

Para las cuatro variantes recién presentadas es válido el axioma de Éxito. Pero recordemos que tanto Goldman como Harman señalaron que, en ocasiones, es conveniente rechazar la conclusión de una inferencia (quizás para revisar alguna de las premisas). De igual manera podemos suponer que la nueva información que promueve la revisión es a fin de cuentas rechazada en base a otras consideraciones. Hansson propuso una forma de revisión, llamada semi-revisión, que se distingue de las anteriores porque no satisface Éxito. La idea es sencilla: en lugar de ordenar las creencias con una función selección o incisión y después agregar la nueva entrada, ahora el sujeto aceptará esta y después someterá sus creencias alguno de los mecanismos de selección provistos por esas funciones. Esta segunda parte del proceso se llama consolidación. Como resultado de la consolidación la entrada puede ser finalmente rechazada.

Veamos primeramente la semirevisión kernel. De nuevo representaré con ϕ a una función incisión, es decir, que cuando se aplica a $\Gamma \amalg \alpha$ y este no es vacío tomará un elemento (al menos) de cada elemento no vacío de $\Gamma \amalg \alpha$. En caso contrario dará como resultado $\{\Gamma\}$. Supondremos que el lenguaje contiene un enunciado $\perp$ y diremos que un conjunto de enunciados Σ es inconsistente si $\perp \notin \mathfrak{Cn}(\Sigma)$. Tal presuposición suele hacerse para poder aplicar la teoría a lógicas no-clásicas. Una semi-revisióm kernel se define de la siguiente manera:

Definición 121. Si ϕ es una función incisión, la semi-revisión kernel de Γ por α (determinada por ϕ), a la que denotaré por $\Gamma?_{\phi}$, está definida por:

$$\Gamma?_{\phi}\alpha =_{Def} (\Gamma + \alpha) \setminus \phi((\Gamma + \alpha)\amalg \perp)$$

De los postulados que caracterizan esta operación algunos son conocidos, otro son versiones reformadas de axiomas anteriores, y dos son completamente nuevos. Los enlisto a continuación:

1. Inclusión. $\Gamma?_{\phi}\alpha \subset \Gamma \oplus \alpha$

2. Pre-expansión. $(\Gamma \oplus \alpha)?_\phi\alpha = \Gamma?_\phi\alpha$

3. Retención del núcleo$^\perp$. Si $\beta \in \Gamma$, pero $\beta \notin \Gamma?_\phi\alpha$, entonces hay un conjunto Σ consistente tal que $\Sigma \subseteq \Gamma$ y $\Sigma \cup \{\beta\}$ es inconsistente.

4. Consistencia. Si α es consistente, entonces $\Gamma?\alpha$ también lo es.

5. Intercambio interno. Si $\alpha \in \Gamma$ y $\beta \in \Gamma$ entonces $\Gamma?\alpha = \Gamma?\beta$

Veamos por último la semi-revisión parcial meet. Se define de la siguiente manera:

Definición 122. Si ϵ es una función elección, la semi-revisión parcial meet de Γ por α (determinada por ϵ), a la que denotaré por $\Gamma?_\epsilon$, está definida por:

$$\Gamma?_\epsilon\alpha =_{Def} \bigcap(\Gamma + \alpha)\perp(\perp))$$

Los postulados que caracterizan esta operación son los siguientes:
1) Inclusión. $\Gamma?_\epsilon\alpha \subset \Gamma \oplus \alpha$
2) Pre-expansión. $(\Gamma \oplus \alpha)?_\epsilon\alpha = \Gamma?_\epsilon\alpha$
3) Relevancia$^\perp$. Si $\beta \in \Gamma$, pero $\beta \notin \Gamma?_\phi\alpha$, entonces hay un conjunto Σ consistente tal que $\Gamma?_\epsilon\alpha \subseteq \Sigma \subseteq \Gamma$ y $\Sigma \cap \{\beta\}$ es inconsistente.
4) Consistencia. Si α es consistente, entonces $\Gamma?\alpha$ también lo es.
5) Intercambio interno. Si $\alpha \in \Gamma$ y $\beta \in \Gamma$ entonces $\Gamma?\alpha = \Gamma?\beta$

No ahondaré más en el tema de la semirevisión. Mi intención en este breve presentación fue mostrar cómo usando las mismas construcciones que en las secciones anteriores es posible mejorar aún más la teoría. Frente a una información nueva el agente considera la posibilidad de agregarla consistentemente a su cuerpo de creencias, lo que parece más racional que una operación que finaliza inevitablemente con la creencia en esa información.

4.2.2. Evaluación

Las revisiones sobre conjuntos base representan un avance considerable con respecto a la teoría AGM original. Recordemos que los postulados respectivos no determinan una operación particular sino un grupo de operaciones y que una revisión específica está dada por una función selección o de una función incisión determinada. Por lo pronto, ¿cómo debemos leer sus postulados? Me parece que la manera más natural de leerlos es como prescribiendo lo que el agente debe hacer una vez que de hecho tiene un conjunto de creencias y se enfrenta a una información novedosa o desea retirar una de sus creencias. Es decir, si Γ es el conjunto de creencias explícitas de un individuo $\Gamma \circledast \alpha$, es el conjunto de creencias que debe tener si quiere incorporar α a Γ consistentemente. Así consideradas, las cuatro revisiones que acabamos de ver no suponen ni generan agentes lógicamente omniscientes, gracias a que no obedecen al postulado de cerradura. Con ello, evitan o disminuyen el problema de las demandas excesivas y permiten hacer una distinción entre creencias implícitas y explícitas. Si el conjunto de creencias de un individuo A en un momento dado es Γ, A cree explícitamente α si $\alpha \in \Gamma$, mientras que cree esa proposición implícitamente si $\alpha \in \mathfrak{Cn}(\Gamma)$. De ese modo, la teoría ofrece salvaguarda contra la objeción de trivialidades.

La generación de conjuntos lógicamente cerrados es excesivamente demandante en un doble sentido: porque el estado inicial en cada paso es un conjunto infinito, lo que complica acomodarlo a la entrada y porque el resultado debe de nuevo ser un conjunto lógicamente cerrado. Esta dificultad, que podemos asmilar a la objeción de demandas excesivas, ha sido señalada por varios autores. Por ejemplo, Hans Rott, dice refiriéndose a los agentes epistémicos supuestos por la teoría AGM: "...son idealmente competentes con respecto a temas de lógica. Deben aceptar las consecuencias de las creencias que sostienen... y ver que sus teorías son consistentes" (Rott, 2000) A ello replicó Isaac Levy (1991) proponiendo que el conjunto creencia contiene los enunciados que el agente está comprometido a creer, más bien que aquellos que de hecho cree. Dicho en otras palabras, Levy propone una interpretación normativa más bien que descriptiva de la teoría. Sin embargo, ya hemos visto en qué sentido esto tampoco es del todo correcto.

Las revisiones sobre conjuntos base evitan la omnisciencia lógica. ¿Cómo lo consiguen? Recordemos que en la teoría AGM, las revisiones provienen de las contracciones y estas, a su vez, de operaciones sobre el conjunto $\Gamma\perp\alpha$. La cerradura reside, en última instancia, en que los elementos de $\Gamma\perp\alpha$ son cerrados bajo consecuencia lógica. Recordemos que Γ era un conjunto cerrado y si Σ es un subconjunto máximo de Γ que no implica α, y $\beta \in \mathfrak{Cn}(\Sigma)$, entonces obviamente $\beta \in \Gamma$, y, por la definición de $\Gamma\perp\alpha$, si β no estuviera en Σ, entonces $\alpha \in \mathfrak{Cn}(\Sigma \cup \{\beta\}) = \mathfrak{Cn}(\Sigma)$ lo que es contradictorio. Ahora bien, si Γ es un conjunto base, $\Sigma \subseteq \Gamma$ y $\beta \in \mathfrak{Cn}(\Sigma)$ entonces no tenemos garantía de que $\beta \in \Gamma$, así que solo podremos realizar el razonamiento anterior para $\beta \in \Gamma$. Algo similar ocurre con la propiedad de uniformidad que radica en que $\Gamma\perp\alpha = \Gamma\perp\beta$ cuando y solo cuando α y β son lógicamente equivalentes dentro de Γ, es decir, solo desde el punto de vista de las creencia explícitas del agente. Aún podríamos estar exigiendo demasiado al creyente, pero hay una sensible mejora con respecto a la versión AGM que conduce al postulado de extensionalidad. Sería demasiado exigente si el conjunto de creencias del agente es muy grande.

En este punto las ventajas del enfoque con conjuntos base pueden apreciarse desde otra perspectiva. Advirtamos que los conjuntos-creencia son infinitos, lo que torna muy complicado el uso de criterios de selección de elmentos de $\Gamma\bot\alpha$ y de $\bigcup(\Gamma \amalg \alpha)$. En cambio, los conjuntos base pueden ser finitos, aún si para representar situaciones reales se requieren conjuntos grandes. En ese caso, el número de las funciones selección e incisión es también es finito. Así es posible determinar una por una todas las posibles revisiones de un conjunto base finito con una proposición dada.

Veamos con ejemplos muy sencillos cómo proceden estos sistemas con las consecuencias lógicas de un cojunto de creencias. Supongamos que Pedro cree que (1) Christian Coleman tiene el record de los 100 metros planos y (2) que su prima segunda la mayor es veterinaria. En el enfoque AGM también creerá que (3) Christian Coleman tiene el record de los 100 metros planos o Aristóteles escribio *La República*; y también (4) que Christian Coleman tiene el record de los 100 metros planos si y solo si su prima segunda la mayor es veterinaria. Ahora desea remover de sus creencias el enunciado (1). Tendrá que remover también (4) o (2). En el enfoque AGM estas dos creencias están en pie de igualdad y la función selección deberá descartar una de ellas. En el enfoque de los conjuntos base, si (1) y (2) fuesen las únicas creencias explícitas del sujeto, creería (3) y (4) solo de manera implícita. La retracción no trata (2) y (4) por igual. Al eliminar (1) de sus creencias explícitas, (4) también desaparece de sus creencias implícitas.

Pondré otro ejemplo que ilustra las ventajas y defectos del postulado de inclusión. Supongamos que las creencias del sujeto comprenden que (5) ninguno de los seres humanos que vivieron antes de Cristo puede entrar al reino de los cielos y (6) que Sócrates vivió antes de Cristo. Tendría que aceptar que Sócrates no puede entrar a ese sitio, pero el individuo quiere remover este enunciado de sus creencias. En el enfoque AGM podrá eliminar (5) o podría substituirlo por la creencia de que casi ninguno de los seres humanos anteriores a Cristo pudo ingresar a tan selecto lugar. Dado que su conjunto creencia es infinito tiene muchas posibilidades. En cambio, en el enfoque de los conjuntos base el resultado de la retracción tiene que ser un subconjunto del conjunto original, lo que limita notablemente las posibilidades de elección.

Lo anterior es tal vez una limitación importante, pero asimismo le otorga a la teoría una ventaja sobre su predecesora. En el enfoque AGM si para cualquier enunciado α, $\alpha \in \Gamma$ si y solo si $\alpha \in \Sigma$, entonces para cualesquier revisión $\circledast$ y retracción $\ominus$, $\Gamma - \alpha = \Sigma - \alpha$ y $\Gamma \circledast \alpha = \Sigma \circledast \alpha$. Es decir, en esas condiciones la teoría no distingue entre dos conjuntos con las mismas consecuencias lógicas. En cambio, la teoría de los conjuntos base distingue entre $\{P, Q, R\}$ y $\{(P \wedge Q), R\}$, como se ve por el postulado de inclusión.

Por eso mismo, la teoría AGM no trata adecuadamente los conjuntos lógicamente inconsistentes. Por ejemplo, si nuestra lógica es clásica
$\mathfrak{Cn}(p, q, r, \neg r) = \mathfrak{Cn}(\neg p, \neg q, r, \neg r)$

Sin embargo, una retracción con r (cuando el sujeto descubre la inconsistencia) dará diferentes resultados en la teoría de Hansson cuando se aplique a $p, q, r, \neg r$ que cuando se aplique a $\neg p, \neg q, r, \neg r$. La teoría que opera con conjuntos base es mejor en este sentido.

En cuanto al rigor, la objeción no se aplica aquí con la misma fuerza porque, aunque supongamos implícito un operador de obligación, la teoría provee al agente con una adecuada guía para incoroporar o eliminar información. Sin embargo, el agente no deja de ser racional porque se aparte un poco de lo que la teoría le prescribe. En ese sentido difiere su situación de la tortuga de lewis Carroll. De igual forma, la teoría no funciona para la creencia como atribución pues el que alguien no siguiera las "prescripciones" de la misma no contaría como un criterio para no atribuirle ciertas creencias.

En la última sección busqué ilustrar brevemente cómo la teoría puede mejorarse o adapatarse a otras situaciones modificando alguno de los postulados pero reteniendo las ideas centrales. Las semi-revisiones no requieren que la revisión de un conjunto Γ por una proposición α resulte en una incorporación de α a Γ. Una consideración del caso podría llevar al agente a rechazar α. Esta mejora puede verse como una respuesta generalizada a la objeción del constreñimiento de Harman. Otro refinamiento, mucho más antiguo de la teoría propone la revisión de un conjunto no por una proposición sino por otro conjunto de enunciados.

La teoría de revisión de creencias ofrece una interesante respuesta a las objeciones de Harman.

4.3. El Sistema de Duc

En un artículo de 1997, Ho Ngoc Duc presenta un sistema para modelar el razonamiento de un agente no lógicamente omnisciente. Caracteriza el problema de la omnisciencia lógica por el hecho de que en los sistemas de lógica epistémica tradicionales son verdaderas las afirmaciones siguientes:

a) si $\models \alpha$ entonces $\models \mathcal{K}\alpha$ (NEC),

b) si $\models \alpha \rightarrow \beta$ entonces $\models \mathcal{K}\alpha \rightarrow \mathcal{K}\beta$ (MON) y

c) si $\models \alpha \leftrightarrow \beta$ entonces $\models \mathcal{K}\alpha \leftrightarrow \mathcal{K}\beta$ (CGR).

Para Duc, los intentos previos de resolver este problema, basados en su mayoría en la estrategia de debilitar la lógica del agente, tienen dos desventajas opuestas: una es que, al invalidar NEC, MON, y CGR, empobrecen excesivamente las capacidades del agente. La otra es que, aún así, los postulados son demasiado exigentes para encontrar sujetos reales que los satisfagan. Eluden la omnisciencia, pero a un costo muy alto:

> [...] de este modo... muchas intuiciones acerca del conocimiento y la creencia se pierden. Mostraré que los axiomas para las lógicas epistémicas deben tener la siguiente forma: si el agente sabe todas las premisas de una regla de inferencia, y si piensa lo suficiente, alcanzará la conclusión. Para formalizar tal idea, me propongo dinamizar la lógica. (Duc, 1997, p. 633)

El enfoque es dinámico y permitirá registrar la adquisición de nueva información por el uso que hace un sujeto de una regla de inferencia en un momento determinado. Inicialmente la idea era introducir en el lenguaje, además de los símbolos usuales de la lógica epistémica, operadores de la forma $\langle R \rangle$ (y sus duales $[R]$) donde R representa alguna regla de inferencia lógica. Una fórmula del tipo $\langle R_i \rangle \alpha$ significará que después de una aplicación de la regla R por el agente i, α será verdadera. Como siempre $[R_i]\alpha$ significará que α es verdadera después de cualquier aplicación de la regla R por el agente i. Normalmente nos interesarán fórmulas en que siguiendo a la aparición de uno de estos operadores vendrá una fórmula epistémica. Por ejemplo, supongamos que el sistema contiene un operador $\langle MP \rangle_i$ que representa un aplicación por el agente i de *modus ponens*, entonces serán válidas fórmulas del tipo:

$$(\mathcal{K}_i\alpha \wedge \mathcal{K}_i(\alpha \rightarrow \beta)) \rightarrow \langle MP_i \rangle \mathcal{K}_i\beta.$$

Es decir el individuo sabrá β si sabe α y sabe $(\alpha \rightarrow \beta)$, pero no en automático, sino como resultado de una aplicación concreta del *modus ponens*. Asimismo el sistema podrá modelar la aplicación sucesiva de varias reglas de inferencia. Por ejemplo, (omitiendo por el momento el subíndice) podremos tener un esquema válido del siguiente tipo.

$$\mathcal{K}(\neg\neg\neg\alpha \wedge (\alpha \vee \beta)) \rightarrow \langle \mathcal{SC} \rangle \langle \mathcal{EDN} \rangle \langle \mathcal{SD} \rangle \mathcal{K}\beta$$

donde "SC", "EDN" y "SD" simbolizan respectivamente la simplificación de la conjunción, la eliminación de la doble negación y el silogismo disyuntivo. Duc hace varias simplificaciones, dos de las cuales subyacen a la validez de estas fórmulas. Estamos suponiendo que mientras el agente aplica una regla de inferencia, a) no olvida las premisas y b) ninguna de estas se vuelve falsa. En particular, la operación que hace el agente no puede cambiar el valor de verdad de las proposiciones que la regla en cuestión se aplica (a diferencia de lo que ocurre en la lógica del anuncio público). Hay otra simplificación que hace Duc y que, en mi opinión, debilita su propuesta. Si fuese a seguir el curso de acción hasta aquí indicado, podría elegir una axiomatización de la lógica clásica y tratar cada axioma y cada regla de inferencia como una operación atómica que el agente puede realizar. Dos inconvenientes disuaden a Duc de seguir esta vía, que luego otros autores tomarán: la arbitrariedad en la elección de estas acciones atómicas y la complejidad del sistema resultante. Duc prefiere utilizar un solo operador inferencial $\langle F_i \rangle$ (y su dual) en donde F_i representa cualquier regla de inferencia (de las que el sistema admite) o, incluso, una sucesión de varias de ellas. Más precisamente, supongamos que se le concede al agente i la capacidad de inferir usando alguna de las reglas R_i^1, R_i^2,...R_i^n. Entonces F_i puede ser considerada como:

$$(R_i^1 \cup R_i^2 \cup ... \cup R_i^n)^+$$

donde $\cup$ representa (1) la elección de alguna R_i^j y (2) su aplicación, mientras que el exponente + simboliza la reiteración (un número finito de veces) de (1) y (2) en ese orden. Dicho en otras palabras, F_i representa la acción de aplicar consecutivamente una sucesión de esas reglas elegidas arbitariamente. Ahora al lenguaje epistémico clásico agregamos un solo operador $\langle F_i \rangle$ (y su dual $[F_i]$) por cada agente i, con la idea de que $\langle F_i \rangle \alpha$ signifique que, después de una inferencia del agente i, α es verdadero. También podríamos darle una lectura subjuntiva: "si el agente i realizara cierta acción inferencial, entonces α sería verdadero". La definición de "fórmula bien formada" es la usual excepto por un punto: el operador inferencial $\langle F_i \rangle$ no debe figurar en el alcance de un operador epistémico. Tal restricción parece razonable. Una fórmula del tipo $\mathcal{K}_i \langle F_i \rangle \mathcal{K}_i \alpha$ parece absurda. Que el agente supiera que si realizara una acción epistémica sabría tal y cual cosa (que ahora no sabe), es poco intuitivo.

Veamos ahora el sistema $DES4_n$ que Duc propone para regular el comportamiento de su operador de acción inferencial. Para ello, es necesario definir el fragmento L_E^+ del lenguaje:

Definición 123. 1) Las fórmulas construidas recursivamente de letras proposicionales y conectivos veritativo-funcionales pertenecen a L_E^+,

2) Si $\{\alpha, \beta\} \subset L_E^+$, entonces $\{(\alpha \vee \beta), (\alpha \wedge \beta), K_i\alpha\} \subset L_E^+$

Además de una base axiomática para el cálculo de proposiciones, $DES4_n$ contiene los siguientes axiomas:

TL1 $[F_i](\alpha \to \beta) \to ([F_i]\alpha \to [F_i]\beta)$

TL2 $[F_i]\alpha \to [F_i][F_i]\alpha$

DE1 $(\mathcal{K}_i\alpha \wedge \mathcal{K}_i(\alpha \to \beta)) \to \langle F_i\rangle\mathcal{K}_i\beta$.

DE2 $\mathcal{K}_i\alpha \to \alpha$

DE3 $\mathcal{K}_i\alpha \to [F_i]\mathcal{K}_i\alpha$ si $\alpha \in L_E^+$

DE4 $\langle F_i\rangle\mathcal{K}_i(\alpha \to (\beta \to \alpha))$

DE5 $\langle F_i\rangle\mathcal{K}_i(((\alpha \to (\beta \to \gamma)) \to ((\alpha \to \beta) \to (\alpha \to \gamma)))$

DE6 $\langle F_i\rangle\mathcal{K}_i((\neg\beta \to \neg\alpha) \to (\alpha \to \beta))$

DE7 $\langle F_i\rangle\mathcal{K}_i(\mathcal{K}_i\alpha \to \alpha)$

DE8 $\mathcal{K}_i\alpha \to \langle F_i\rangle\mathcal{K}_i\mathcal{K}_i\alpha$ si $\alpha \in L_E^+$

Además del *modus ponens*, el sistema tiene como regla de inferencia a la "necesitación" (es decir, de $\vdash \alpha$ se puede pasar a $\vdash [F_i]\alpha$).

DE3 es válido bajo el supuesto que hicimos de que α no cambia de valor de verdad como resultado de la acción del agente. Duc llama a las fórmulas con esta propiedad "persistentes" y observa que no pertenecen a esta categoría las que son de la forma $\neg K_i\beta$ o equivalentes, pues puede ocurrir que el individuo se sepa ignorante de algo pero, como efecto de una inferencia, llegue a saberlo y esté consciente de ello. De allí la restricción a L_E^+. Tal vez la medida sea excesiva, pero es una estrategia de seguridad. Lo mismo puede decirse de la salvedad impuesta en DE8. Vemos que haciendo un esfuerzo el agente puede llegar a conocer cada una de las tautologías. Que con algún trabajo puede enterarse de los axiomas del cálculo de enunciados es el contenido de los axiomas DE4, DE5 y DE6.

Antes de proseguir hago dos observaciones. Al principio de su artículo Duc dice que para él las leyes de la lógica no son proposiciones sobre el mundo, sino reglas, y conocerlas solo significa que el sujeto es capaz de recurrir a ellas para obtener información a partir de lo que ya sabe. Sin embargo, vemos que si el sujeto se esfuerza lo suficiente puede llegar a conocer leyes de la lógica (todas, potencialmente), lo que contraviene la declaración de Duc. Por otro lado, es un poco sorprendente la expresión "$[F_i][F_i]\alpha$", pues si bien cada figuración de "F_i" puede aludir a diferentes procesos inferenciales, su concatenación significa la puesta en marcha secuencial de dos de estas operaciones, lo que a su vez es un solo proceso inferencial que podría ser representado por una sola figuración de "$[F_i]$".

Lema 124. *Si α es una tautología, entonces* $\vdash_{DES4_n} \langle F_i\rangle\mathcal{K}_i\alpha$

Demostración. Lo probaré por inducción en la longitud de la prueba. Si α es alguno de los axiomas del cálculo de enunciados que son subfórmulas de DE4, DE5 o DE6, sabemos que con un poco de esfuerzo, el agente los conocerá. Para el paso de inducción, supongamos ahora que tenemos $\langle F_i\rangle\mathcal{K}_i\alpha$ y $\langle F_i\rangle\mathcal{K}_i(\alpha \to \beta)$. Por necesitación, el axioma TL1 y por dualidad, es válida la regla: si $\vdash \alpha \to \beta$ entonces $\vdash \langle F_i\rangle\alpha \to \langle F_i\rangle\beta$. La demostración seguiría así:

1. $(\mathcal{K}_i\alpha \to (\mathcal{K}_i(\alpha \to \beta) \to \langle F_i\rangle\mathcal{K}_i\beta))$ (DE1).
2. $\langle F_i\rangle\mathcal{K}_i\alpha \to \langle F_i\rangle(\mathcal{K}_i(\alpha \to \beta) \to \langle F_i\rangle\mathcal{K}_i\beta)$. (Regla derivada)
3. $\langle F_i\rangle\mathcal{K}_i\alpha \to (\langle F_i\rangle(\mathcal{K}_i(\alpha \to \beta) \to \langle F_i\rangle\langle F_i\rangle\mathcal{K}_i\beta)$. (Regla derivada).
4. $\langle F_i\rangle K_i\alpha \to (\langle F_i\rangle(K_i(\alpha \to \beta) \to \langle F_i\rangle K_i\beta)$. (Dual de TL2).

Aplicando dos veces *modus ponens* concluimos que $\langle F_i\rangle K_i\beta$. Observe que de 2) a 3) aplicamos la regla derivada al condicional del axioma $DE3$, lo que es válido porque β es una fórmula proposicional. □

Cor $\vdash_{DES4_n} \langle F_i\rangle\mathcal{K}_i(\alpha \to \beta) \to (\langle F_i\rangle\mathcal{K}_i\alpha \to \langle F_i\rangle\mathcal{K}_i\beta)$ y
$\vdash_{DES4_n} \mathcal{K}_i(\alpha \to \beta) \to (\langle F_i\rangle\mathcal{K}_i\alpha \to \langle F_i\rangle\mathcal{K}_i\beta)$

Además, el agente puede saber, después de algunas inferencias, que su conocimiento es verídico (DE5) y, de ciertas proposiciones que sabe, puede saber que las sabe, siempre después de algún esfuerzo cognitivo (DE8). Que el sistema es consistente puede apreciarse eliminado los operadores inferenciales y epistémicos de los axiomas y de las reglas de inferencia. Lo que obtenemos son tautologías y la regla de *modus ponens*. De ese modo todas los teoremas se convierten en tautologías, por lo que una fórmula y su negación no podrán ser ambas teoremas. Por otro lado, es fácil probar ni NEC, ni MON, ni CGR son válidas para $DES4_n$ como tampoco lo son las fórmulas que expresan la omnisciencia lógica (como las que aparecen en el segundo capítulo).

4.3.1. Evaluación

El agente que describe este sistema no es lógicamente omnisciente, aunque sí puede llegar a saber cada tautología y algunas otras verdades lógicas del sistema, como la veracidad de su conocimiento de una proposición determinada. No podrá saber, en cambio, que su conocimiento siempre es verídico, ni, por ejemplo, proposiciones de la forma DE7.

El objetivo explícito de Duc era descriptivo, pues para él las proposiciones epistémicas deben ser interpretadas como condicionales del tipo: si el agente sabe X y si se esfuerza lo suficiente, llegará a saber Y, donde Y es consecuencia de X según una regla de inferencia. De nuevo, con axiomas diferentes podríamos modelar las inferencias de un agente con una lógica diferente. Por ejemplo, en lugar de DE4, DE5 y DE6, pudimos poner esquemas de la forma$\langle F_i \rangle \mathcal{K}_i \alpha$, donde α es un axioma de la lógica intuicionista, lo que desde luego no nos impediría usar la lógica clásica para tratar los conocimientos del agente. Podríamos demostrar que, dado cualquier enunciado válido de la lógica intuicionista, el agente lo conocerá si se esfuerza lo suficiente. No podríamos probar, en cambio, que sea incapaz de conocer otros principios lógicos (por ejemplo, la ley del tecero excluido). En cualquier caso el propósito es descriptivo. Una lectura normativa de las fórmulas sería inadecuada. por ejemplo, vimos que si α es una tautología, $\langle F_i \rangle \mathcal{K}_i \alpha$ es un teorema del sistema, lo que podríamos leer como que, en tal caso, el agente llegará a saberlo, si hace ciertas inferencias. Interpretarlo como aseverando que el sujeto deberá saber esa tautología es incorrecto. Tampoco es posible interpretar los teoremas del sistema como reglas de atribución de creencias.Así es que debemos preguntar qué tan bien el sistema representa el concepto de creencia o describe las creencias de un sujeto.

La introdución de un solo operador $\langle F_i \rangle$ que condensa la aplicación sucesiva de diversas operaciones inferenciales es una simplificación que permite a Duc utilizar axiomas de la lógica temporal para su propio sistema. Sin embargo, el costo es muy alto. Es posible probar que si $\vdash_{PC} \alpha \rightarrow \beta$ entonces $\vdash_{DES4_n} \mathcal{K}_i \alpha \rightarrow \langle F_i \rangle \mathcal{K}_i \beta$. Es decir que si β y γ son consecuencias de α, una trivial y la otra muy complicada, el sistema tendrá como teoremas $\mathcal{K}_i \alpha \rightarrow \langle F_i \rangle \mathcal{K}_i \beta$ y $\mathcal{K}_i \alpha \rightarrow \langle F_i \rangle \mathcal{K}_i \gamma$ sin que quede registrado el esfuerzo respectivo que tomó al agente llegar a saber β y llegar a saber γ ni el modo particular en que cada una de estas proposiciones fue conocida. En ese sentido, me parece que la lógica de justificación es mejor porque mantiene un registro de las operaciones que el agente lleva a cabo para alcanzar un resultado.

Recordemos que Duc se ha propuesto eludir la omnisciencia lógica conservan-do ciertas intuiciones sobre el conocimiento y la creencia que en intentos previos al suyo se perdían. Por supuesto que, con un pequeño cambio en los axiomas el sistema puede adaptarse sin problemas, al caso de la creencia. Sin embargo, su enfoque no provee una definición o representación de estos conceptos. Me parece que una semántica de mundos posibles tipo Kripke funciona mejor a este respecto. El sistema de Duc nos indica las situaciones epistémicas a las que puede llegar el sujeto por el uso reiterado de las reglas de inferencia que le otorgamos. Si tenemos un sistema formal SF con un conjunto de axiomas AX y reglas de inferencia de la lógica clásica, entonces si T es un terorema de SF, sabemos que si el sujeto de Duc sabe AX podría llegar a saber T. Sin embargo, no sabemos si podrá hacerlo en el tiempo que le queda de vida, ni qué operaciones le permitirán acceder a estas creencias. Eso no dice nada acerca del conocimiento o la creencia. Además para que el observador pueda saber que si el agente sabe los axiomas, con algún esfuerzo sabrá el teorema, es necesario que él mismo haya hecho ese esfuerzo. La lógica del observador es prácticamente la misma que la del agente.

4.4. El sistema de Agotnes y Alechina

Agotnes y Alechina introdujeron en 2007 unas "lógicas... que son... capaces de expresar la propiedad de que, aunque el conocimiento del agente no es cerrado bajo consecuencia lógica, el agente puede conocer algunas reglas de inferencia y ser capaz de derivar a la larga consecuencias de sus creencias" (Agotnes y Alechina, 2007, p. 84). Se trata de "modelar el razonamiento como una alternativa a la omnisciencia lógica" (ibidem). Ellos mismos califican su enfoque como sintáctico. En contraste con el recurso a mundos posibles de Kripke para representar las creencias del agente, el enfoque sintáctico adscribe directamente un valor de verdad arbitrario a fórmulas que atribuyen creencias. Entre las ventajas que los autores encuentran a este tratamiento están la flexibilidad que evita la omnisciencia lógica y permite atribuir a un individuo creencias contradictorias. Sin embargo, la presentación no es (primariamente) axiomática, sino a través de modelos constituidos por "mundos posibles", entendidos esta vez como funciones que asignan valores de verdad a fórmulas, α, del lenguaje del agente (lo que a su vez determina los valores de verdad de fórmulas atómicas, $\mathcal{B}\alpha$, del lenguaje del obervador). Los mundos representan las creencias o los conocimientos de un agente en un momento determinado, mientras que la relación de accesibilidad entre ellos simboliza acciones que llevan al individuo de un estado epistémico a otro. Las proposiciones que el sujeto cree están expresadas en un cierto lenguaje LO. Por otro lado, el lenguaje ML del sistema está formado a partir de átomos epistémicos, es decir, enunciados de la forma $\mathcal{B}\alpha$ (donde α es una proposición de LO) con la aplicación reiterada de los conectivos proposicionales clásicos, un operador modal ($\diamond$) y dos nuevos operadores epistémicos ($\triangle$ y $\triangledown$) cuyo significado explicaré enseguida. Un modelo de ML es una estructura de la forma $M = \langle W, R, V\rangle$, donde $W \neq \emptyset$ y $R \subseteq W \times W$ y v es una función que otorga a cada átomo epistémico en cada mundo un valor de verdad, con la peculiaridad de que solo un número finito de estos átomos son verdaderos en cada mundo posible. Es decir, las creencias del individuo en cada momento forman un conjunto finito. Otra novedad la consituyen los dos operadores que introducen Agotnes y Alechina. Ambos se aplican no a proposiciones individuales, sino a conjuntos de proposiciones de LO. El primero es $\triangle$. Si $\{\rho_1, \rho_2, \ldots \rho_n\}$ es un conjunto de proposiciones de LO, entonces la fórmula $\triangle\{\rho_1, \rho_2, \ldots \rho_n\}$ no es más que una abreviatura de $(\mathcal{B}\rho_1 \wedge \mathcal{B}\rho_2 \wedge \ldots \wedge \mathcal{B}\rho_n)$ y puede ser parafraseada como "el agente cree al menos $\rho_1, \rho_2 \ldots$ y ρ_n". El segundo operador es $\triangledown$. La fórmula $\triangledown\{\rho_1, \rho_2, \ldots \rho_n\}$ significará que el agente cree a lo más $\rho_1, \rho_2, \ldots$y ρ_n. Siguiendo a los autores en un primer momento solo introduciré $\triangle$.

Supongamos dado un conjunto infinito de proposiciones del lenguaje objeto LO.

Definición 125. El conjunto de proposiciones de ML está definido por la siguiente especificación recursiva:

1) Si ρ es una fórmula de LO, $\mathcal{B}\rho$ es una fórmula (atómica) de ML.

2) Si α y β son fórmulas de ML también lo son $\neg\alpha$ y $(\alpha \wedge \beta)$

3) $\diamond\alpha$ es una fórmula de ML, si α también lo es.

Una fórmula del tipo $\diamond\alpha$ simbolizará que el agente puede hacer alguna inferencia en un estado como resultado de lo cual α será verdadero en el siguiente estado. Los otros conectivos y $\Box$ se definen de la manera usual, mientras que:

$$\triangle\{\rho_1, \rho_2, \ldots \rho_n\} \equiv_{Def} (\mathcal{B}\rho_1 \wedge \mathcal{B}\rho_2 \wedge \ldots \wedge \mathcal{B}\rho_n)$$

Una vez que tenemos el conectivo $\triangle$, el símbolo $\mathcal{B}$ resulta superfluo, pues $\triangle\{\alpha\}$ es lo mismo que $\mathcal{B}\alpha$. En lugar de $\triangle\{\alpha\}$ usaré $\triangle\alpha$. Por otro lado, si

$$\Sigma = \{\rho_1, \rho_2 \ldots \rho_n\}$$

es un conjunto de proposiciones de LO, denotaré por $\triangle(\Sigma)$ al conjunto

$\{\mathcal{B}\rho_1, \mathcal{B}\rho_2, \ldots \mathcal{B}\rho_n\}$ de proposiciones atómicas de ML. Análogamente emplearé $\triangle^{-1}$ para la operación inversa.

Definición 126. Un modelo es una triada $M = \langle W, R, V\rangle$ en que W es un conjunto cuyos elementos serán llamados mundos, R una relación binaria de elementos de W ($R \subset WXW$) y V una función que a cada mundo asocia un conjunto finito de proposiciones de LO.

Si eliminamos la palabra "finito" de la definición anterior, obtenemos la definición de "modelo general".

Definición 127. Dado un modelo $M = \langle W, R, V\rangle$, y un mundo w, el que una fórmula α de ML sea verdadera en w ($M, w \models \alpha$) está definido por la siguiente especificación recursiva:

- $M, w \models \mathcal{B}\rho$ si y sólo si $\rho \in V(w)$, donde ρ es una fórmula atómica de LO.
- $M, w \models (\alpha \wedge \beta)$ si y sólo si $M, w \models \alpha$ y $M, w \models \beta$
- $M, w \models \neg\alpha$ si no es el caso que $M, w \models \alpha$
- $M, w \models \diamond\alpha$ si y solo si existe $v \in W$ tal que wRv y $M, v \models \alpha$

Como siempre, la validez es verdad en todos los mundos posibles de todos los modelos. Sin sorpresa, la lógica que obtenemos es **K**, es decir, el mínimo sistema normal de lógica modal proposicional. La axiomatización acostumbrada de **K** que consiste en una axiomatización del cálculo de proposiciones, el axioma de distribución ($\Box(\alpha \to \beta) \to (\Box\alpha \to \Box\beta)$), el *modus ponens* y la regla de necesitación (de $\vdash \alpha$ pase a $\vdash \Box\alpha$) es adecuada y suficiente para este sistema. La única dificultad para demostrarlo es que, en el modelo canónico de K, un mundo puede validar un conjunto infinito de fórmulas, es decir, es un modelo general. Algunas maniobras son necesarias para demostrar que cualquier fórmula consistente de ML tiene un modelo. Ahora bien, que tengamos un sistema normal no significa que una fórmula del tipo ($\mathcal{B}\alpha \to (\mathcal{B}(\alpha \to \beta) \to \mathcal{B}\beta)$) sea válida, pues debemos recordar que las fórmulas epistémicas $\mathcal{B}\gamma$ son átomos de ML. El esquema K de distribución es válido con respecto al operador $\Box$. Es decir, en el lenguaje ML del observador tenemos la lógica clásica, pero al agente no le hemos adscrito ninguna lógica. Es posible expresar algunas relaciones epistémicas particulares. Lo autores dan los siguientes ejemplos:

- $\triangle\{p,(P \to Q)\} \to \diamond\triangle Q$ el agente puede hacer *modus ponens* con P y $P \to Q$
- $\triangle P \to \diamond\neg\triangle P$ el agente puede olvidar P en el siguiente estado.
- $\triangle P \to \Box\triangle P$ el agente debe recordar P en el siguiente estado.

Ninguna de esas fórmulas es válida. Sería conveniente otorgarle al agente al menos la capacidad de hacer inferencias en cualquier situación. Para ello podemos pedir que los modelos sean seriales, es decir, que cada mundo tenga acceso a otro (o a él mismo). Podemos axiomatizar la validez correspondiente agregando a K el esquema axiomático $\mathbf{D}$ (es decir, $\Box\alpha \to \diamond\alpha$). Si queremos que cada mundo tenga a lo más un sucesor (es decir que si wRv y wRz entonces $v = z$) y que, por tanto, el razonamiento sea determinístico, podemos axiomatizar la correspondiente lógica agregando a los axiomas de $\mathbf{K}$ el esquema axiomático F, a saber, $\diamond\alpha \to \Box\alpha$.

Para formar el lenguaje $ML^{\triangledown}$, incorporemos al vocabulario de ML el operador $\triangledown$ y, a sus reglas, la condición de que si $\{\rho_1, \rho_2, \ldots \rho_n\}$ es un conjunto finito de fórmulas de LO, entonces $\triangledown\{\rho_1, \rho_2, \ldots \rho_n\}$ es una fórmula de ML. A la definición de verdad de una fórmula en un mundo de un modelo de ML, anexamos la siguiente cláusula:

$$M, w \models \triangledown\{\rho_1, \rho_2, \ldots \rho_n\}$$

si y sólo si

$$V(w) \subset \{\rho_1, \rho_2, \ldots \rho_n\}$$

Claramente

$$M, w \models \triangle\{\rho_1, \rho_2, \ldots \rho_n\} \text{ si y sólo si } \{\rho_1, \rho_2, \ldots \rho_n\} \subset V(w)$$

A la combinación

$$\triangledown\{\rho_1, \rho_2, \ldots \rho_n\} \wedge \triangle\{\rho_1, \rho_2, \ldots \rho_n\}$$

los autores la abrevian $\boxtimes\{\rho_1, \rho_2, \ldots \rho_n\}$. Evidentemente esta fórmula expresa que el agente conoce (o cree) exactamente las proposiciones de LO contenidas en el referido conjunto.

Una peculiaridad del conectivo $\triangledown$ es que puede una fórmula de la forma $\triangledown\{\rho_1, \rho_2, \ldots \rho_n\}$ ser verdadera en un mundo w bajo una asignación V y volverse falsa bajo una asignación V' aunque V y V' coincidan en validar los elementos de $\{\rho_1, \rho_2, \ldots \rho_n\}$. Por ejemplo, si $V(w) = \{\rho_1, \rho_2, \ldots \rho_n\}$ y $V'(w) = \{\delta, \rho_1, \rho_2, \ldots \rho_n\}$ con δ

una proposición de LO distinta de cada ρ_i, entonces $M, w \models \triangledown\{\rho_1, \rho_2, \dots \rho_n\}$ y $M', w \not\models \triangledown\{\rho_1, \rho_2, \dots \rho_n\}$ (si V y V' son respectivamente las valuaciones de M y M') . De eso se valen Agotnes y Alechina para demostrar que el operador $\triangledown$ no es definible en términos de los otros operadores y conectivos del lenguaje.

Si los mundos de un modelo solo verifican un número finito de proposiciones de LO y, por lo tanto, solo un número finito de proposiciones atómicas de ML (pues $\mathcal{B}\alpha$ es verdadero en w si y solo si α es verdadero en w), entonces no parece que sea necesario el operador $\triangledown$. Sin embargo, es útil para expresar que el agente sabe exactamente tales y cuales proposiciones (y ninguna otra) en un mundo dado, pero asimismo se puede emplear para expresar a través de esquemas ciertos rasgos epistémicos del agente. Agotnes y Alechina dan el siguiente ejemplo:

$$\triangledown(\{\phi, \neg\phi\} \cup \Gamma) \to \Box(\triangledown(\{\phi\} \cup \Gamma) \vee \triangledown(\{\neg\phi\} \cup \Gamma)) \qquad (*)$$

La fórmula expresa el hecho de que si el agente llegara a creer una contradicción siempre podría resolverla en el siguiente estadio.

Para el sistema $K^{\triangledown}$, que resulta al agregar este operador a ML la referida condición semántica, Agotnes y Alechina ofrecen la axiomatización que resulta de agregar a K los siguientes axiomas:

E1) Si $X \nsubseteq Y$, $\triangle X \to \neg\triangledown Y$

E2) $(\triangledown(Y \cup \{\gamma\}) \wedge \neg B\gamma) \to \triangledown Y$

E3) Si $X \subseteq Y$, $(\triangledown X \to \triangledown Y)$

De nuevo, la demostración de completud (débil) de este sistema axiomático con respecto a la semántica dada requiere de algunos artificios. El modelo canónico es un modelo general (es decir, donde $V(w)$ puede ser infinito). Para convertirlo en un modelo hay que utilizar un tipo particular de filtraciones, es decir, substituir en $V(w)$ las fórmulas atómica irrelevantes (para la evaluación de una fórmula particular) por una fórmula fija $\mathcal{B}\rho$. No es necesario aquí entrar en esos detalles.

4.4.1. Evaluación

Vemos que este sistema respeta los límites entre el agente y el observador, tanto en lo que se refiere al lenguaje como a las inferencias que cada uno puede hacer. Ofrece un lenguaje que permite describir los estados doxásticos o epistémicos del agente y el paso entre unos y otros por efecto del razonamiento o de otros factores (como el olvido). Como en el sistema de Duc, podemos atribuirle al agente el conocimiento de ciertas reglas lógicas. Por ejemplo, si una regla R permite inferir γ de β y α:

$$\boxtimes(\{\alpha,\beta\}\cup\Sigma)\rightarrow\diamond\boxtimes(\{\alpha,\beta,\gamma\}\cup\Sigma) \qquad (**)$$

expresa que el agente es capaz de aplicar R. Es decir, si algún estadio sabe Σ y α y β entonces puede ocurrir que en el estadio siguiente sepa eso y además γ.
Por otro lado, si hay un regla que al aplicarse a un conjunto de premisas elimina la última (lo que llamaré más tarde una regla diminutiva), podemos expresar su conocimiento por parte del agente de la siguiente manera. Supongamos que la regla R^* aplicada a α y β (en ese orden), produce α (es decir, elimina β). Entonces

$$(\triangle\{\beta\}\rightarrow\neg\nabla\Sigma)\wedge(\boxtimes(\{\alpha,\beta\}\cup\Sigma)\rightarrow\diamond\boxtimes(\{\beta\}\cup\Sigma)) \qquad (***)$$

expresa que el individuo conoce la regla R^*. El segundo conyunto dice que si en una situación sabe exactamente Σ y α y β, puede ocurrir que en el siguiente estadio sepa Σ y β y nada más. Sin embargo, podría ser que β se hubiera conservado en Σ (en cuyo caso el agente no habría aplicado R^*). Para evitarlo, el primer conyunto expresa que $\{\beta\}$ no es subconjunto de Σ. Sin embargo, el sistema no describe propiamente la acción del agente. Este podría haber eliminado β, no por un razonamiento del tipo R^*, sino por alguna otra causa. Lo mismo puede decirse del ejemplo previo. Hecha esta salvedad el sistema es completamente liberal en este punto, es decir, puede adaptarse (como el de Rantala) a cualquier lógica que queramos atribuirle al agente. Hasta aquí no lo hemos limitado en ningún aspecto, pero podría hacerse restringiendo el conjunto de mundos posibles a los que satisfacen propiedades determinadas.

En términos de los principios puente los ejemplos anteriores admiten una lectura en términos de modalidad de tener razones o autorización. Por ejemplo, podría leerse la fórmula (**) anterior como: si el agente tiene entre sus creencias a α, a β y a los elementos de Γ está autorizado a tener esas mismas creencias, además de Γ. Por supuesto que también podríamos interpretarlo bajo la modalidad de obligación. Por ejemplo, (***) diría que si el individuo cree α y β simultáneamente, debe dejar de creer β. Sin embargo, los autores parecen interpretarlo en términos de la competencia razonante del sujeto. Leen (*) como diciendo que si el individuo llega a creer una contradicción siempre podrá resolverla en el siguiente estadio. Sobre el sistema de Duc, el de Agotnes y Alechina presenta algunas ventajas. Una es que el individuo solo tiene en cada momento un número finito de creencias, lo que parece adecuado pues se trata de creencias explícitas y se alivia el problema de las demandas excesivas. Por otro lado, $\diamond$ (a diferencia de $\langle F_i\rangle$) representa un solo paso de razonamiento. Además el agente podría perder algunas de sus convicciones como resultado de actos de inferencia. Nos permite, si así lo deseamos, atribuirle creencias inconsistentes y describir lo que podrá hacer en tales casos.

El sistema no ofrece una correcta representación de la creencia o el conocimiento. El agente puede tener las creencias que queramos atribuirle arbitrariamente. Nada nos impide que crea proposiciones completamente desconectadas. Al dejarnos en completa libertad para atribuirle al agente una lógica, este sistema no resuelve el problema de la omnisciencia lógica, pero ofrece un patrón general para hacerlo. Por ejemplo, ¿qué podemos predecir que el agente creerá con cierta verosimilitud en el futuro inmediato dadas sus creencias actuales? o, dicho en el modo normativo, ¿qué es razonable exigirle o atribuirle dado lo que ya cree? Eso no está resuelto. Por otro lado, como dije antes, Agotnes y Alechina presentan un marco general para representar las creencias de un sujeto y cómo varían, o deben cambiar, con el tiempo, pero aún nos faltan los detalles para implementar su estrategia. Veremos enseguida que el sistema de Jago se propone como un posible complemento al que acabamos de ver.

4.5. El sistema de Jago

En 2009 Mark Jago presentó un sistema de los que he denominado "sintácticos" para resolver la cuestión la omnisciencia lógica. El sistema comparte algunos rasgos con el de Agotnes y Alechina, al grado de ser compatible con él. El autor se propone explícitamente modelar las creencias de un sujeto de carne y hueso: "Tomaré como punto de partida el requisito de modelar adecuadamente las creencias de agentes reales con recursos acotados" (p. 132). Es plenamente consciente de las objeciones que he llamado demandas excesivas y trivialidades:

> La mayoría de los razonadores humanos tienen la habilidad de aplicar modus ponens a cualquier par de premisa $(\phi \to \psi)$ y ϕ, pero frecuentemente tienen excelentes razones para no llevar realmente a cabo la inferencia, por ejemplo, cuando las que serían las conclusiones son irrelevantes al curso ordinario de sus pensamientos, o bien porque las facultades cognitivas del agente están ocupadas en otras tareas. (Ibid.)

Como hemos visto, otros sistemas de tipo sintáctico habían sido hasta entonces creados con el mismo fin pero, de acuerdo a Jago, no resultaban suficientemente interesantes desde el punto de vista de la filosofía (o aún de la computación), o al menos sus autores no habían mostrado su pertinencia a este respecto. Jago se propuso subsanar esta carencia y modelar con su lógica la manera en que las creencias de un individuo evolucionan por la aplicación paso a paso de reglas de inferencia. La inferencia será representada como un proceso no determinístico, en el sentido en que el sujeto podrá emplear cualquiera de las reglas de inferencia que tiene a su disposición a cualquier conjunto de creencias suyas a las que dicha regla pueda aplicarse, sea para producir creencias nuevas, sea, al contrario, para eliminar algunas de las que ya tenía. En cambio, supondrá que el conjunto de reglas de inferencia no se altera con el paso del tiempo. No hace Jago esta restricción para simplificar el sistema sino porque piensa en casos en que el número de reglas es demasiado grande como para que el agente tuviese tiempo de revisarlas, o bien, en que las reglas han sido generadas por el consenso de un grupo de expertos al que nuestro sujeto no pertenece. Las reglas, en general, podrán ser del tipo $\pi_1\pi_2\ldots\pi_n\Upsilon\gamma_1\gamma_2\ldots\gamma_m$ donde $\pi_1\pi_2\ldots\pi_n$ son premisas y $\gamma_1\gamma_2\ldots\gamma_m$ forman un conjunto de conclusiones que resultan de aplicar Υ a esas premisas. En realidad, Jago solo considera reglas con una sola conlusión, pero de dos tipos, que llamaré respectivamente aumentativas y diminutivas. Las primera serán denotadas por un supríndice '+'; la segundas por un superíndice '-'. Una regla aumentativa de la forma $\pi_1\pi_2\ldots\pi_n\Upsilon^+\gamma$ permite al individuo agregar γ a su conjunto de creencias, si las premisas de la regla figuraban en ese conjunto, pero no la conclusión. En cambio, una regla del tipo diminutivo $\pi_1\pi_2\ldots\pi_m\Upsilon^-\gamma$, autoriza al sujeto a eliminar γ de su sistema de creencias si antes creía $\pi_1\pi_2\ldots\pi_m$ y también γ. El sistema no representará la heurística (o las estrategias) que el individuo en cuestión sigue realmente, ni las que debiera seguir, para alcanzar una determinada conclusión partiendo de sus creencias en un momento dado.

Jago insiste mucho en que su enfoque es enunciativo (y no proposicional), es decir que representa las creencias del agente con un conjunto de enunciados de un lenguaje determinado. A diferencia de otras lógicas epistémicas, la suya trata con una noción fina de información por lo que no podrá ser presentada por medio de mundos posibles. Cita a autores que se han pronunciado por un enfoque similar (Davidson, Loewer y Lepore, Quine y Fodor, entre otros). Es decir, una regla podrá aplicarse a un conjunto de enunciados y no a otro aunque ambos transmitan la misma información. En esta lógica la creencia no está definida (como lo hacen Hintikka o Baltag y Smets), sino que es primitiva y no puede ser especificada a través de una semántica kripkeana. Lo que el sistema realmente modela es la evolución paso a paso de un conjunto de creencias por la aplicación de reglas de inferencia.

Como antecesores en esta vía, Jago menciona a Konolige, en la vertiente de la lógica epistémica estática. El sistema de este autor otorgaba a cada agente un sistema de reglas de deducción. Dicho sistema podía ser incompleto en cuyo caso el agente no era lógicamente omnisciente, pero sus creencias eran cerradas bajo la aplicación reiterada de las reglas de las que se le proveía. Por el lado dinámico, alude Jago a Duc, a Agotnes y Alechina. Del sistema de estos último afirma que es muy general y que es compatible con el que ahora propondrá. También alude a algunos autores que desarrollaron ideas similares en inteligencia artificial.

En el sistema de Jago habrá, por un lado, un lenguaje interno L (con su lógica propia) que es el del agente, y, por otro, el lenguaje ML con que describiremos las acciones inferenciales y estados de dicho sujeto. Sea LP un conjunto infinto de letras proposicionales. Una literal de LP es un elemento de LP o su negación. El lenguaje interno sólo contiene literales λ_i (con $i \in \mathbb{N}$) y reglas, que son enunciados de la forma $\lambda_1\lambda_2\ldots\lambda_n \rhd^+ \lambda$ o $\lambda_1\lambda_2\ldots\lambda_n \rhd^- \lambda$ (con λ_i literales). Denotaremos por $\rhd^{Pr}$ al conjunto $\{\lambda_1\lambda_2\ldots\lambda_n\}$ y por $\rhd^{Cn}$ a λ, si $\lambda_1\lambda_2\ldots\lambda_n \rhd \lambda$ es una regla, sea aumentativa o diminutiva. El lenguaje externo, o metalenguaje, ML, será construido a partir de los enunciados de L. Más precisamente:

Definición 128. Los enunciados atómicos de ML son todos los enunciados de las formas $\mathcal{B}\lambda$ y $\mathcal{B}\varrho$ donde λ es una literal y ϱ una regla ($\lambda \in L$ y $\varrho \in L$)

Los enunciados de ML se forman recursivamente, de la manera usual, a partir de los atómicos y a través de los operadores proposicionales $\neg$, $\wedge$ y del operador modal $\Diamond$.

Definición 129. Un modelo es una terna $M = \langle S, T, V\rangle$, donde S es un conjunto no vacío cuyos elementos serán llamados estados, T es una relación binaria de elementos de S (es decir, $T \subseteq S^2$) y V una asignación a cada estado de un conjunto de literales y reglas del lenguaje interno L (es decir, $V : S \Rightarrow 2^L$). A V se le llama función de etiquetado. Si wTv diremos que v es sucesor de w. De los elementos de $V(s)$ diremos que etiquetan s. A $V(s)$ lo llamaremos el etiquetado de s.

Como era de esperarse, la función de etiquetado indica qué literales el agente cree y de qué reglas dispone en cada estado. La relación T conectará una estado con otro cuando el individuo pueda pasar de uno a otro aplicando a lo más una vez una regla de inferencia. Esto quedará claro más adelante.

Definición 130. Dados dos modelos $M = \langle S, T, V\rangle$ y $M' = \langle S', T', V'\rangle$y dos estados $s \in S$ y $s' \in S'$ diremos que están idénticamente etiquetados si $V(s) = V'(s')$ y lo denotaremos por $s \leftrightsquigarrow L \leftrightsquigarrow s'$.

Definición 131. Las siguientes cláusulas definen qué significa que un enuciado α de ML sea verdadero en un estado s de un modelo $M = \langle S, T, V\rangle$ (es decir, que $M, s \models \alpha$:

a) $M, s \models \mathcal{B}\gamma$ si y solo si $\gamma \in V(s)$

b) $M, s \models \neg\alpha$ si y sólo si $M, s \not\models \alpha$

c) $M, s \models \alpha \wedge \beta$ si y sólo si $M, s \models \alpha$ y $M, s \models \beta$

d) $M, s \models \Diamond\alpha$ si y sólo si existe $v \in S$ tal que sTv y $M, v \models \alpha$.

Definición 132. Dados dos modelos $M = \langle S, T, V\rangle$ y $M' = \langle S', T', V'\rangle$ y dos estados $s \in S$ y $s' \in S'$ diremos que s y s' son modalmente equivalentes ($s \leftrightsquigarrow s'$) si y solo si para cada fórmula α de ML, $M, s \models \phi$ si y solo si $M', s' \models \phi$.

Hasta aquí el lenguaje y la semántica difiere poco de la del sistema de Agotnes y Alechina, salvo por diferencias notacionales y porque ahora las fórmulas atómicas pueden ser formadas también a partir de reglas. La semántica no es aún adecuada para representar relaciones inferenciales entre estados. Para adaptarla a este fin serán necesarias algunas otras cláusulas y algunas definiciones. Que una regla ϱ sea compatible con un estado s significará que es susceptible de aplicarse provechosamente a las fórmulas que etiquetan s. Esto significa, si la regla es aumentativa, que en s se hallan las premisa de ϱ, pero no la conclusión. En cambio, si es diminutiva, significará que en s se encuentra tanto las premisas como la conclusión de ϱ.

Definición 133. Dado un estado s de un modelo y una regla ϱ, decimos que ϱ es s-compatible con s si $\varrho^{Pr} \subseteq V(s)$ pero $\varrho^{Cn} \notin V(s)$, si ϱ es aumentativa; o bien que $\varrho^{Pr} \cup \{\varrho^{Cn}\} \subseteq V(s)$, si ϱ es diminutiva.

Recordemos que en los modelos que serán pertinentes para los fines de Jago todos los mundos tendrán las mismas reglas. Si no es posible aplicar ninguna de esas reglas a un estado s diremos que este es terminal.

Definición 134. Un estado s de un modelo M es terminal si ninguna regla de M es s-compatible.

Ahora definimos los modelos $\mathbb{M}^+$, $\mathbb{M}^-$ y $\mathbb{M}$ que nos permitirán modelar las transiciones inferenciales. Para ello, ϱ^+ y ϱ^- representarán respectivamente una regla aumentativa y una diminutiva.

Definición 135. Dado un modelo $M = \langle S, T, V \rangle$ diremos que es un $\mathbb{M}$-modelo o que pertenece a la clase $\mathbb{M}$ si y solo si:

1) Si $\varrho^+ \in V(s)$ y ϱ es s-compatible, existe $u \in S$ tal que sTu y $V(u) = V(s) \cup \{\varrho^{Cn}\}$.

1') Si $\varrho^- \in V(s)$ y ϱ es s-compatible, existe $u \in S$ tal que sTu y $V(u) = V(s) - \{\varrho^{Cn}\}$.

2) Si $s \in S$ es terminal entonces existe $v \in S$ tal que $s \leftrightsquigarrow L \leftrightsquigarrow v$ y sTv.

3) Si $s, v \in S$ y sTv entonces, a) o hay una regla aumentativa $\varrho^+ \in V(s)$ s-compatible y $V(v) = V(s) \cup \{\varrho^{Cn}\}$, b) o hay una regla diminutiva $\varrho^- \in V(s)$ s-compatible y $V(v) = V(s) - \{\varrho^{Cn}\}$, o bien, c) s es terminal y $s \leftrightsquigarrow L \leftrightsquigarrow v$.

4) Dado un modelo $M = \langle S, T, V \rangle$ y cualquier regla $\varrho \in L$ se cumple que para cualesquiera dos mundos $s, v \in S$, $\varrho \in V(s)$ si y solo si $\varrho \in V(v)$.

Una formulación equivalente habría sido substituir 2) por la condición de que si s es terminal entonces sTs y habríamos modificado 3) en consecuencia. Note también que si 4) hubiese sido formulada en primer término nos habríamos evitado las condiciones "si $\varrho \in V(s)$", pero queda así claro que Jago podría haber incluido solo los tres primeros incisos si quisiera modelar un agente que también, bajo ciertas condiciones, agregara o eliminara reglas.

Definición 136. Un $\mathbb{M}^+$-modelo es un $\mathbb{M}$ que solo tiene reglas aumentativas y, por lo tanto, para el cual no son aplicables las cláusulas 1') ni 3a).

Definición 137. Un $\mathbb{M}^-$-modelo es un $\mathbb{M}$ que solo tiene reglas diminutivas y, por lo tanto, para el cual no son aplicables las cláusulas 1) ni 3b).

Definición 138. Sea $\mathcal{R}$ un conjunto finito de reglas. Un $\mathbb{M}_{\mathcal{R}}$-modelo es un $\mathbb{M}$ modelo tales que las reglas de cada mundo son exactamente los elementos de $\mathcal{R}$. Similarmente definimos los $\mathbb{M}^+_{\mathcal{R}}$ modelos y los $\mathbb{M}^-_{\mathcal{R}}$ modelos; en el primer caso, todas las reglas de $\mathcal{R}$ son aumentativas; en el segundo, diminutivas.

Definición 139. Si $\Gamma \cup \{\phi\}$ es un conjunto de fórmulas de ML diremos que ϕ es $\mathbb{M}$-consecuencia lógica de Γ ($\Gamma \models_{\mathbb{M}} \phi$) si y solo si para cada $\mathbb{M}$-modelo M, si $M \models \alpha$ para cada $\alpha \in \Gamma$, entonces $M \models \phi$.

Definiciones análogas pueden proveerse para $\mathbb{M}^+$ consecuencia lógica y $\mathbb{M}^-$ consecuencia lógica. Análogamente tomando solo en cuenta los $\mathbb{M}_\mathcal{R}$-modelos, $\mathbb{M}_\mathcal{R}^+$ modelos o los $\mathbb{M}_\mathcal{R}^+$ modelos, definimos respectivamente la $\mathbb{M}_\mathcal{R}$-consecuencia, $\mathbb{M}_\mathcal{R}^+$-consecuencia y la $\mathbb{M}_\mathcal{R}^-$ consecuencia lógica.

Con un ejemplo ilustraré una peculiaridad del sistema de Jago. Supongamos que a cualquier paciente que llega al hospital con fiebre, se le aplican las pruebas de diagnóstico rápido (PDR) para determinar si tiene malaria pues la incidencia de la enfermedad es relativamente alta en esa zona. A la vez se le realizan ciertas pruebas de laboratorio. Si las (PDR) resultan positivas y los resultados de laboratorio encuentran Pasmodium-Falciparum se le administra doxiciclina o bien arteméter. En cambio, si las PDR son positivas y se halla Pasmodium Vivax se le prescriben citoquinas. Si advierten que sea le han prescrito a la vez el arteméter y la doxicicliclina, se recomienda dejarle solo la primera. Entre las reglas del protocolo se encuentran las siguientes:

R1) Pedro-Fiebre, Pedro-PDRPositivo $\rhd^+$ Pedro-Malaria,

R2) Pedro-Malaria, Pasmodium-Falciparum $\rhd^+$ Pedro-Doxiciclina

R3) Pedro-Malaria, Pasmodium-Falciparum $\rhd^+$ Pedro-Arteméter

R4)) Pedro-Malaria, Pasmosium-Vivax $\rhd^+$ Pedro-Citoquinas

R5) Pedro-Arteméter, Pedro-Doxiciclina$\rhd^-$Pedro-Arteméter

Supongamos que Pedro llegó con fiebre, sus pruebas de diagnóstico rápido dieron positivo y su examen de laboratorio halló Pasmodium Falciparum en su sangre. Entonces la base de datos inicial es:

S_1) Pedro-Fiebre, Pedro-PDRPositivo, Pasmodium-Falciparum.

La única regla compatible con S_1 es $R1$. Habrá entonces una situación S_2 tal que S_1TS_2:

S_2) Pedro-Fiebre, Pedro-PDRPositivo, Pasmodium-Falciparum, Pedro-Malaria

Ahora R_1, R_2 y R_3 son compatibles con S_2. Dejando de lado el primer caso, habrá dos estados S_3 y S_4 tales que S_2TS_3 y S_2TS_4:

S_3) Pedro-Fiebre, Pedro-PDRPositivo, Pasmodium-Falciparum, Pedro-Malaria, Pedro-Doxiciclina.

S_4) Pedro-Fiebre, Pedro-PDRPositivo, Pasmodium-Falciparum, Pedro-Malaria, Pedro-Arteméter.

Ahora $R1$, $R2$ y $R3$ son compatibles con S_3 y con S_4. Tomemos la tercera ruta (aplicar R3 a S_3 para llegar a S_5):

S_5) Pedro-Fiebre, Pedro-PDRPositivo, Pasmodium-Falciparum, Pedro-Malaria, Pedro-Doxiciclina, Pedro-Arteméter.

$R5$ es compatible con $S5$, Por lo tanto habrá un estado $S6$ tal que S_5TS_6:

S_6) Pedro-Fiebre, Pedro-PDRPositivo, Pasmodium-Falciparum, Pedro-Malaria, Pedro-Doxiciclina.

Este ejemplo ilustra las inferencias que puede hacer el agente. Los enunciados anteriores no perteneces a ML. El sistema no es determinístico porque pudimos seguir otras rutas. Vemos también que las reglas en el sistema de Jago no son esquemas, sino que conectan algunas proposiciones atómicas (las premisas) con una proposición también específica (la conclusión). Jago demuestra enseguida algunas propiedades de este sistema. Me concentraré en una sola que es una peculiaridad del sistema.

Teorema 140. *Sean $M = \langle W, R, V\rangle$ un $\mathbb{M}^+$-modelo y $w, u, v \in W$ tales que u y v son accesibles desde w, entonces existen mundos $r, s \in W$ tales que r y s son accesibles desde u y v respectivamente tales que $r \leftrightsquigarrow s$.*

Demostración. Si v es accesible desde u y z es sucesor de v (es decir, vTz) entonces trivialmente el teorema se cumple si $r = s = z$. El caso en que u es accesible desde v es análogo. Supongamos que hay dos reglas en M, ϱ_1 y ϱ_2, ambas compatibles con w y dos mundos $u \in W$ y $v \in W$ tales que wTu, wTv, $V(u) = V(w) \cup \{\varrho_1^{Cn}\}$ $V(v) = V(w) \cup \{\varrho_2^{Cn}\}$. Dado que las premisas de ambas reglas se conservan tanto en u como v es posible ahora aplicar ϱ_1 a v y ϱ_2 a u. Como M es un $\mathbb{M}^+$-modelo, existirán mundos $r, s \in W$, tales que uTr, vTs y $V(r) = V(s) = V(w) \cup \{\varrho_1^{Cn}\} \cup \{\varrho_2^{Cn}\}$. El caso más general se prueba de manera similar. $\square$

Es decir, que si el agente solo sigue reglas aumentativas, entonces si partiendo de una situación dadas y siguiendo dos cursos de acción puede llegar a situaciones distintas, entonces partiendo a su vez de ellas puede en algún tiempo arribar a una misma situación (o al menos dos situaciones que para efectos prácticos son idénticas).

Este teorema también es válido para los $\mathbb{M}^-$-modelos. No lo es, en cambio, para un $\mathbb{M}$-modelo pues si, por ejemplo $V(w) = \{p, q, \varrho_1^+, \varrho_2^-\}$ donde $\varrho_2^+ = p, q \rhd r$ y $\varrho_2^- = p \rhd q$, entonces aplicando sucesivamente a w la regla aumentativa seguida de la diminutiva accedemos a un estado z tal que $V(z) = \{p, r, \varrho_1^+, \varrho_2^-\}$, mientras que aplicando a w la regla diminutiva llegamos a un mundo s tal que $V(s) = \{p, \varrho_1^+, \varrho_2^-\}$. z y s son terminales pero no son modalmente equivalentes.

Enseguida Jago analiza las propiedades de modelos en que el número de reglas es finito. El estudio parece muy pertinente si queremos aplicar el sistema a sujetos reales. Sin embargo, no es así. Recordemos que las reglas de su sistema se refieren a enunciados atómicos particulares. Cualquier ser cuyas habilidades inferenciales puedan ser modeladas con un esquema de inferencia rebasa los casos que Jago estudia en la mencionada sección.

Veamos ahora cómo caracteriza axiomáticamente el conjunto de enunciados que son válidos en todos los $\mathbb{M}_\mathcal{R}$ modelos (donde $\mathcal{R}$ es un conjunto dado, finito, de reglas). Para ello, utilizaré las siguientes convenciones. Sean R^+ y R^- respectivamente los subconjuntos de reglas aumentativas y diminutivas de R. Los símbolos $\rhd^+$ y $\rhd -$ indicarán que la regla a las que me estoy refiriendo es una regla aumentativa o una diminutiva respectivamente.
Si $\varrho^+ = \lambda_1\lambda_2 \ldots \lambda_n \rhd^+ \lambda$, las expresiones $[\varrho^+]^{Comp}$ y $[\lambda_1\lambda_2, \ldots \lambda_n \rhd^+ \lambda]^{Comp}$ abreviarán la expresión:

$$\mathcal{B}\lambda_1 \wedge \mathcal{B}\lambda_2 \wedge \ldots \wedge \mathcal{B}\lambda_n \wedge \neg\mathcal{B}\lambda$$

y si $\varrho^- = \lambda_1\lambda_2 \ldots \lambda_n \rhd^- \lambda$, las expresiones $[\varrho^-]^{Comp}$ y $[\lambda_1\lambda_2, \ldots \lambda_n \rhd^- \lambda]^{Comp}$ abreviarán la expresión:

$$\mathcal{B}\lambda_1 \wedge \mathcal{B}\lambda_2 \wedge \ldots \wedge \mathcal{B}\lambda_n \wedge \mathcal{B}\lambda$$

Los siguientes son los (esquemas de) axiomas del sistema $\Lambda\mathbb{M}_\mathcal{R}$:

1) Todas las instancias de tautologías clásicas.

2) $\Box(\alpha \to \beta) \to (\Box\alpha \to \Box\beta)$

3) $\mathcal{B}\varrho$ si $\varrho \in \mathcal{R}$

4) $\neg\mathcal{B}\varrho$ si $\varrho \notin \mathcal{R}$

5) $\Diamond T$

6+) $[\varrho^+]^{Comp} \wedge \Diamond \mathcal{B}\varrho^{Cn}$

6-) $[\varrho^-]^{Comp} \wedge \Diamond \neg \mathcal{B}\varrho^{Cn}$

7+) $\Diamond(\mathcal{B}\lambda_1 \wedge \mathcal{B}\lambda_2) \rightarrow (\mathcal{B}\lambda_1 \vee \mathcal{B}\lambda_2)$

7-) $\Diamond(\neg\mathcal{B}\lambda_1 \wedge \neg\mathcal{B}\lambda_2) \rightarrow (\neg\mathcal{B}\lambda_1 \vee \neg\mathcal{B}\lambda_2)$

7) $\Diamond(\mathcal{B}\lambda_1 \wedge \neg\mathcal{B}\lambda_2) \rightarrow (\mathcal{B}\lambda_1 \vee \neg\mathcal{B}\lambda_2)$

8+) $\Diamond\mathcal{B}\lambda \rightarrow (\mathcal{B}\lambda \vee \bigvee_{\varrho^+\in\mathbb{R}}[\varrho^+]^{Comp} \wedge \varrho^{Cn} = \lambda)$

8-) $\Diamond\neg\mathcal{B}\lambda \rightarrow (\neg\mathcal{B}\lambda \vee \bigvee_{\varrho^-\in\mathbb{R}}[\varrho^-]^{Comp} \wedge \varrho^{Cn} = \lambda)$

9) $[\varrho_1^+]^{comp} \wedge \ldots \wedge [\varrho_n^+]^{Comp} \wedge [\varrho_1^-]^{comp} \wedge \ldots \wedge [\varrho_n^-]^{Comp} \wedge$

$\bigwedge_{\varrho\in\mathcal{R}-\{\varrho^+_\infty,\ldots\varrho^+_\setminus,\varrho^-_\infty,\ldots\varrho^+_\Updownarrow\}} \neg[\varrho]^{Comp} \rightarrow$

$\Box(\mathcal{B}\varrho_1^{+Cn} \vee \ldots \vee \mathcal{B}\varrho_n^{+Cn} \vee \neg\mathcal{B}\varrho_1^{-Cn} \vee \ldots \vee \neg\mathcal{B}\varrho_m^{-Cn})$

Enseguida Jago demuestra teoremas de completud y corrección del sistema formado por los axiomas 1), 2), 3), 4), 5), 6+), 7+) 8+) y 9 (restringido a las reglas aumentativas con respecto a la $\mathbb{M}^+{}_\mathcal{R}$ consecuencia. Lo mismo demuestra del sistema completo en relación a la $\mathbb{M}^+{}_\mathcal{R}$ consecuencia. Por simetría cabe suponer que el sistema formado por los axiomas 1), 2), 3), 4), 5), 6-), 7-) 8-) y 9 (restringido a las reglas diminutivas) es correcto y completo con respecto a la $\mathbb{M}^-{}_\mathcal{R}$ consecuencia.

4.5.1. Evaluación

Como el propio autor se propone, su sistema busca describir la competencia razonante de seres humanos reales. El individuo modelado por Jago puede seguir programas fijos que le indican uno o varios cursos de acción en situaciones concretas. También podríamos suponer que las reglas en cuestión son normativas, como en el ejemplo que puse, que guían a un individuo a realizar ciertas acciones para ciertos efectos. Por ejemplo, podría ser que el programa contenga las instrucciones que deba seguir el sujeto para diagnosticar a un enfermo. Con reglas bien diseñadas podrían evitarse los problemas de las demandas excesivas y de la trivialidad. En el ejemplo de arriba nada impide que el razonador quede atrapado en un bucle. Razonaría de acuerdo a las reglas, pero no llegaría al fin propuesto. Creo que también podría resolverse el problema del constreñimiento porque una regla adecuada podría indicarle al sujeto que renunciara una conclusión y, en consecuencia, eliminara una premisa, bajo ciertas condiciones (como vimos con el ejemplo de las medicinas). El autor señala asimismo algunas virtudes computacionales del sistema. Sin embargo, hay un rasgo del sistema que complica su aplicación, a saber, que las reglas solo se aplican a literales que representan proposiciones concretas incapaces de ningún tipo de combinación o composición. Supongo que Jago introdujo este rasgo para evitar el siguiente problema. Una regla puede prescribir que de P se pase a Q en un estado s. El agente en el estado s podría recibir una información que es lógica o analíticamente equivalente a P, ¿Debe pasar a Q? Por ejemplo, la regla dice que Juan puede contraer matrimonio solo si es soltero. El agente (que es juez del registro civil) sabe que Juan es casado. Para aplicar correctamente la regla debe conocer el significado de las palabra "soltero" y aplicar una sencilla inferencia lógica. Pero si estamos modelando su competencia de raciocinio la inferencia en cuestión también debe estar explícitamente codificada. Creo que a ello obedece la idea de trabajar con una noción más fina de información. Pero no se ve cómo podría formularse un sistema de reglas lógicas que solo contuviera enunciados concretos.

Dejando de lado esta cuestión, podemos otorgarle al sujeto las regla de inferencia que nos plazca, incluso alguna que indique qué hacer en caso de que al seguir una regla infiera la negación de una proposición que tiene en su base de datos. Si es un dialetista puede quedarse con la contradicción.

Como en el caso de otros tratamientos sintácticos, el de Jago adolece de una adecuada representación de las nociones semánticas cuyo funcionamiento busca codificar. Ambas nociones, conocimiento y creencia, se representan aquí con la metáfora muy simple de un cajón de enunciados que no necesariamente tiene una estructura interna. El sistema más bien representa la disminución o incremento del conjunto de enunciados como resultado de un paso inferencial.

4.6. El sistema de Rasmussen

En un artículo de 2015, Mattias Skipper Rasmussen afronta de nuevo el problema de la omnisciencia lógica desde un punto de vista dinámico retomando algunas de las ideas que Duc sugiere, pero no desarrolla, en el texto que antes expuse. Explica que la lógica epistémica clásica representa las creencias implícitas del agente, es decir, las consecuencias lógicas de las creencias que el agente sostiene explícitamente, o bien, la forma en que el mundo sería según lo que el agente dice creer. Su objetivo es modelar las creencias explícitas, aquellas que el agente defiende y que guían sus decisiones. Parte de la suposición de que el ser humano razona siguiendo reglas de inferencia, dejando de lado si esto realmente es el caso. Se propone entonces desarrollar un sistema que represente las creencias explícitas de un individuo satisfaciendo las siguientes tres condiciones:

1) El conocimiento del agente no debe ser cerrado bajo ninguna regla de inferencia no trivial[2] .

2) "No debemos sacrificar el conocimiento lógico en el altar de la no omnisciencia lógica", es decir, el agente no debe ser desprovisto de toda capacidad inferencial.

3) Debemos definir el conocimiento como verdad en todos los mundos que el individuo concibe como posibles.

El punto número uno alude a un argumento expuesto por Jago y por Bjerring cuya conclusión es que ningún sistema podr'ia distinguir entre inferencias sencillas y complejas de tal manera que el conocimiento del individuo representado fuese cerrado sólo con respecto a las primeras. En la exposición del sistema de Rasmussen y Bjerring, más adelante, veremos con mayor detenimiento este argumento. Baste por ahora decir que se basa en la observación de que toda inferencia compleja puede descomponerse en última instancia en una cadena de inferencias simples. Sin embargo, hay un sentido en el que el sistema de Rasmussen podría servirnos para hacer la división entre inferencias sencillas y difíciles, como más adelante diré.

El segundo punto coincide con el *desideratum* de Duc de que el agente pueda acrecentar sus conocimientos por el uso de algunas reglas de inferencia. Como vimos, parece haber una dificultad para satisfacer 1) y 2) simultáneamente desde una perspectiva estática.

La última condición obedece a la alta estima que Rasmussen tiene de la representación del conocimiento a la manera de Hintikka. Piensa que si el enfoque en términos de mundos posibles ha sido tan exitoso en tantas disciplinas no debería descartarse al resolver el problema de la omnisciecia lógica. Yo coincido con él en que una buena solución al problema no debe ser simplemente un instrumento que se ajuste a nuestras adscripciones de creencia (o de conocimiento), sino que debe ofrecer un análisis de los conceptos epistémicos. Sin embargo, pudiera ser que alguna otra manera de definir la creencia, por ejemplo, la de Baltag y

[2]Por supueso que esta es una idea intuitiva que se busca caracterizar.

Smets, fuese superior a la de Hintikka. En todo caso, esto le sirve a Rasmussen para desechar los enfoques sintácticos como esencialmente errados. Sin embargo, la tercera condición es la única que no puede aquí satisfacer y que deja para un desarrollo posterior.

Rasmussen utiliza las condiciones 1) y 2) para evaluar y recusar algunos intentos previos (no sintácticos) de tratar con la omnisciencia lógica. Los clasifica en tres grandes categorías según que propongan modificar la noción de verdad, la noción de mundo o la de conocimiento. En la primera especie entran, por ejemplo, el sistema de Levesque y sus variantes que reemplazan la lógica clásica con la relevante. Allí las proposiciones pueden adquirir cuatro valores de verdad y, en ese sentido, no operan con la noción clásica de la verdad. No satisfacen el primer criterio porque su agente es lógicamente omnisciente desde el punto de vista de una lógica no-clásica. En la segunda clase se hallan los sistemas que postulan mundos imposibles, como el de Rantala. Si esos mundos son completamente arbitrarios no se cumple el segundo criterio. En cambio, si en aras de una mayor verosimilitud se les restringe con alguna cláusula sintáctica (por ejemplo, que siempre que otorguen la verdad a dos proposiciones también hagan verdadera a su conjunción) entonces, evidentemente no satisfarán el primer criterio. Por último, ejemplifica la tercera categoría, aquella que modifca la noción de conocimiento, con la lógica de la consciencia porque define el conocimiento explícito como verdad en todos los mundos posibles de la que el agente es consciente. De nuevo, pueden establecerse, o no, restricciones a la función de consciencia, lo que hará que no se satisfaga una u otra de las dos primeras cláusulas que estableció Rasmussen. En su opinión, el defecto de estos enfoques radica en su naturaleza estática "porque es la habilidad del agente para llevar a cabo un razonamiento lógico no trivial lo que lo hace a la vez lógicamente no omnisciente y lógicamente no ignorante". Se propone formular un sistema que permita codificar las relaciones entre el conocimiento que el agente tiene en un momento y su capacidad de argumentar. Menciona entre quienes lo han precedido en este camino a Konolige, a Duc y a Agotnes y Alechina. Sin embargo, su designio, que no conseguirá en este artículo, es resolver el problema satisfaciendo también el tercer requisito impuesto, es decir, caracterizar el conocimiento, no desde un enfoque sintáctico, sino como verdad en todos los mundos concebibles.

La idea central es similar a la de Duc, a saber, introducir operadores para aplicaciones de reglas de inferencia, pero con algunas ventajas. Aunque el sistema será presentado con esquemas generales, podrá ser aplicado a casos en que haya un operador para cada regla particular que el agente es capaz de emplear. Además el leguaje permitirá llevar un registro de la sucesión de operaciones inferenciales que un individuo lleva a cabo en una ampliación concreta de su conocimiento, y además del costo cognitivo de cada una de ellas. Así podremos decir que si un agente sabe tal y cual cosa, llegará a saber tal otra si realiza tales y cuales operaciones de inferencia en un orden determinado y con un costo cognitivo tal. Descartando esta última cláusula la idea estaba ya expuesta en la primera parte del artículo de Duc, solo que en su desarrollo posterior sufrió una simplificación en la que se eliminaron factores valiosos.

El lenguaje del sistema, además de los usuales símbolos de la lógica epistémica clásica, contendrá un número finito de operadores unarios de la forma $\langle\Upsilon_i\rangle^{\lambda_i}$ (y sus correspondientes duales $[\Upsilon_i]^{\lambda_i}$) en donde Υ_i es una regla de inferencia específica y λ_i un número. Para facilitar la exposición supondremos que hay un solo agente. Una fórmula de la forma $\langle\Upsilon_i\rangle^{\lambda_i}\alpha$ representará el hecho de que, después de una aplicación de la regla Υ_i que hizo el agente, con un esfuerzo cognitivo λ_i, la fórmula α es verdadera. Una aplicación se distinguirá de otra por sus premisas y su conclusión. Por ejemplo, si tenemos la regla de eliminación de la conjunción, el paso de "hoy es lunes y llueve" a "hoy es lunes" es una aplicación distinta del tránsito de "la nieve es blanca y está nublado" a "la nieve es blanca". $[\Upsilon_i]^{\lambda_i}\alpha$ significará que después de cualquier aplicación de la regla Υ_i por el agente la proposición α es verdadera.

Para presentar su sistema axiomáticamente Rasmussen emplea algunas abreviaturas. Una expresión de la forma $\langle\ddagger\rangle^n$ representará una sucesión cualquiera de operadores diamante, de tal manera que si $\langle\ddagger\rangle^n$ representa la secuencia

$$\langle\Upsilon_{i_1}\rangle^{\lambda_1}\langle\Upsilon_{i_2}\rangle^{\lambda_2}\ldots\langle\Upsilon_{i_m}\rangle^{\lambda_m} \text{ entonces } n = \lambda_1 + \lambda_2 + \ldots\lambda_m.$$

Similarmente $[\ddagger]^n$ simbolizará una sucesión de operadores $[\Upsilon_i]^{\lambda_i}$ y n la suma de sus costos cognitivos. Por supuesto, estas sucesiones pueden ser vacías o incluir operadores repetidos. Asimismo utiliza $\langle\dagger\rangle^n$ para la figuración de otra secuencia. Los axiomas son los siguientes:

a) Todas las instancias de substitución de tautologías

b) $\langle\ddagger\rangle^n\mathcal{K}\alpha \rightarrow \alpha$

c) $\langle\ddagger\rangle^n\mathcal{K}\alpha \rightarrow \langle\ddagger\rangle^n[\dagger]^m\mathcal{K}\alpha$

d) $\langle\ddagger\rangle^n\alpha \wedge \langle\dagger\rangle^m\beta \rightarrow \langle\ddagger\rangle^n\langle\dagger\rangle^m(\alpha \wedge \beta)$

e) $\langle\ddagger\rangle^n(\alpha \wedge \beta) \rightarrow \langle\ddagger\rangle^n\alpha.$

La única regla de inferencia es *modus ponens*.
El segundo axioma obedece a la concepción de que el conocimiento siempre es de algo que es verdadero, sea que el agente lo obtenga después de un proceso inferencial o no (recuerde que $\langle\ddagger\rangle$ puede representar la sucesión vacía). El tercero representa una forma de monotonía. Suponemos, como en el caso de Duc, que una proposición inferida no deja de ser verdadera mientras se realiza otro proceso de razonamiento. d) y e) son intuitivamente claros, pero debe observarse que no implican ninguna forma de cerradura epistémica bajo alguna regla de inferencia. Note, por ejemplo, que $\mathcal{K}\alpha \wedge \mathcal{K}\beta \rightarrow \mathcal{K}(\alpha \wedge \beta)$ no es una instancia de substitución de d). Tendremos como teorema

$$\langle\ddagger\rangle^n\mathcal{K}\alpha \rightarrow (\langle\dagger\rangle^m\mathcal{K}(\alpha \rightarrow \beta) \rightarrow \langle\ddagger\rangle^n\langle\dagger\rangle^m(\mathcal{K}\alpha \wedge \mathcal{K}(\alpha \rightarrow \beta))$$

pero nada que nos asegure que el individuo sabe β. Ahora bien, a estos axiomas generales pueden agregarse otros más específicos para modelar la capacidad de un agente particular para inferir de acuerdo a determinada regla de inferencia que queramos atribuirle. Si el individuo tiene a su disposición el *modus ponens* entonces podrá concluir β si sabe α y $(\alpha \rightarrow \beta)$, pero solo si realiza la inferencia pagando su respectivo costo cognitivo. Por ejemplo, si quisiéramos atribuirle al agente la regla de inferencia conocida como contrapositiva (y su costo cognitivo es de 5 unidades) podríamos agregar al lenguage el operador $\langle CP\rangle$ y el esquema axiomático:

$$\langle\ddagger\rangle^i\mathcal{K}\alpha \wedge \langle\dagger\rangle^j\mathcal{K}(\neg\beta \rightarrow \neg\alpha) \rightarrow \langle\ddagger\rangle^i\langle\dagger\rangle^j\langle CP\rangle^5(\mathcal{K}\alpha \wedge \mathcal{K}(\neg\beta \rightarrow \neg\alpha) \wedge \mathcal{K}\beta)$$

Podemos formular un esquema axiomático para una regla de inferencia R_i de la siguiente manera. Si $\alpha_1, \alpha_2, \ldots \alpha_n R_i \beta$, es decir, si la aplicación de R_i al conjunto de premisas $\alpha_1, \alpha_2, \ldots \alpha_n$ produce la fórmula β y el costo cognitivo de R_i es n entonces el siguiente puede ser un esquema axiomático:

$$\langle\ddagger\rangle^\lambda(\gamma \wedge \mathcal{K}\alpha_1 \wedge \mathcal{K}\alpha_2 \wedge \ldots \wedge \mathcal{K}\alpha_m) \rightarrow \langle\ddagger\rangle^\lambda\langle R_i\rangle^n(\gamma \wedge \mathcal{K}\alpha_1 \wedge \mathcal{K}\alpha_2 \wedge \ldots \wedge \mathcal{K}\alpha_m \wedge \mathcal{K}\beta)$$

donde γ es cualquier fórmula. Dado que los operadores $\langle R_i\rangle$ se distribuyen sobre la conjunción tendremos como teorema:

$$\langle\ddagger\rangle^\lambda(\gamma \wedge \mathcal{K}\alpha_1 \wedge \mathcal{K}\alpha_2 \wedge \ldots \wedge \mathcal{K}\alpha_m) \rightarrow \langle\ddagger\rangle^\lambda\langle R_i\rangle^n\mathcal{K}\beta$$

Supongamos, por ejemplo que, además de CP, introducimos el siguiente axioma:

AD $\langle\ddagger\rangle^\lambda\mathcal{K}\neg\alpha \rightarrow \langle\ddagger\rangle^\lambda\langle AD\rangle^8(\mathcal{K}\neg\alpha \wedge \mathcal{K}(\alpha \rightarrow \gamma))$

Tendríamos entonces como teoremas:

(CP') $\langle\ddagger\rangle^{i}\mathcal{K}\alpha \wedge \langle\dagger\rangle^{j}\mathcal{K}(\neg\beta \rightarrow \neg\alpha) \rightarrow \langle\ddagger\rangle^{i}\langle\dagger\rangle^{j}\langle CP\rangle^{5}\mathcal{K}\beta$

(AD') $\langle\ddagger\rangle^{\lambda}\mathcal{K}\neg\alpha \rightarrow \langle\ddagger\rangle^{\lambda}\langle AD\rangle^{8}\mathcal{K}(\alpha \rightarrow \gamma)$

En este sistema axiomático específico que podríamos denominar $L_{D}^{CP,AD}$ siguiendo a Rasmussen, tenemos como teorema a la fórmula:

$$\mathcal{K}\alpha \rightarrow (\langle\ddagger\rangle^{\lambda}\mathcal{K}(\neg\neg\beta \rightarrow \neg\alpha) \rightarrow (\langle\ddagger\rangle^{\lambda}\langle CP\rangle^{5}\langle AD\rangle^{8}\mathcal{K}(\beta \rightarrow \gamma))$$

Sabiendo que el teorema de la deducción es válido en $L_{D}^{CP,AD}$ un esbozo de la prueba es el siguiente:

$K\alpha$	hipótesis.
$\langle\ddagger\rangle^{\lambda}\mathcal{K}(\neg\neg\beta \rightarrow \neg\alpha)$	hipótesis.
$\mathcal{K}\alpha \wedge \langle\ddagger\rangle^{\lambda}K(\neg\neg\beta \rightarrow \neg\alpha)$	por cálculo de enunciados.
$\langle\ddagger\rangle^{\lambda}(\mathcal{K}\alpha \wedge \mathcal{K}(\neg\neg\beta \rightarrow \neg\alpha))$	por el axioma d) y *modus ponens*.
$\langle\ddagger\rangle^{\lambda}\langle CP\rangle^{5}\neg\beta$	por CP' y *modus ponens*.
$\langle\ddagger\rangle^{\lambda}\langle CP\rangle^{5}\langle AD\rangle^{8}(\beta \rightarrow \gamma)$	por AD' y *modus ponens*.

Es muy evidente que si concedemos al agente la capacidad de emplear algunas reglas de inferencia no será lógicamente ignorante. Por otro lado, su conocimiento no será cerrado bajo ninguna regla de inferencia, ni aún bajo las que tiene a su disposición. Por ejemplo, incluso si tenemos un axioma que le concede el conocimiento del *modus ponens*, $(\mathcal{K}\alpha \wedge \mathcal{K}(\alpha \rightarrow \beta)) \rightarrow \mathcal{K}\beta$ no será un teorema. Habrá, en cambio, teoremas de la forma $(\mathcal{K}\alpha \wedge \mathcal{K}(\alpha \rightarrow \beta)) \rightarrow \langle MP\rangle^{\lambda}\mathcal{K}\beta$. Podemos modelar a un agente con los recursos de un sistema deductivo incompleto, como hizo Konolige, o con uno completo (con respecto a la lógica proposicional clásica), a la manera de Duc, en cuyo caso el individuo será lógicamente omnisciente, pero solo en potencia.

Evaluación

Rasmussen cumplió dos de sus objetivos al resolver el problema de la omnisciencia lógica sin despojar al agente de toda capacidad inferencial, pero no el de hacerlo a través de una semántica de mundos posibles. Su sistema es dinámico porque permite representar el efecto de ciertas acciones (inferencias) realizadas por el agente. Aunque de forma meramente simbólica, toma en cuenta el grado de dificultad que una serie de deducciones conlleva. Con la distinción, apenas esbozada, entre inferencias arduas y sencillas, afronta la objeción de las demandas excesivas y, de ese modo, el problema de la omnisciencia lógica. Recordemos cómo Field y Steinberger propusieron hacer una diferenciación similar para encarar esa objeción. De nuevo, podemos atribuir al hablante una lógica que no coincida con la nuestra, aunque solo caracterizada por reglas de inferencia.

Ahora bien, el sistema está confeccionado para describir los estados doxásticos explícitos del sujeto que resultan de acciones inferenciales y difícilmente se presta a una lectura normativa. No podríamos decir que delínea la capacidad razonante del sujeto, porque esta debe conocerse desde el principio y se traduce en las reglas de deducción que le conferimos. Tampoco se trata de las creencias implícitas del hablante. Si cree un condicional y el antecedente, creerá implícitamente el consecuente, pero el sistema no le atribuye la creencia en el consecuente hasta que haya llevado a cabo el paso deductivo correspondiente. Si decimos que quien cree y cree que si P entonces Q, cree Q si hace un *modus ponens*, de ninguna manera estamos normando la conducta o el pensamiento de nadie, sino simplemente describiendo dicha regla. Si acaso, podríamos usar la contrapositiva de este tipo de enunciados para hacer atribución de creencias. Tal vez podr'iamos decir que si el sujeto cree explícitamente que P y también que no P, pero hay una proposición que no cree, entonces no podemos atribuirle que siga el principio de explosión. Sin embargo, esta es una lectura muy rebuscada.

Más verosímil sería emplear la distinción entre inferencias sencillas y arduas, distinción que no tiene porque ser tajante, para dar una lectura normativa del siguiente tipo. El sujeto está autorizado u obligado a creer Q, si Q puede derivarse fácilmente de sus creencias por una cadena de inferencias. Sin embargo, no podríamos evitar la objeción de las trivialidades (muy probablemente el paso de Q a $(Q \vee R)$ puede agregarse a esa secuencia inferencial sin volverla compleja). Asimismo el sistema no provee, al menos no en su presentación original, reglas que permitan al sujeto desechar alguna creencia aunque sea una fácil consecuencia lógica de otras proposiciones que cree. Si fuera interpretado normativamente, en la forma que acabo de decir, la objeción del constreñimiento seguiría en pie. Por último, como señala el propio autor, el sistema no ofrece una representación adecuada de la creencia.

Veamos a continuación un sistema similar que se propone subsanar esta última deficiencia.

4.7. El Sistema de Bjerring y Rasmussen

Recientemente Bjerring y Rasmussen (2019) propusieron un sistema que pretende mejorar el que presenté en la sección anterior. Ellos enfatizan, como yo lo he hecho, que "...la suposición de la omnisciencia lógica es un problema si queremos desarrollar una teoría normativa de la creencia que sea sensible a las limitaciones cognitivas de los agentes ordinarios" (Bjerring, Rasmussen 2019: 502). Proponen un sistema dinámico que combina los dos enfoques, sintáctico y semántico, que han aparecido en uno u otro de los sistemas anteriores. En el enfoque semántico el conocimiento (o la creencia) de un agente A en un tiempo determinado es definido como verdad en todos los mundos que A concibe en ese instante. En el sintáctico, cada mundo es un conjunto de proposiciones que el agente sabe (o cree) en un momento determinado. El doble problema de la omnisciencia y de la ignorancia lógica será resuelto a través de un mismo recurso, a saber, la postulación de mundos imposibles, es decir, mundos en que las fórmulas reciben valores arbitarios, sin importar la forma en que están compuestas. Los mundos imposibles puede tener un doble papel, representar posibilidades epistémicas para un individuo y ser identificados con el conjunto de enunciados que en él son verdaderos. Sin embargo, hay una dificultad: que la relación de un modelo en un enfoque representa indistinguibilidad, mientras que, en el otro, simboliza una operación inferencial. Veremos cómo Bjerring y Rasmussen armonizan estos dos enfoques.

¿Por qué pretenden que el agente no sea lógicamente ignorante o, dicho de otro modo, que sea lógicamente competente? Su respuesta es que es por las mismas razones por las que rechazamos la omnisciencia lógica: queremos, en lo posible, considerar a un individuo real, que tiene alguna facultad, aunque limitada, de raciocinio. ¿Qué contará como competencia lógica? A este respecto, la respuesta de Bjerring y Rasmussen es un poco vaga: "Diremos que un agente es lógicamente competente cuando no pierde ninguna consecuencia sencilla[3] de lo que cree" (Bjerring y Rasmussen 2019, 502-503). Más precisamente sea Q una consecuencia sencilla de una creencia que el agente tiene. Si ahora le preguntamos si Q es es el caso, ¿responde inmediatamente que sí? En eso caso él no pierde esa consecuencia inmediata. Si esto ocurre con cada consecuencia sencilla de sus creencias entonces el agente es lógicamente competente. La prueba tiene algunas dificultades[4] pero, si las dejamos de lado por el momento, la idea central es clara. Con

[3]En realidad, los autores usan la plabra "trivial", pero como yo la he empleado para etiquetar una de las objeciones de Harman, prefiero decir "sencilla".

[4]Básicamente la dificultad es dar un criterio para separar competencia de desempeño (*performance)*

ello han tomado en cuenta la objeción de las trivialiades a la manera en que lo propusieron Field y Steinberger. El individuo puede defender explícitamente P y no haber considerado $(P \vee Q)$, pero en el momento en que le preguntamos si cree esta proposición la sometemos a su consideración. Queda abierta la cuestión de qué deba considerarse como una consecuencia sencilla. Más adelante, Rasmussen y Bjerring dirán que la aplicación de esta noción dependerá del agente cuya competencia queremos evaluar.

En el sistema al agente (o a cada agente) será otorgado un conjunto $\mathcal{R}$ de reglas de inferencia. Toda inferencia que haga será una sucesión de aplicaciones de los elementos de $\mathcal{R}$. Cada aplicación es un paso y habrá un cierto número m (que depende de los recursos cognitivos del agente) tal que si la proposición β puede ser inferida de un conjunto Γ de proposiciones en n pasos (de $\mathcal{R}$), β será una consecuencia sencilla de Γ si $n \leq m$. Qué deba incluirse en $\mathcal{R}$ y cuál deba ser la cota m no es ahora relevante. Al igual que los autores, supondré que tenemos un solo agente.

Los enfoques estáticos que emplean mundos imposibles oscilan, como ya habíamos visto, entre dos problemas extremos. Si permiten que los mundos imposibles asignen valores arbitrarios a las fórmulas entonces el agente modelado será logicamente ignorante. Si introducen la restricción de que al interior de los mundos imposibles valga una lógica no clásica, entonces el agente será omnisciente de acuerdo a esta lógica. ¿Por qué no restringirlos de tal manera que se amolden a la distinción entre consecuencias sencillas y pesadas? Por ejemplo, podría estipularse que si β es una consecuencia sencilla de α, entonces en cualquier mundo, posible o no, si α tiene el valor verdadero, también β lo tiene. Pero entonces el sujeto sería lógicamente omniscente con respecto a esta lógica “sencilla”.

Aquí es relevante el argumento de Bjerring y Jago que más arriba mencioné. La conclusión es que es imposible hacer que el conocimiento del agente sea cerrado solo bajo consecuencias lógicas sencillas. El problema es que si el individuo cree el conjunto de proposiciones Γ y de Γ se puede inferir en n pasos (de los consignados en $\mathcal{R}$ con $n < m$) α entonces α también debe ser creído por el agente, pero si, a su vez, de α se puede llegar en n pasos de inferencia a γ entonces esta proposición también debe formar parte del conjunto de creencias del individuo. Como este proceso puede iterarse indefinidamente, dicho conjunto será cerrado bajo $\mathcal{R}$ y no sólo bajo las inferencias sencillas. A este argumento lo llaman los autores “el resultado del colapso”. La única manera de evitar este problema es emplear un marco dinámico, como los usados por Duc o el propio Rasmusssen. El que el sujeto conozca una regla de inferencia no implica que su conocimiento sea cerrado bajo ella en ningún momento, sino que puede irlo ampliando a través de sucesivas aplicaciones de la misma. Ellos resuelven el problema haciendo la siguiente suposición. Por definición, si un agente lógicamente competente cree P y Q es consecuencia lógica sencilla de P, entonces cuando le preguntamos si Q es el caso, responde inmediatamente que sí, pero "dado el resultado del colapso, sabemos que no podemos modelar esta habilidad diciendo que el agente creía Q antes de que le hiciéramos la pregunta. Si lo hiciéramos tendría que creer todas las consecuencias lógicas de sus creencias" (Bjerrin y Rasmussen 2019, p. 508) Es decir, vamos a modelar la situación como si la pregunta que le hicimos hubiese inducido en él la inferencia correspondiente. Es decir, supondremos que en algunos de estos tests se produce un proceso de aprendizaje, y no de memoria.

El sistema que proponen Bjerring y Rasmussen combina, como ya dije, los dos enfoques mencionados. Por un lado, el conocimiento será verdad en todos los mundos accesibles y, por otro, habrá un mecanismo que permita ampliar un conjunto de fórmulas (lo que es un mundo imposible) por m o menos aplicaciones sucesivas de las reglas de $\mathcal{R}$. Veamos cómo unir ambos enfoques.

El lenguaje L_{BR} del sistema contiene los símbolos usuales de la lógica proposicional, el operador $\mathcal{B}$ de creencia y por cada $n \in \mathbb{N}$ un operador $\langle n \rangle$ (y su dual $[n]$). Las reglas de formación de fórmulas son las esperadas. $\langle n \rangle \alpha$ significará que después de n pasos de razonamiento lógico α es el caso. En particular, $\langle n \rangle B\alpha$ dice que después de n pasos de razonamiento lógico el agente cree α.

Definición 141. Un modelo doxástico es una terna $M = \langle W, R, V \rangle$ donde W es la unión de dos conjuntos disjuntos W^P y W^I, $R = WXW$ y V una función que otorga a cada elemento de W^P un conjunto de letras proposicionales y a cada miembro de W^I un conjunto de fórmulas de L_{BR}.

Los elementos de W^P y los de W^I son respectivamente llamado mundos posibles y mundos imposibles. La asignación de valores a fórmulas que no contienen operadores $\langle n \rangle$ (o $[n]$) procedrá como es usual, de manera composicional, en mundos posibles. En mundos imposibles, en cambio, la asignación depenerá solo de la función V. Para dar la definición de "enunciado verdadero en un mundo posible" y, en particular, para determinar cómo evaluar las fórmulas que contienen dichos operadores, es necesario dar algunas definiciones previas.

La expresión $\Gamma \vdash^n_{\mathcal{R}} \Sigma$ significará que cada una de las fórmulas de Σ pueden ser inferidas de Γ por a lo más n pasos de las reglas de inferencia que pertenecen a $\mathcal{R}$. En particular, $\Gamma \vdash^0_{\mathcal{R}} \alpha$ si y solo si $\alpha \in \Gamma$. Dicha relación es monótona, es decir que si $\Gamma \vdash^n_{\mathcal{R}} \alpha$, $k > n$ y $\Gamma \subset \Delta$, entonces $\Delta \vdash^n_{\mathcal{R}} \alpha$ y $\Gamma \vdash^k_{\mathcal{R}} \alpha$. En lo sucesivo supondré que un cierto conjunto $\mathcal{R}$ de reglas de inferencia ha sido fijado. Por ello hablaré de las reglas de inferencia sin más y eliminaré el subíndice correspondiente (excepto cuando sea necesario recordarlo). Para que el sistema funcione adecuadamente es necesario suponer un principio de comprensión según el cual, en cada modelo, dado cualquier conjunto A incompleto o inconsistente de enunciados, existe un mundo imposible w tal que $V(w) = A$. La idea central de la semática presentada por los autores es que

(*) La fórmula $\langle n \rangle \mathcal{B}\alpha$ será verdadera en un mundo w si α se sigue en al menos n pasos de las reglas de inferencia de cada mundo accesible (por la relación R) desde w[5].

Más formalmente: $w \models \langle n \rangle \mathcal{B}\alpha$ si y solo si para cada v tal que wRv, $v \vdash^n \alpha$, es decir, si y solo si el conjunto de fórmulas verdaderas en cada uno de los mundos que el agente ve puede ser expandido, de alguna manera, por n aplicaciones de reglas de inferencias, hasta obtener un conjunto que contiene a α. Estas expansiones serán representadas a su vez por mundos. Por ello, si en w $\langle n \rangle \mathcal{B}\alpha$ es verdadera y wRv entonces hay un mundo z que "está a no más de n pasos inferenciales" de v y que valida α. Veamos cómo formulan Bjerring y Rasmussen esta situación.

Definición 142. Si $w \in W^P$, entonces $w^n = w$ para todo n.
Si $w \in W^I$, entonces $w^n = \{v \in W^I : V(w) \subseteq V(v)$ y $V(w) \vdash^n_{\mathcal{R}} V(v)\}$

w^n es llamado el n-radio de w y de cada uno de sus miembros diremos que es es una n-expansión de w

Tomemos el conjunto de fórmulas $V(w)$ que valida elmundo w y apliquemos a algunos de sus elementos una vez una regla de inferencia de R. Agreguemos a $V(w)$ las fórmulas resultantes. El resultado es una 1-expansión de w (o de $V(w)$). Reiteremos el proceso y obtendremos una 2-expansión de w, etc. Si w es un mundo posible nunca saldremos de él. Si no, el principio de comprensión posibilita que el conjunto de fórmulas que genera una n-expansión sea un mundo del modelo. Una sola cláusula pudo definir ambos casos pues cada mundo posible es deductivamente cerrado. Nótese que en el caso de un mundo imposible w una n-expansión de w verifica solamente fórmulas que pueden ser obtenidas en n pasos de inferencias lógicas (desde $n = 0$). Ahora bien, para combinar los mundos que son opciones que el creyente no puede distinguir del mundo real, de los mundos como conjuntos de fórmulas que se obtienen de la expansión, Bjerring y Rasmussen introducen una idea que, parcialmente, se asemeja a la de Plaza en su lógica del anuncio público. En algún sentido la idea es evaluar una fórmula del tipo $\langle n \rangle \alpha$

[5]Es decir, *cada* mundo accesible a w tiene *una* n-expansión en que vale α.

en un mundo de un modelo, evaluando α en el mismo mundo, pero en otro modelo que se obtiene por una conveniente transformación del modelo original. La idea es reemplazar cada mundo v tal que wRv por una n-expansión de v. Para ello, los autores definen una relación $(M, w) \sim^n (M', w')$ entre modelos puntuados que se produce si, cuando el individuo está en la situación (M, w), puede con n pasos de inferencia llegar al estado (M', w'). Veamos cómo definir formalmente esta idea.

Definición 143. Sea $\mathcal{C}$: $2^{2^W} \rightarrow 2^{2^W}$ una función tal que aplicada a un conjunto de mundos $\mathcal{W}$= $\{W_1, W_2, \ldots W_n\}$ da como valor el conjunto $\mathcal{C}(\mathcal{W})$ de conjuntos de mundos que resultan de todos los modos en que exactamente un elemento puede ser seleccionado de cada $W_i \in (\mathcal{W})$. Cada elemento de $\mathcal{C}(\mathcal{W})$ es llamado una selección de $\mathcal{W}$

Por ejemplo, si $\mathcal{W}$= $\{\{a, b, c\}, \{d, e\}, \{f\}\}$, entonces
$\mathcal{C}(\mathcal{W})$= $\{\{a, d, f\}, \{a, e, f\}, \{b, d, f\}, \{b, e, f\}, \{c, d, f\}, \{c, e, f\}\}$

Definición 144. Dado un modelo $M = \langle W, R, V\rangle$ y un mundo $w \in W$, definimos una n-variante de R en w como una relación $R' \subseteq WXW$ tal que o

$$R'zv = \begin{cases} Rzv & \text{si } z \neq w \\ c & \text{si } z = w \text{ con } c \in \mathcal{C}\{\ddagger^{\setminus}|\sqsupseteq\mathcal{R}\ddagger\} \end{cases}$$

Al conjunto de n-variantes de R en w lo denotamos por R^n_w. Es decir, si tenemos un modelo $M = \langle W, R, V\rangle$ y un mundo $w \in W^I$, una n-variante de R en w es una relación binaria R de elementos de W que coincide con R en todo, excepto que si w estaba relacionado con s_1, s_2, $\ldots s_k$ por R, ahora w estará relacionada con t_1, t_2, $\ldots t_k$ donde cada t_i valida un conjunto de fórmulas que se obtuvo por n pasos de las reglas de inferencia aplicadas a los enunciados validados por s_i. En cambio, si w es un mundo posible $R^n_w = R$

Definición 145. Sean $M = \langle W, R, V\rangle$ y $M = \langle W', R', V'\rangle$ dos modelos, entonces $(M, w) \sim^n (M', w')$ si y sólo si $w = w'$, $V = V'$ y $R' \in R^n_w$

Si $(M, w) \sim^n (M', w')$ diremos que (M', w') es n-accesible desde (M, w).

Con estos elementos podemos dar la definición de “verdadera” y “falsa” para una fórmula en un mundo de un modelo.

Definición 146. Dado un modelo $M = \langle W, R, V\rangle$ y un mundo $w \in W$, que una fórmula α sea verdadera en w ($M, w \models \alpha$), está determinado por las siguientes

claúsulas:
Para cualquier mundo posible w:

1) Si $\rho \in LP$, $M, w \models \rho$ si y sólo si $\rho \in V(w)$,
2) $M, w \models \neg\alpha$ si y solo si $M, w \not\models \alpha$,
3) $M, w \models \alpha \wedge \beta$ si y solo si $M, w \models \alpha$ y $M, w \models \beta$,
4) $M, w \models \mathcal{B}\alpha$ si y solo si para todo $v \in W$, si wRv entonces $M, v \models \alpha$,
5) $M, w \models \langle n \rangle \alpha$ si y solo si para algún M', w' tal que $(M, w) \sim^n (M', w')$, $M', w' \models \alpha$.

Para cualquier mundo imposible w:

6) $M, w \models \alpha$ si y sólo si $\alpha \in V(w)$

Definición 147. Si $w \in W^P$, α es falsa en M, w si $M, w \not\models \alpha$.
Si $w \in W^P$, α es falsa en M, w si $M, w \models \neg\alpha$

Claramente en un mundo imposible una fórmula puede ser verdadera y falsa o ni verdadera ni falsa. La combinación de las cláusulas (4) y (5) formaliza las ideas antes expuestas para evaluar fórmulas del tipo $\langle n \rangle B\alpha$. Por ejemplo, supongamos que deseamos evaluar la fórmula $\langle n \rangle B\alpha$ en w y que wRv y wRz y no hay otro mundo aparte de v y z que sean accesibles desde w. Supongamos, para ilustar el procedimiento que v es un mundo posible y z uno imposible. Entonces debemos considerar todas las n-expansiones de v y z. v solo tiene una que es v mismo. Para que la fórmula original sea verdadera en w, α debe ser verdadera en v y en alguna (al menos) de las n-expansiones de z.

Si α es una fórmula proposicional $\langle n \rangle \alpha$ es verdadera en w si y solamente si α es verdadera en w. Por otro lado, es claro que $\langle n \rangle \langle n \rangle \beta$ es equivalente a $\langle 2n \rangle \beta$. En cambio, no está definido qué significado tendría una fórmula del tipo $\langle n \rangle \mathcal{B}\mathcal{B}\alpha$[6] .

Los autores prueban el teorema:

Teorema 148. *Si* $\{\alpha_1, \alpha_2, \ldots \alpha_k\} \vdash^n_{\mathbb{R}} \beta$ *entonces*
$\{\langle h_1 \rangle \mathcal{B}\alpha_1, \langle h_2 \rangle \mathcal{B}\alpha_2, \ldots \langle h_k \rangle \mathcal{B}\alpha_k\} \models \langle h + n \rangle \mathcal{B}\beta$
donde $h = h_1 + h_2 + \ldots + h_k$.

Del cual son obvios corolarios: Cor Si $\{\alpha_1, \alpha_2, \ldots \alpha_k\} \vdash^n_{\mathbb{R}} \beta$ entonces $\{\mathcal{B}\alpha_1, \mathcal{B}\alpha_2, \ldots \mathcal{B}\alpha_k\} \models \langle n \rangle \mathcal{B}\beta$

Cor Si $\vdash^n_{\mathbb{R}} \beta$ entonces $\models \langle n \rangle B\beta$

[6]Agradezco al Dr. Fernández de Castro esta observación.

4.7.1. Evaluación

El sujeto modelado por el sistema anterior no es, efecto, lógicamente omnisciente, ni siquiera de manera potencial. El último corolario parecería establecer lo contrario (si $\mathbb{R}$ contiene exactamente las reglas inferenciales que le concedemos). Sin embargo, debemos recordar que, si la cota correspondiente a nuestro sujeto es n y $n > m$, entonces $\models \langle n \rangle \mathcal{B}\beta$ no significa que β sea una de sus creencias explícitas, a menos que haya otro camino inferencial de longitud menor que n del que pueda derivarse β. De esta manera, evitamos las demandas excesivas. Con la definición de "inferencia sencilla" se elude también el problema de las trivialidades. En parte porque el sujeto solo deduce $P \vee Q$ de P cuando traemos a su atención esa proposición; y, en parte, porque la deducción de trivialidades también requiere de una cadena de pasos cuya longitud puede ir más allá de n.

El sistema contempla perfectamente que el agente pueda tener su propia lógica. Sin embargo, hay dos reservas que debo hacer a este respecto. La primera es que estamos suponiendo que la lógica del agente puede ser codificada por un sistema de prueba, suposición que desde Konolige es común encontrar en la literatura y que es admisible. La segunda, más seria, es que aunque las restricciones en el tipo de mundos imposibles que pueden emplearse permiten modelar la lógica del agente, el sistema no dice nada al respecto y, en ese sentido, es como otros que echan mano de este recurso. Sin embargo, hay una diferencia importante. En el caso de Rantala, por aludir a un sistema que antes he expuesto, es el uso de mundos imposibles lo que permite escapar a la omnisciencia lógica. No es así en este sistema en donde además de ese recurso se emplean operadores de acción y una separación entre inferencias según su grado de complejidad.

Dicha flexibilidad, por supuesto, también tiene sus ventajas. Podemos modelar la situación de un agente que cree una contradicción y podemos dotarlo de una lógica no explosiva. Con ello, podemos prescribir lo que debe de hacer en esa situación dada la lógica que tiene.

El sistema tolera muy bien una lectura normativa pues si Q es una inferencia sencilla desde P en la lógica del agente, la creencia en Q será razonable o permitida u obligatoria al agente (desde el punto de vista subjetivo, de su propia lógica). Sin embargo, esto depende de qué se pueda dar una medida razonable de la complejidad de una inferencia. A ese respecto los autores no ofrecen ninguna propuesta de solución.

¿El sistema ofrece una adecuada representación de la creencia? Creo que solo parcialmente. Es verdad que tenemos mundos posibles o imposibles para describir un estado doxástico de un sujeto. Podemos pensar que tenemos dos tipos de relaciones de accesibilidad. La relación R del modelo es la única de la primera especie. De la segunda son, para cada n, $V(w) \vdash_{\mathbb{R}}^{n} V(z)$ y se da cuando en no más de n aplicaciones de las reglas de inferencia es posible deducir de las fórmulas que valida w cada uno de los elementos de $V(z)$. Llamemos a la primera la relación doxástica y, a la otra, la (n)-deductiva. La primera puede representar el estado de creencia en una situación dada (a la manera habitual). Los mundos que son accesibles a w desde R representan alternativas que el sujeto A ignora si son reales. Supongamos que desde w solo se puede acceder (vía R) a v y z. Si α es verdadera en ambos A cree que α es el caso. La segunda es similar pues v^n y z^n son respectivamente n-expansiones de v y de z, y en ambos es verdadero β, entonces el sujeto también creerá esta proposición. Sin embargo, no se especifica cómo estos nuevos mundos se relacionan (via R) con los anteriores y sin ello no está claro cómo debemos concebirlos. Esto puede apreciarse al intentar evaluar una fórmula del tipo $\langle n\rangle\mathcal{B}\mathcal{B}\alpha$ en un mundo w. Debemos modificar el modelo cambiando cada sucesor de w por una n-expansión, y debemos evaluar en esos mundos $\mathcal{B}\alpha$. Eso supone que evaluemos α en los sucesores vía R de los mundos recién introducidos, pero no está definido cuáles son sus R-sucesores.

Aunque ignoro si estos defectos podrán subsanarse, me parece que el sistema de Bjerring y Rasmussen ofrece una esperanzadora respuesta al problema de la omnisciencia lógica.

4.8. Otra representación formal de la creencia

Desde principio de siglo varios autores han propuesto una manera alternativa de modelar la creencia. La idea es utilizar una semántica de mundos posibles ordenados según cuán plausibles los considera el agente. La creencia no será entonces verdad en todos los mundo concebibles, sino solo en aquellos más verosímiles. Hay diversas variantes de esta propuesta. Yo seguiré la de Baltag y Smets (2008) o, más precisamente, la versión que de ella ofrece Fernando Velázquez-Quesada (2014). Aunque dichos autores no pretendieron resolver el problema de la omnisciencia lógica, a partir de sus ideas, sistemas similares han aparecido en la literatura que sí afrontan de forma novedosa esta dificultad. La mayoría de ellos implementa los modelos de plausibilidad con otros artificios, los que combinados de diversas maneras dan lugar a sistemas lógicos interesantes.

4.8.1. Preámbulo. La Lógica del Anuncio Público

En lo que sigue revisaré muy brevemente la lógica del anuncio público que Plaza propuso en 1989. El sistema que se genera aún supone agentes lógicamente omniscientes. Sin embargo, mi intención ahora es introducir una forma de construir modelos semejante a otra que más adelante veremos. Este sistema ofrece una forma limitada de revisión de creencias. No siendo relevante para otros de mis objetivos, mi exposición será escueta.

Plaza pretende dar cuenta del incremento del conocimiento que se genera en una comunidad cuando un miembro da a conocer públicamente una información verdadera. Podremos también suponer que la información proviene de una fuente externa pero que es del dominio público. Por último, si solo tenemos un individuo, también podría, en principio, servir para representar cómo cambian sus creencias cuando aprende algo. En este sentido, se trata de una operación de revisión de creencias, con la peculiaridad de que produce un cambio monótono de cierto tipo, a saber, en que el individuo no pierde ninguna de las creencias no epistémicas[7] que tenía como resultado de esta operación. Hay dos restricciones importantes a este tratamiento cuando se emplea en los sistemas Hintikka. Para que esta lógica represente el fenómeno que pretende se requiere a) que lo que se anuncio sea verdadero, y b) que el agente no crea antes del anuncio la negación de aquello que se le notificará. En este último caso, la teoría predice que al aceptar el anuncio, las creencias del agente formarán un conjunto inconsistente.

Partimos de un conjunto finito Λ de agentes por cada uno de los cuales tendremos un operador $\mathcal{K}$ de conocimiento que se comporta como $\mathcal{K}$ en S_5. Por otro lado, para cada fórmula α del lenguaje habrá un operador $[\alpha!]$ y su dual $\langle\alpha!\rangle$. El lenguaje de la lógica epistémica clásica se extenderá para incluir fórmulas del tipo $[\alpha!]\beta$ y $\langle\alpha!\rangle\beta$. La primera simbolizará que después de cualquier anuncio de α, β es verdadera, mientras que la segunda representará que después de algún anuncio de α, β es verdadera. $\langle\alpha!\rangle\beta$ puede ser definida como $\neg[\alpha!]\neg\beta$ .La diferencia entre ambos operadores quedará más clara un poco más adelante. Así como una fórmula que se inicia con un operador modal es evaluada en un mundo a partir de sus valores en otros mundos, una fórmula del tipo $[\alpha!]\beta$ sera verdadera en un mundo w de un modelo M, si α es falsa en w o β verdadera en ese "mismo" mundo, pero en otro modelo que se obtiene por la supresión en M de todos los mundos en que α es falsa. Es decir, el efecto del anuncio público verdadero será la eliminación de todos los mundos en que la fórmula anunciada es falsa.

[7]Es decir que no se refieren a otras creencias o conocimientos.

Supongamos fijo el conjunto Λ de agentes. El lenguaje LPAL de la lógica PAL(Λ),está dado por la siguiente definición recursiva:

Definición 149. 1) Una letra proposicional es una fórmula
2) Si α y β son fórmulas, también lo son $\neg\alpha$, $(\alpha \vee \beta)$, $(\alpha \wedge \beta)$, $(\alpha \rightarrow \beta)$, $(\alpha \leftrightarrow \beta)$, $\langle\alpha!\rangle\beta$ y $\mathcal{B}\alpha$ (con $i \in \lambda$).

Definimos $[\alpha!]\beta$ como $\neg\langle\alpha!\rangle\neg\beta$ y $\mathcal{K}_{\rangle}\beta$ como $\neg\mathcal{B}_i\neg\beta$.
Antes de presentar la semántica del sistema es necesario definir la restricción $M \upharpoonright \alpha$ de un modelo M a una fórmula α.

Definición 150. Sea $M =< W, R_i, V >$ un modelo normal y sea α una fórmula de LPAL, entonces $M \upharpoonright \alpha =< D, R'_i, V' >$ donde:

a) $D =\parallel \alpha \parallel_M$ es decir $\parallel D \parallel_M= \{w \in W \mid w \models \alpha\}$.

b) $R'_i = R_i \cap D^2$.

c) $V = V \upharpoonright D$.

Es decir, $M \upharpoonright \alpha$ resultará de recortar el modelo M eliminando todos los mundos en que α es falsa, restringiendo las relaciones R_i a los casos en que conecta dos nodos en que α es verdadera y, por último, una valuación de una letra proposicional seleccionará de los mundos que antes eran su imagen aquellos en que α es verdadera.

Definición 151. Dado un modelo S_5, $M =< W, R_i, V >$, y un mundo $w \in W$, definimos que una fórmula α de LPAL sea verdadera en w ($M, w \models \alpha$) por medio de las siguientes cláusulas recursivas:

1) $M, w \models \rho$ donde $\rho \in LP$ si $w \in V(\rho)$

2) $M, w \models \neg\alpha$ si y solo si $M, w \not\models \alpha$

3) $M, w \models \alpha \wedge \beta$ si y solo si $M, w \models \alpha$ y$M, w \models \beta$

4) $M, w \models \mathcal{B}_i\alpha$ si y solo si para todo $v \in W$ tal que wR_iv, $M, v \models \alpha$

5) $M, w \models \langle\alpha!\rangle\beta$ si y solo si $M, w \models \alpha$ y $M \upharpoonright \alpha, w \models \beta$

Como una consecuencia tenemos la cláusula
$M, w \models [\alpha!]\beta$ si y solo si (si $M, w \models \alpha$ entonces $M \restriction \alpha, w \models \beta$)
Es decir, si el enunciado α que se anuncia es falso en el mundo real, entonces independientemente del valor de β, $\langle\alpha!\rangle\beta$ será falsa, mientras que $[\alpha!]\beta$ será verdadero por vacuidad. En cambio, si α es verdadero en w $\langle\alpha!\rangle\beta$ y $[\alpha!]\beta$ serán equivalentes y su valor será el de $M \restriction \alpha, w \models \beta$. Como siempre, la validez es verdad en todos los mundos de todos los modelos.
Como ya vimos:
Si $M, s \models \alpha$ entonces $M, s \models [\alpha!]\beta \leftrightarrow \langle\alpha!\rangle\beta$.
Por otro lado, si α es falsa, claramente $[\alpha!]\beta$ es verdadero, mientras que $\langle\alpha!\rangle\beta$ es verdadero.

Lema 152. *Las siguientes fórmulas son equivalentes:*

1) $\alpha \to [\alpha!]\beta$

2) $\alpha \to \langle\alpha\rangle\beta$

3) $[\alpha!]\beta$

Ahora bien dado que $\models [\alpha!]\beta \leftrightarrow (\alpha \to \langle\alpha!\rangle\beta)$ tenemos que
$\models [\alpha!]\neg\beta \leftrightarrow (\alpha \to \langle\alpha!\rangle\neg\beta)$ o, equivalentemente:
$\models [\alpha!]\neg\beta \leftrightarrow (\alpha \to \neg[\alpha!]\beta)$.
Vemos que el operador del anuncio público de α puede ser conmutado con la negación si el anuncio es verdadero. Recuerde que definimos $K_i\beta$ como $\neg B_i\neg\beta$. Podemos mostrar que lo anterior sucede también con los operadores epistémicos:

Lema 153. $(\alpha \to \mathcal{K}_i[\alpha!]\beta) \equiv [\alpha!]\mathcal{K}_i\beta$

Los siguientes lemas describen cómo los operadores de anuncio público interactuan con otros operadores lógicos.

Lema 154. $[\alpha!](\beta \wedge \gamma) \leftrightarrow [\alpha!]\beta \wedge [\alpha!]\gamma$

Lema 155. $[\alpha!](\beta \to \gamma) \leftrightarrow ([\alpha!]\beta \to [\alpha!]\gamma)$

Lema 156. *si ρ es una letra proposicional,* $\models [\alpha!]\rho \leftrightarrow (\alpha \to \rho)$

¿cómo funciona un operador de anuncio público al interior de otro? Lo veremos en el siguiente

Lema 157. $[(\alpha \wedge [\alpha!]\beta)!]\delta \equiv [\alpha!][\beta]\delta$

Con las equivalencias anteriores se puede mostrar que cada fórmula del lenguaje es equivalente a una fórmula sin operadores de anuncio público.

Observa Velázquez-Quesada que si quisieramos utilizar un modelo similar para representar cómo un anuncio público verdadero afecta las creencias del individuo, podríamos fácilmente obtener resultados inesperados. Si, por ejemplo, el sujeto cree algo falso y se le anuncia que es verdadero, podría el conjunto de sus creencias tornarse inconsistente, sin haberlo sido antes del anuncio. En efecto, supongamos que en su situación actual, w, solo tiene acesso a mundos que validan P, pero esta proposición es falsa en w. Entonces el anuncio elimina todos esos sucesores de w y este se convierte en un mundo "ciego" y, como resultado, toda proposición que inicia con el operadror $\mathcal{B}$ es verdadera en w. Es decir, el agente paga muy caro por corregir una creencia errónea. Tira al niño con todo y el agua de la bañera.

Una axiomatización de esta lógica contiene los axiomas de S_5 para cada K_i y fórmulas válidas que exhiben el comportamiento de los operadores $[\alpha!]$ en presencia de los otros operadores o conectivos lógicos, a saber:
$[\alpha!][\beta]\gamma \leftrightarrow\equiv [(\alpha \wedge [\alpha!]\beta)!]\gamma$ y
$(\alpha \to \mathcal{K}_i[\alpha!]\beta) \leftrightarrow [\alpha!]\mathcal{K}_i\beta$
Además de los axiomas de S_5 para cada K_i y de los esquemas de los dos teoremas anteriores, los siguientes esquemas completan una axiomatización de PAL:

1. $[\alpha!]\rho \leftrightarrow \rho$ (donde ρ es una letra proposicional).
2. $[\alpha!](\beta \wedge \gamma) \leftrightarrow ([\alpha!]\beta \wedge [\alpha!]\gamma)$
3. $[\alpha!]\neg\beta \leftrightarrow (\alpha \to \neg[\alpha!]\beta)$.

Las reglas de inferencia son *Modus Ponens* y "necesitación" para cada K_i (es decir si α es teorema, también lo es $K_i\alpha$). Con ello obtenemos un sistema correcto y completo para la semántica ofrecida.

Claramente el problema de la omnisciencia lógica reaparece en la lógica del anuncio público, pues los axiomas que rigen a los operadores B_i son los mismos que en S_5. Además surgen formas suplementarias del problema. Por ejemplo, si α y β son equivalentes de acuerdo a PAL, y en un mundo w de un modelo M se tiene que:
si $M, w \models \langle\alpha!\rangle\mathcal{B}_i\gamma$ entonces $M, w \models \langle\beta!\rangle\mathcal{B}_i\gamma$.
Es decir, si como resultado de un anuncio de α el agente aprendió algo, lo mismo

habría aprendido si en lugar de α hubiese sido hecho del conocimiento público alguna proposición equivalente.
Como sugerí también podríamos aplicar este operador al caso de un individuo aislado. De ser así, $[\alpha!]\mathcal{K}\beta$ podría interpretarse como: después de que el agente se da cuenta (de cualquier manera) que α, sabe (o aprende) β. El operador $[\alpha!]$ sería similar al operador de consciencia. El sujeto no se había percatado de que $\alpha \to \beta$. Cuando se percata entonces cree tal cosa. En este caso $[\alpha \to \mathcal{K}\beta!]$ puede interpretarse como que una vez que el agente advierte de una relación de consecuencia lógica)y no simplemente de implicación material), sabe β. Digo que no es una implicación material lo que está en juego pues la operación correspondiente suprimirá mundos del modelo. La omnisciencia lógica aparecerá de nuevo, por ejemplo, de la siguiente manera:
Si $\alpha \models \beta$ entonces $\models [\alpha!]\mathcal{B}\beta$
Si $\alpha \models \beta$ entonces $\models \beta \wedge [\beta!]\mathcal{B}\gamma \to [\alpha!]\mathcal{B}\gamma$
De nuevo, esto es excesivo para un agente ordinario. Más adelante presentaré la operación de actualización con una fórmula que se aplica a modelos de plausibilidad y que es similar a la del anuncio público.

4.8.2. Los modelos de verosimilitud

En los modelos de Hintikka, y en muchas de sus variantes, el conocimiento y la creencia se representan de la misma manera, a saber, como verdad en todos los mundos posibles. La única variación importante entre ambos conceptos es que, en el caso del conocimiento, el mundo w en que el se evalúa una fórmula es accesible a sí mismo, mientras que, en el caso de la creencia, solo se exige que el conjunto de mundos accesibles desde w no sea vacío[8]. En 2008 Baltag y Smets introdujeron otra forma de modelar la creencia. En lo que sigue, no seguiré directamente la presentación de estos autores, sino la que hace Fernando Velázquez-Quesada en su texto de 2014.

Definición 158. Sea W un conjunto no vacío y $R \subseteq W \times W$. Dado $w \in W$ definimos $C_R(w)$ (que llamaré la clase de comparabilidad de w) como: $C_R(w) = \{z \in W | Rwz \text{ o } Rzw\}$.

Definición 159. Sea W un conjunto no vacío y $R \subseteq W \times W$. Si $\emptyset \neq U \subseteq W$, $Max_R(U) = \{z |$ para cada $v \in U$ $Rvz\}$. Los elementos de $Max_R(U)$ serán llamados los R-máximos de U.

[8] Desde luego esto también se cumple para los mundos en que deben ser evaluldadas las subfórmulas de la fórmula original.

La referencia a R será eliminada si está claro a qué relación estoy aludiendo.

Definición 160. Sea $W \neq \emptyset$ y $R \subseteq W \times W$. Diremos que R es un preorden localmente bueno de W, si y solo si:
a) R es reflexiva y transitiva y
b) para cada U, tal que $\emptyset \neq U \subseteq W$, $Max_R(U) \neq \emptyset$.

Si R es un preorden localmente bueno y $w \in W$, entonces $w \in C_R(w)$ y, por lo tanto, $Max_R(C_R(w)) \neq \emptyset$, pero el máximo de esa clase no tiene por qué ser único. Sin embargo si $u, v \in Max_R(C_R(w))$, entonces $u \leq v$ y $v \leq u$.

Se puede demostrar que la siguiente definición es equivalente:

Definición 161. Sea $W \neq \emptyset$ y $R \subseteq W \times W$. R es un preorden localmente bueno de W, si y solo si se satisfacen las tres siguientes condiciones:
a) R es reflexiva y transitiva.
b) Si $u, v \in C_R(w)$ (para cualquier w), entonces uRv o vRu.
c) No existe un subconjunto infinto $U = \{u_1, u_2, u_3, \ldots\}$ de W tal que $u_i \neq u_j$ (con $i \neq j$) y $R(u_i, u_{i+1})$ para todo $i \in \mathbb{N}$

Advierta que para ser una relación de equivalencia a R le falta la simetría. Sin embargo, si R es un preorden localmente bueno de W, la relación $xRy \vee yRx$ es una relación de equivalencia.

Definición 162. Un modelo de plausibilidad es un modelo de Kripke $M = \langle W, \leq, V\rangle$ en que $\leq$ es un preorden localmente bueno de W.

Si $u, v \in W$, diré que a) u es al menos tan verosímil como v si $v \leq u$, b) que u y v son igualmente verosímiles (lo que denotaré por $u \cong v$) si $v \leq u$ y $u \leq v$, y, por último, c) que u es más verosímil que v si $v \leq u$ pero no $u \leq v$.
Por otro lado, denotaré por $\sim$, a la relación de equivalencia que se da x y y, si y solo si $x \leq y \vee y \leq x$. Cada mundo u tendrá su clase de equivalencia $\bar{u}$, que es lo que antes llamaba la clase de comparabilidad de u. Es decir, $\bar{u} = \{z | z \leq u \text{ o } u \leq z\}$.

Ahora introducimos dos nuevos operadores unarios, $\langle\leq\rangle$ y $\langle\sim\rangle$, en el lenguaje. A las reglas de fórmación de fórmulas, agregamos las siguiente cláusula: Si α es una fórmula, también lo son $\langle\leq\rangle\alpha$ y $\langle\sim\rangle\alpha$.
Supongamos dados un modelo $M = \langle W, \leq, V\rangle$ y $w, v \in W$. A la definición de "enunciado verdadero en w" agregamos las cláusulas:

1. $M, w \models \langle\leq\rangle\alpha$ si y solo si para algún $v \in W$, tal que $u \leq v$ se tiene $M, v \models \alpha$ y

2. $M, w \models \langle\sim\rangle\alpha$ si y solo si hay algún $v \in W$ tal que $w \sim v$ y $M, v \models \alpha$.

Como siempre, la validez es verdad en todos los mundos de todos los modelos. Ahora podemos introducir los operadores duales $[\leq]$ y $[\sim]$, definidos de manera usual, a saber:

$$[\leq]\alpha \equiv_{Def} \neg\langle\leq\rangle\neg\alpha \text{ y } [\sim]\alpha \equiv_{Def} \neg\langle\sim\rangle\neg\alpha$$

Ahora bien, ¿cómo debemos interpretar estos operadores desde el punto de vista doxástico? Si en una clase de equivalencia están u y v, el sujeto no puede distinguir cuál de esos mundos es el actual aunque, por ejemplo, si $u < v$ entonces pensará que v es más verosímil que u. La idea central de esta representación de la creencia es que un agente cree una proposición porque esta es verdadera en todos los mundos que considera más verosímiles en una situación dada. Más precisamente, el individuo cree α en un mundo w si para toda $v \in Max(\bar{w})$, $v \models \alpha$. Por otro lado, sabrá α en w si para todo $v \in \bar{w}$, $v \models \alpha$. La cuestión está en cómo expresar estos hechos en el lenguaje objeto. Ello se resuelve con el siguiente lema:

Lema 163. *Para todo $v \in Max(\bar{w})$, $v \models \alpha$ si y solo si $w \models \langle\leq\rangle[\leq]\alpha$*

Demostración. Supongamos que $w \models \langle\leq\rangle[\leq]\alpha$ y sea v tal que $v \in Max(\bar{w})$. Por lo primero, existe z tal que $w \leq z$ y para toda x tal que $z \leq x$, $x \models \alpha$. Dado que $\leq$ conecta a todos los miembros de $\bar{w}$ y que v es máximo en esta clase, $z \leq v$. Por lo tanto, $v \models \alpha$.
Ahora supongamos que para todo $v \in Max(\bar{w})$, $v \models \alpha$. Sea $z \in Max(\bar{w})$, puesto que $\leq$ conecta a todos los miembros de $\bar{w}$, tenemos que $w \leq z$. Sea x tal que $z \leq x$, entonces $x \in Max(\bar{w})$ y, por lo tanto, $x \models \alpha$. En consecuencia, $z \models [\leq]\alpha$ y $w \models \langle\leq\rangle[\leq]\alpha$. $\square$

Con esto podemos definir operadores epistémicos $\mathcal{K}$ y $\mathcal{B}$ por medio de la siguientes convenciones:

$\mathcal{K}\alpha \equiv_{Def} [\sim]\alpha$
$\mathcal{B}\alpha \equiv_{Def} \langle\leq\rangle[\leq]\alpha$

Evidentemente K satisface todos los esquemas T, 4 y 5 pues $\sim$ es una relación de equivalencia, mientras que B cumple D, 4 y 5. Lo más importante es que para ambos es verdadero el axioma de distribución. En particular,
$\models \mathcal{B}(\alpha \to \beta) \to (\mathcal{B}\alpha \to \mathcal{B}\beta)$

es decir, el agente será lógicamente omnicreyente. Para evitarlo, aún serán necesarios otros artificios que veremos más adelante.

Una presentación axiomática de este sistema puede contribuir a esclarecer su funcionamiento. Se obtiene de la siguiente manera:
Axiomas: Si α y β son fórmulas, entonces son axiomas:

1) Cualquier ejemplificación de una tautología.
2) $[\sim](\alpha \to \beta) \to ([\sim]\alpha \to [\sim]\beta)$
3) $([\sim]\alpha \to \alpha)$
4) $([\sim]\alpha \to [\sim][\sim]\alpha)$
5) $(\langle\sim\rangle\alpha \to [\sim]\langle\sim\rangle\alpha)$
6) $[\leq](\alpha \to \beta) \to ([\leq]\alpha \to [\leq]\beta)$
7) $([\leq]\alpha \to \alpha)$
8) $([\leq]\alpha \to [\leq][\leq]\alpha)$
9) $\langle\leq\rangle\alpha \to \langle\sim\rangle\alpha$
10) $(\langle\sim\rangle\alpha \wedge \langle\sim\rangle\beta) \to (\langle\sim\rangle(\alpha \wedge \langle\leq\rangle\beta) \vee (\langle\sim\rangle(\beta \wedge \langle\leq\rangle\alpha)$

y las siguientes son reglas de inferencia:

R_1 de $\vdash \alpha$ y $\vdash \alpha \to \beta$ concluya $\vdash \beta$

R_2 de $\vdash \alpha$ se sigue $\vdash [\sim]\alpha$

R_3 de $\vdash \alpha$ se sigue $\vdash [\leq]\alpha$

Los axiomas (2)-(5) con la regla R_1 corresponden al hecho de que $\sim$ es una relación de equivalencia. Los axiomas (6)-(8) con la regla R_3 corresponden a que $\leq$ es una relación reflexiva, transitiva y cerrada (bajo implicación material). El axioma (9) refleja que $\leq$ está contenida en $\sim$. Por último, el décimo axioma asevera que si en una clase de equivalencia $\bar{w}$ hay un mundo v que hace verdadero a α y otro, v, que hace lo propio con β, entonces (dado que $u \leq v$ o $v \leq u$) habrá en la clase un mundo en que valga α y y tenga arriba un mundo en que β sea verdadero o al revés.

Me parece que esta forma de representar la creencia y el conocimiento es más atractiva que la de los modelos de Kripke. En primer lugar es más intuitiva. En lo que se refiere al conocimiento es idéntica a la de S_5. Para distinguir este mundo de otro, el agente concibe variaciones, pero deja fijo todo lo que sabe. En cambio, la creencia en una proposición P, no lleva al agente a descartar como posibilidad epistémica los mundos en que P no sucede. Yo creo que ayer tembló en Oaxaca. Concibo que puede ser falso, pero no lo descarto. Simplemente doy más peso a los mundos en que eso ocurrió cuando se trata de conjeturar qué mundo es el real. Otra ventaja es que en este enfoque el agente siempre sabe qué es lo que cree. En efecto, advierta que $\models (\mathcal{B}\alpha \rightarrow \|\mathcal{B}\alpha)$. De la misma nanera, si no cree algo, sabe que no lo cree. Esto porque desde cualquier mundo de una clase de equivalencia el acceso al anillo más verosímil es el mismo. Es un resultado que concuerda con nuestras intuiciones. Nadie diría: "no sé si creo P".[9] Otras ventajas aparecerán en las siguiente secciones donde veremos cómo Baltag Smets modelan el cambio de creencias con modelos de plausibilidad y evitan la omnisciencia lógica.

4.8.3. Modelos de plausibilidad y consciencia

El primer paso consiste en introducir en los modelos de plausibilidad una función unaria A que asocia a cada mundo posible un conjunto de fórmulas. Como vimos, esta idea fue propuesta por primera vez por Fagin y Halpern. Sin embargo, hay una diferencia importante. Esta vez, que la fórmula $\alpha \in A(w)$ no significa simplemente que el agente es consciente en el mundo w de la fórmula α, sino que en esa situación él la reconoce como verdadera. ¿Significa esto que α en w es verdadera y el agente la juzga como tal? Hay varias posibilidades. En la que veremos enseguida $\alpha \in A(w)$ se interpreta intuitivamente como que el agente considera a esa fórmula como verdadera (en w), lo sea realmente o no. Correspondiente a este elemento de los modelos, hay un conectivo unario $\mathcal{A}$ tal que:

$$w \models \mathcal{A}\alpha \text{ si y sólo si } \alpha \in A(w)$$

Con este operador agregado al lenguaje podemos representar la diferencia entre las variantes explícita e implícita tanto del conocimiento como de la creencia. Para el caso implícito conservamos $[\sim]\alpha$ y $\langle\leq\rangle[\sim]\alpha$ respectivamente. Para el conocimiento explícito hay varias opciones. Recordemos que para Fagin y Halpern

[9] En la acepción de "creer" que me interesa, es decir, la que es compatible con la noción clásica de conocimiento.

saber α en w significa que, en todos los mundos accesibles desde w, α es verdadera y que el individuo es consciente de α en w, lo que podemos simbolizar por $\mathcal{A}\alpha \wedge [\sim]\alpha$. El propio Velázquez-Quezada propuso en un artículo previo $[\sim]\mathcal{A}\alpha$ como representación del conocimiento. Por supuesto que correspondiendo con estas dos opciones podríamos tener las correspondientes para la creencia explícita (a saber, $\mathcal{A}\alpha \wedge \langle\leq\rangle[\leq]\alpha$ y $\langle\leq\rangle[\leq]\mathcal{A}\alpha$). En el artículo que ahora examinamos, el autor explora las opciones $[\sim](\alpha\wedge\mathcal{A}\alpha)$ para el conocimiento explícito y $\langle\leq\rangle[\leq](\alpha\wedge\mathcal{A}\alpha)$ para la creencia explícita.

Como siempre, la validez es verdad en todos los mundos de todos los modelos. Las fórmulas válidas serán exactamente las mismas que teníamos antes de la introducción del operador $\mathcal{A}$. En consecuencia, el sistema puede ser axiomatizado con la lista de fórmulas y reglas de la sección anterior. Esto se debe a que no hemos impuesto ninguna condición sobre los conjuntos $\mathcal{A}(v)$ (con $v \in W$). Lo que hemos ganado, en cambio, es que con las definiciones dadas ni el conocimiento explícito ni la creencia explícita son cerrados bajo consecuencia lógica, dado que cualquier fórmula puede ser invalidada por un agente que no la reconoce como tal. ¿Qué otras propiedades tendrán los conceptos epistémicos así definidos? Representemos por $\mathcal{K}_E$ y $\mathcal{K}_I$ respectivamente a las versiones explícita e implícita del conocimiento (y con $\mathcal{B}_E$ y $\mathcal{B}_I$ a las correspondientes a la creencia). Entonces es fácil demostrar los siguientes lemas.

Lema 164. *1)* $\models \mathcal{B}_E\alpha \to \mathcal{B}_I\alpha$.

2) $\models \mathcal{K}_E\alpha \to \mathcal{K}_I\alpha$.

3) $\models \mathcal{K}_I\alpha \to \mathcal{B}_I\alpha$.

4) $\models \mathcal{K}_E\alpha \to \mathcal{B}_E\alpha$,

Lema 165. *Dados mundos w y v de un modelo M tales que $v \in \bar{w}$, entonces*

1) $M, w \models \mathcal{K}_I\alpha$ *si y solo si* $M, v \models \mathcal{K}_I\alpha$

2) $M, w \models \mathcal{K}_E\alpha$ *si y solo si* $M, v \models \mathcal{K}_E\alpha$

3) $M, w \models \mathcal{B}_I\alpha$ *si y solo si* $M, v \models \mathcal{B}_I\alpha$

4) $M, w \models \mathcal{B}_E\alpha$ *si y solo si* $M, v \models \mathcal{B}_E\alpha$

Si vemos ahora cómo el sistema hasta ahora implementado enfrenta el problema de la omnisciencia lógica y evaluamos en él las fórmulas que empleamos para otros sistemas, obtenemos el mismo resultado que en la lógica de la consciencia general. En ambos casos, es posible falsificar una fórmula en una situación w eliminándola de $A(w)$. Solo debemos recordar que ahora $\alpha \in A(w)$ no significa que α aparece en el campo de atención del sujeto, sino que este la reconoce como verdadera (en esa situación). De nuevo el sujeto no creerá nunca una contradicción (o no debe hacerlo).

¿Qu'e pasa con la creencia de orden superior? A pesar de que los modelos son transitivos y euclideanos ni $\mathcal{B}_E\alpha \rightarrow \mathcal{B}_E\mathcal{B}_E\alpha$ ni $\neg\mathcal{B}_E\alpha \rightarrow \mathcal{B}_E\neg\mathcal{B}_E\alpha$ son válidos. Para verlo basta considerar que, por ejemplo, el agente puede ser consciente de α en los mundos máximos de una clase, pero no de $\mathcal{B}_E\alpha$[10]

4.8.4. Cambio de creencias en modelos de plausibilidad

Los modelos de plausibilidad resultan adecuados para modelar la revisión de creencias cuando el agente recibe nueva información (sea esta verdadera o no). Mostré antes brevemente cómo representa Plaza el anuncio público verídico. La operación correspondiente a anunciar α consiste en evaluar la fórmula en el mismo mundo, pero ahora como parte de un submodelo del modelo original, si es que α es verdadera. El modelo en cuestión es la restricción del original a los mundos en que α es verdadera. Si α es falsa en w, $[\alpha!]\beta$ es verdadera (y $\langle\alpha!\rangle\beta$ es falsa) en w para cualquier β. Por ello, el modelo parece solo apropiado para un anuncio verdadero que, además, el agente toma como tal. Es posible combinar el anuncio público de Plaza con los modelos de plausibilidad que acabamos de ver. Sin embargo, no me detendré en este punto sino que pasaré a una forma más adecuada de representar el cambio de creencias en los modelos de verosimilitud.

Enseguida veremos una operación similar a la del anuncio público, llamada actualización, que es más interesante desde el punto de vista de este trabajo. Hay diversas variantes, pero todas ellas consisten en un reordenamiento de los mundos dentro de cada clase de equivalencia. Ahora una fuente fiable, aunque no infalible, anuncia o sugiere que α es verdadera. Esta vez el agente no puede descartar los mundos en α es falsa, pues no tiene certeza de que lo anunciado sea el caso,

[10]Bajo ciertas condiciones es posible tener estas propiedades. La condición de la introspección positiva es más fácil de aceptar que su contraparte neagativa. Ver Velázquez-Quesada (2014). Como dije, las creencias de orden superior no son parte primordial de mi estudio.

aunque sí un alto grado de confianza. La idea es que el anuncio altera la relación $\leq$ sin que deje de ser un subconjunto de $\sim$. Hay varias propuestas, algunas más acertadas que otras, para representar este fenómeno. Siguiendo a Fernando Velázquez-Quesada (2014) expondré la noción llamada actualización radical, que consiste en el siguiente reordenamiento. La actualización radical con α (es decir, producto de un anuncio u observación de α) generará un nuevo modelo en que dentro de cada clase de equivalencia los mundos en que α se da tendrán mayor plausibilidad que todos aquellos en que α es falsa, y por lo demás el orden se conserva. Aquí tenemos, de nuevo, un principio de conservatividad. Si representamos por $\leq$ al orden original y con $\leq^{\alpha}$ al nuevo, entonces si $w \leq v$ y α es verdadera en v o falsa en w, entonces $w \leq^{\alpha} v$; mientras que si α es falsa en v y verdadera en w, entonces $v \leq^{\alpha} w$.

Si $\|\alpha\|^M$ es el conjunto de mundos en que α es verdadero en un modelo M, entonces el modelo que resulta por una actualización con α, $M \Uparrow \alpha$, puede definirse de la siguiente manera:

Definición 166. Si $M =< W, R, \leq, V >$ es un modelo de plausibilidad y α un fórmula entonces:
$M \Uparrow \alpha =< W, R, \leq^{\alpha}, V >$ donde:

$$\leq^{\alpha} = (\leq \cap (W \times \|\alpha\|^M)) \cup (\leq \cap (\|\neg\alpha\|^M \times W)) \cup (R \cap (\|\neg\alpha\|^M \times \|\alpha\|^M))$$

Se puede demostrar que $M \Uparrow \alpha$ es un modelo de plausibilidad, si M lo es. Denotemos por $\sim^{\alpha}$ a la correspondiente clase de equivalencia del nuevo modelo (es decir, si $w \sim^{\alpha} v \equiv_{Def} w \leq^{\alpha} v \wedge v \leq^{\alpha} w$) y supongamos que $w \sim v$. Si en w y v, α es verdadera o, en ambos, α es falsa, entonces $w \sim^{\alpha} v$. Si $w \models \alpha$ y $v \not\models \alpha$ entonces $v \leq^{\alpha} w$, mientras que si $w \not\models \alpha$ y $v \models \alpha$, entonces, $w \leq^{\alpha} v$.

Ahora agreguemos al lenguaje un operador $[\alpha \Uparrow]$ (y su dual $\langle \alpha \Uparrow \rangle$), por cada fórmula α, con la cláusula:
Si β es una fórmula, entonces $[\alpha \Uparrow]\beta$ es una fórmula.
A su vez la definición de "enunciado verdadero" se complementa con la cláusula:

$$M, w \models [\alpha \Uparrow]\beta \text{ si y solo si } M \Uparrow \alpha, w \models \beta$$

y, puesto que $\langle \alpha \Uparrow \rangle \beta \equiv \neg[\alpha \Uparrow]\neg\beta$,
$M, w \models \langle \alpha \Uparrow \rangle \beta$ si y solo si $M \Uparrow \alpha, w \not\models \neg\beta$
Es decir que $M, w \models [\alpha \Uparrow]\beta$ si y solo si $M, w \models \langle \alpha \Uparrow \rangle \beta$

Veamos cómo se comporta el conectivo $[\alpha \Uparrow]$ cuando se combina con otros operadores.

Vimos que $[\alpha \Uparrow]\beta$ es equivalente a $\langle\alpha \Uparrow\rangle\beta$. Por definición, $\neg\langle\alpha \Uparrow\rangle\beta$ es equivalente a $[\alpha \Uparrow]\neg\beta$. De allí puede concluirse que

$$M, w \models [\alpha \Uparrow]\neg\beta \text{ si y solo si } M, w \models \neg[\alpha \Uparrow]\beta$$

Evidentemente si β es una fórmula sin símbolos modales y sin el operador $[\leq]$ (o su dual) entonces:

$$M, w \models [\alpha \Uparrow]\beta \text{ si y solo si } M, w \models \beta$$

pues la reordenación que provocará la actualización con α en nada altera la evaluación de β en w.
Lo mismo sucede con el operador $\mathcal{A}$: $M, w \models [\alpha \Uparrow]\mathcal{A}\beta \leftrightarrow \mathcal{A}\beta$.
Por otro lado,

$$M, w \models [\alpha \Uparrow](\beta \vee \gamma) \text{ si y solo si } M, w \models [\alpha \Uparrow]\beta \text{ o } M, w \models [\alpha \Uparrow]\gamma$$

pues si luego de la reordenación que provocó la actualización con α en w es verdadera una disyunción, tiene que ser verdadero uno de sus disyuntos y viceversa. De lo anterior obviamente podemos deducir que

$$M, w \models [\alpha \Uparrow](\beta \rightarrow \gamma) \text{ si y solo si } M, w \not\models [\alpha \Uparrow]\beta \text{ o } M, w \models [\alpha \Uparrow]\gamma$$

Todo ello era de esperarse porque la actualización solo afecta el orden de los mundos, no las evaluaciones internas a cada mundo. Por supuesto que el caso interesante es cómo funciona el operador $[\alpha \Uparrow]$ en la cercanía de $\langle\leq\rangle$ y, por lo tanto, de los operadores epistémicos. La relación entre $\Uparrow$ y $\leq$ está condensada en la siguiente fórmula (*) válida:

$$\langle\alpha \Uparrow\rangle\langle\leq\rangle\beta \leftrightarrow \langle\leq\rangle(\alpha \wedge \langle\alpha \Uparrow\rangle\beta) \vee (\neg\alpha \wedge \langle\leq\rangle(\langle\alpha \Uparrow\rangle\beta) \vee (\neg\alpha \wedge \langle\sim\rangle(\alpha \wedge \langle\alpha \Uparrow\rangle\beta))$$

fórmula intrincada pero que admite una lectura relativamente intuitiva. Consideremos un mundo w. La fórmula dice que β será verdadero en un mundo al menos tan plausible como w después de la actualización con α si y solo si pasa alguna de tres cosas. La primera es que en un mundo al menos tan verosímil como w ocurran tanto α como que después de la actualización β sea verdadero. La segunda que en

w α sea falso, pero haya un mundo al menos tan verosímil como w en que después de la actualización β sea verdadero. La tercera es que en w α sea falso, pero en su clase de comparabilidad haya un mundo en que α sea verdadero y en que después de la actualización β sea verdadero.

En cambio, la interacción entre $\langle\alpha \Uparrow\rangle$ y $\langle\leq\rangle$ es más clara:

$$\models \langle\alpha \Uparrow\rangle\langle\sim\rangle\beta \leftrightarrow \langle\sim\rangle\langle\alpha \Uparrow\rangle\beta$$

Las equivalencias anteriores en que se muestra la interacción del operador $\langle\alpha \Uparrow\rangle$ con los otros operadores lógicos incluida la fórmula (*), más la regla de inferencia: de $\vdash \beta$ derive $\vdash [\alpha \Uparrow]\beta$, proveen, junto con los axiomas de PA, una axiomatización completa de esa lógica con la modalidad de actualización.

La actualización con α dará prioridad a los mundos en que esta fórmula es verdadera sobre los demás, pero la operación será la misma, sea α verdadera o falsa. Tampoco podrá llevar a la certidumbre porque no elimina ningún mundo del horizonte subjetivo del agente. Una ventaja de esta forma de representar la creencia es que permite modificaciones no-monotónicas. Nada impide que un agente crea una proposicíon, después de una actualización deje de creerla y, como resultado de otra actualización, vuelva a creerla. De hecho, una actualización con P seguida por una actualización con $\neg P$ dejará el modelo como estaba al comienzo. Otra es que la modificación que conlleva el cambio de creencias depende solo del individuo mismo, del grado de confianza que tiene en cada una de sus creencias. Señala Velázquez-Quesada una ventaja más de la actualización radical: la operación de actualización genera un modelo de plausibilidad cuando se aplica a otro. Es decir, la operación no cambia el tipo de modelo.

Es claro que la actualización con α modela cómo cambian las creencias implícitas del individuo cuando recibe nueva información. En cambio, para la creencia explícita es necesario tomar en cuenta las funciones $\mathcal{A}$. En el contexto del agente no-omnisciente, parece razonable representar la revisión de creencias con α, no a través de $\langle\alpha \Uparrow\rangle$, sino por medio de $\langle(\alpha \wedge \mathcal{A}(\alpha)) \Uparrow\rangle$. A esta operación se le llama "actualización no omnisciente con α". No solo requerimos que α sea verdadera sino que el agente la reconozca como tal.

Comparemos estas dos revisiones con la de la teoría AGM. En el caso de la creencia implícita, seguimos teniendo cerradura pues, como vimos, $\mathcal{B}_I$ se distribuye sobre el condicional material. En cambio, Éxito ($[\alpha \Uparrow]\mathcal{B}_I\alpha$) no siempre se satisface. Si tomamos una fórmula sin los conectivos $\langle\leq\rangle$ y $\langle\sim\rangle$ (y, por tanto, sin operadores epistémicos), después de la actualizacón con ella el agente la creerá implícitamente pues los mundos más verosímiles la validarán. Las cosas pueden ser diferentes con fórmulas epistémicas. Velázquez-Quesada ofrece el ejemplo de un enunciado tipo Moore $(P \wedge \neg \mathcal{B}_I P)$. Supongamos un modelos con dos mundos, w y v tal que $w < v$ y tales que $w \models P$ y $v \not\models P$. Entonces $(P \wedge \neg \mathcal{B}_I P)$ solo será verdadera en w, por lo que al actualizar con este enunciado, v quedará por debajo de w y ahora el agente sabrá P, por lo que $(P \wedge \neg \mathcal{B}_I P)$ será falso en todo el modelo. De paso note que si en el modelo resultante de esta actualizacón actualizamos ahora con $\neg(P \wedge \neg \mathcal{B}_I P)$, el orden de los mundos ya no se altera. Sin embargo, tras una actualización con una fórmula α que no aluda a las creencias del agente, este creerá α, siempre y cuando ya exista en el modelo un mundo en el cual α sea verdadera. No habrá modo de que el individuo rechace esa información, excepto en el caso mencionado. Tampoco se cumple el postulado según el cual una actualización con α en un w genera un subconjunto de $Con(Max(C_{\leq}(w))$. En efecto, puede suceder que el reacomodo de mundos al interior de una clase de comparabilidad finalice con que todos los mundos máximamente veroímiles contengan Q, que antes no era creída. La actualización satisface extensionalidad porque si P y Q son lógicamente equivalentes, todo mundo posible que tenga a una tendrá a la otra y, por lo tanto, estos enunciados sufrirán la misma suerte después de una reordenación. Si consideramos a la actualización con $\neg P$ como una retracción de P, Recovery no es válida. Baste recordar que el ejemplo que vimos en que la actualización con $(P \wedge \neg \mathcal{B}_I P)$ alteraba el orden de los mundos mientras que el modelo así resultante la actualización con $\neg(P \wedge \neg \mathcal{B}_I P)$ no producía ningún cambio. Sin embargo, lo que nos interesa evaluar no es la creencia implícita sino la explícita. Lo que haré al final de esta sección.

En lo que sigue esbozaré una propuesta de Velázquez-Quesada para modelar cómo los actos de inferencia de un sujeto pueden cambiar sus conocimiento y sus creencias y que resultan de una combinación y enriquecimiento de algunos de los artificios hasta ahora vistos.

4.8.5. Modelos de acción, reconocimiento y verosimiltud

El título de la sección refleja los varios recursos que utiliza a Velázquez-Quesada en el sistema que revisaré ahora. La novedad son los modelos de acción propuestos por Baltag, Moss y Solecki (1999) en su combinación con los modelos vistos en la sección anterior.

Definición 167. Un modelo de acción PA (de plausibilidad y reconocimiento) es un cuádruple $O = \langle E, \preceq, Pre, Pos_A \rangle$ donde:

1. $\langle E, \preceq, Pre, \rangle$ es un modelo de plausibilidad con E un conjunto finito cuyos elementos serán llamados eventos. Es decir, que $\preceq$ es un preorden localmente bueno en E y Pre una función que asigna fórmulas a mundos.

2. $Pos_A : (E \times Form) \Rightarrow Form$, donde $Form$ es el conjunto de fórmulas del lenguaje.

Si $O = \langle E, \preceq, Pre, Pos_A \rangle$ y $e \in E$, a la pareja (O, e) la llamaré un modelo de acción puntuado. La idea será aplicar un modelo acción sobre un modelo de plausibilidad para obtener un nuevo modelo de plausibilidad. Cuando un modelo de acción puntuado (E, e) se aplique a un mundo w de M, si en w son verdaderas las fórmulas de $Pre(E)$ (lo que significa que se dan las condiciones para que la acción se lleve a cabo), se generará un nuevo mundo del nuevo modelo en que el agente reconocerá como verdaderas las fórmulas de $POS_A(e)(w)$. Si otro evento e' tiene precondiciones que se satisfacen en w, habrá otra "copia" de w en el nuevo modelo, excepto que allí el agente reconocerá como verdaderas otras formulas (las contenidas en $POS_A(e')(A(w))$. Es decir, el primer efecto de la aplicación del modelo de acción será cambiar el conjunto de fórmulas que el agente reconoce como verdaderas. ¿Qué orden de verosimilitud tendrán los mundos así creados? Tenemos dos órdenes en competencia, a saber, $\leq$ y $\preceq$, el primero del modelo de verosimilitud original y el segundo del modelo de acción. La idea es ordenar los mundos por orden "lexicográfico", o sea decidir tomando en cuenta $\preceq$ y, en caso de empate, será $\leq$ quien determine la relación. Para hacerlo más formal denotemos por $\approx$ a la relación que se establece entre e y e' si $e \preceq e'$ y $e' \preceq e$, mientras que $e \prec e'$ significará que $e \preceq e'$, pero no $e' \preceq e$. Con esta notación podemos formalizar las ideas anteriores mediante las siguientes definiciones.

Definición 168. Sea $M = \langle W, \leq, V, A \rangle$ un modelo de plausibilidad y $O = \langle E, \preceq , Pre, Pos_A \rangle$ un modelo de acción. La operación actualización producto es la que aplicada a M y a O, genera el modelo $M \otimes O = \langle W', \leq', V', A' \rangle$ tal que:

- $W' = \{(w, e) \in E \times E | M, w \models Pre(e)\}$
- $(w, e) \leq' (v, f)$ si y solo si $e \prec f$ y $w \sim v$ o $e \approx f$ y $w \leq v$.
- $V'(w, e) = V(w)$
- $A'(w, e) = Pos_A(e, A(w))$

Es decir, el nuevo modelo tendrá por cada mundo w del modelo original y cada evento de cuyas precondiciones sean satisfechas por w una réplica (w, e). Estas réplicas evaluarán las fórmulas atómicas como lo hacían los mundos originales. El orden de los nuevos mundos será el que antes dije y la función A de reconocimiento producirá conjuntos nuevos de fórmulas según lo que determine la función $Pos_A(e)$ aplicado a $A(w)$.

Ejemplo 169. El siguiente es un ejemplo de modelo de acción que muestra el efecto de hacer un *modus ponens* con implicación y antecedente conocido.
Sea $O^{\nu \to \xi}_{KK} = \langle E, \preceq, Pre, Pos_A \rangle$ donde $E = \{e\}$,
$\preceq = \{(e, e)\}$, $Pre(e) = \{K_E(\nu \to \xi), K_E\nu\}$, $Pos_A(e, X) = X \cup \{\xi\}$.

Supongamos que tenemos un modelo de plausibilidad M y un mundo w de M, tal que en todo $\bar{w}$, son verdaderos $K_E(\nu \to \xi)$ y $K_E\xi$. Tal vez el individuo no sabe explícitamente ξ porque en algunos mundos de $\bar{u}$ aunque, por supuesto vale ξ, él no lo reconoce como verdadero. Sin embargo, al aplicar el modelo de acción anterior, cada mundo w dará lugar a otro w^*, y si ξ ya estaba en $A(w)$ entonces $w = w^*$. Si no era este el caso, w^* será idéntico a w, salvo que $\xi \in A'(w^*)$. Por ello, en el nuevo modelo $K_E\xi$ será verdadera en todo $\bar{w}$.

Como antes, podemos introducir un operador en el lenguaje que codifique el efecto de un modelo de acción puntuado sobre un mundo de un modelo de plausibilidad, a través de la siguiente definición.

Definición 170. Sean M, w y O, e modelos puntuados, de plausibilidad el primero y de acción el segundo.

$$M, w \models \langle O, e \rangle \alpha \text{ si y solo si } M, w \models Pre(e) \text{ y } M \otimes O, (w, e) \models \alpha$$

Ejemplo 171. Si O es el modelo de acción del ejemplo anterior y M, w un modelo puntuado de posibilidad, entonces

$$\text{Si } M, w \models K_E(\nu \to \xi) \wedge K_E\nu \text{ entonces } M \otimes O^{\nu \to \xi}_{KK}, w \models K_E\xi$$

Enseguida tomo otros ejemplos del autor con que ilustra muy claramente la utilidad de esta clase de modelos, ahora para casos en que el individuo hace una inferencia cuyas premisas o implicación son creídas y no necesariamente conocidas.

Ejemplo 172. Supongamos que el agente sabe ν y cree que $(\nu \to \xi)$. La acción y efecto de llevar a cabo un *modus ponens* pueden ser representados por el siguiente

modelo de acción:
$O^{\nu\to\xi}_{BK} = \langle E, \preceq, Pre, Pos_A \rangle$ donde $E = \{e, f\}$,
$\preceq = \{(e,e), (e,f), (f,f)\}$,
$Pre(e) = Pre(f) = \{B_E(\nu \to \xi), K_E\xi\}$
$Pos_A(e, X) = X \backslash \{\nu \to \xi\} \cup \{\neg(\nu \to \xi), \neg\xi\}$, $Pos_A(f, X) = X \cup \{\xi\}$

Supongamos de nuevo un modelo de plausibilidad M y un mundo w de M tal que para todo $v \in \bar{w}$, $v \models \nu \wedge \mathcal{A}(\nu)$, y para todo $z \in Max(\bar{u})$, $z \models (\nu \to \xi) \wedge \mathcal{A}(\nu \to \xi)$. El efecto de aplicar el modelo de acción anterior a M es que cada mundo v en w dará lugar a dos mundos nuevos, (v, e) y (v, f). En los del primer tipo se agregarán $\neg(\nu \to \xi)$ y $\neg\xi$ al conjunto de las fómulas que el sujeto ya reconocía y en cambio, se eliminará del mismo $(\nu \to \xi)$. En los de la segunda clase a las fórmulas que el sujeto reconocía se añade ξ. Para cada $v \in \bar{u}$, (v, f) será más verosímil que (v, e). Así si el individuo cree explícitamente el condicional $(\nu \to \xi)$ y sabe explíctamente ξ, como resultado de la inferencia creerá explícitamente ξ, pero no lo sabrá. Esto puede ser representado de la siguiente manera:

Si $M, w \models \mathcal{B}_E(\nu \to \xi) \wedge \mathcal{K}_E\nu$ entonces $M \otimes O^{\nu\to\xi}_{BK}, w \models (\mathcal{B}_E\xi \wedge \neg\mathcal{K}_E\xi)$

Estas operaciones son globales en el siguiente sentido. Consideremos el ejemplo anterior. El modelo que resulta del producto contiene dos mundos (w, e) y (w, f) por cada mundo w que cumple las precondiciones de e y f (que son las mismas). ¿Qué ocurre con los otros mundos? Resulta que no hay otros mundos porque si v y z son mundos en $\bar{u}$, $v \models \mathcal{B}_E(\nu \to \xi) \wedge \mathcal{K}_\mathcal{E}\nu$ si y solo si $u \models \mathcal{B}_E(\nu \to \xi) \wedge \mathcal{K}_\mathcal{E}\nu$. Es decir, las precondiciones se satisfacen toda la clase (si se satisfacen en un elemento) y los resultados también son válidos en todo $\bar{u}$. Por ejemplo, en el caso anterior, el sujeto cree ξ en cualquier elemento de la clase. Sin embargo, también podemos modelar inferencias “locales” como explicaré enseguida.

En efecto, un modelo de acción puede actuar a nivel local, es decir, que los eventos que lo forman contengan precondiciones que no se satisfagan en todos los mundos. Por ejemplo, puede ocurrir que tengamos dos eventos cuyas precondiciones son complementarias. Cada mundo del modelo de plausibilidad cumplirá una y solo una de esas precondiciones y, por lo, tanto, dará lugar a un solo mundo extra en el nuevo modelo. Por ejemplo, evidentemente todos los mundos del modelo de plausibilidad que validen una fórmula ϕ harán verdadera $\neg\neg\phi$, pero puede ser que en en algunos de ellos las fórmulas $\mathcal{A}\phi$ y $\mathcal{A}(\phi \to \neg\neg\phi)$ sean verdaderas sin que lo sea $\mathcal{A}(\neg\neg\phi)$. En ese caso es posible diseñar un modelo de acción con dos eventos. El primero con precondición $\{\mathcal{A}\phi, \mathcal{A}(\phi \to \neg\neg\phi)\}$ y el otro con el complemento de este conjunto. Para el primero, la instrucción es "agrege $(\neg\neg\phi)$ a las fórmulas que el individuo reconoce" y, para la segunda, "deje todo como estaba". En ese caso, el modelo resultado del producto tendrá los mismos mundos que el original, salvo que aquellos que satisfacen la precondición del primer evento también validarán $\mathcal{A}\neg\neg\phi$. Este es un ejemplo muy sencillo de acción local. Muchos otros refinamientos pueden ser diseñados utilizando las mismas ideas. El sistema admite una axiomatización completa, pero no es necesario exponerla aquí.

Evaluación

Hemos visto cómo pueden combinarse los modelos de plausibilidad, los de acción y el operador de conciencia para ofrecer un modelo muy rico de cómo los estados doxásticos y epistémicos de un individuo se modifican sea por un anuncio o información (no necesariamente cierta), sea por la acción inferencial del propio sujeto. Como dije antes, los modelos de plausibilidad ya constituyen una forma de representar la creencia y el conocimiento más adecuada que la de los sistemas normales. ¿En que se advierte esta mejora? En parte, en la intuición que la respalda de que el conocimiento es de naturaleza distinta de la creencia. En los modelos de plausibilidad, las creencias aparecen esencialmente escalonadas según la convicción con que son mantenidas por el sujeto. Sin embargo, tanto las creencias como los conocimientos del sujeto son cerrados bajo consecuencia lógica en este sistema. Para evitarlo, es necesario introducir el operador de conciencia y hay varias maneras de hacerlo. En la que vimos el sujeto sabe que ϕ en una situación si ϕ es verdadera en toda la clase de comparabilidad de esa situación y el sujeto la reconoce como tal en toda la clase. Por otro lado, el sujeto cree que ϕ si en toda la capa más verosímil de esa clase ϕ es verdadera y el individuo la reconoce como tal en esos mundos. Con ello, tenemos versiones explíctias tanto del conocimiento como de la creencia. Ahora el cambio de creencias generado por una información novedosa puede modelarse a través del cambio en el orden de verosimilitud al interior de cada clase de comparabilidad. Vimos que, tal como lo definimos, el cambio de creencias implícitas no cumple varios de los postulados de la teoría AGM. En el caso de la creencia explícita, ni siquiera Cerradura y Extensionalidad son válidos porque, por ejemplo, dadas dos proposiciones lógicamente equivalentes, el sujeto puede reconocer a una como verdadera y no a la otra, en una situación dada. Así es que la teoría ofrece mucha libertad para la atribución de creencias al sujeto. Como era el caso en el sistema de Halpern y Fagin de la conciencia general que vimos en el tercer capítulo, el empleo de las funciones de consciencia permite enfrentar sin problemas las demandas excesivas y las trivialidades. Sin embargo, mientras no haya una restricción que regule el uso de las funciones A_i no podemos hacer frente a la demanda del rigor, es decir, la lógica no podrá ser normativa. Análogamente, la creencia como atribución impone ciertos límites a las creencias que podemos adscribir a los otros. Ahora bien, dado que ahora dichas funciones no registran simplemente los enunciados de que el individuo es consciente, sino aquellos que reconoce como verdaderos, cabría esperar que algunas reglas pudiesen derivarse de dicha interpretación. Por ejemplo, ¿es posible que un individuo reconozca la verdad de una conjunción sin reconocer la verdad de cada uno de sus conyuntos? Deslizarse por esta pendiente podría llevar a restaurar alguna forma de omnisciencia lógica (para la creencia), pero solo quiero ahora señalar con ello que aún el problema de la omnisciencia lógica no está completamente resuelto con la propuesta que he revisado en esta sección. El uso de operadores que registran la acción y efecto de las inferencias que el sujeto lleva a cabo representan un avance importante con respecto a sistemas semejantes que han aparecido antes en este capítulo. Como en el sistema de Jago, podemos modelar el empleo de reglas de inferencias que agreguen o eliminen conocimientos o creencias o incluso que hagan ambas cosas simutáneamente. De ese modo podemos también enfrentar la objeción del constreñimiento. Una regla puede prescribir aceptar el consecuente de una implicación que el sujeto reconoce y cuyo antecedente sabe, pero limitarlo a ciertas situaciones. En otras tal vez sea mejor que rechace la implicación. Podemos incluso circunscribir las reglas de inferencia a aquellas que él reconoce como correctas, aún cuando sepamos que lo conducirán al error. Por supuesto, esto supone que interpretamos las fórmulas de un sistema de manera normativa. En ese caso, las versiones C (es decir, donde el conectivo deóntico aparece solo en el consecuente. Sin embargo, es obvio que también podemos leerlas de forma descriptiva en un formato del tipo; si el sujeto cree o sabe tales cosas y hace cierta regla de inferencia, como producto de ello sabrá o creerá tal o cual. En ese sentido también es posible leerlos en la versión atributiva: si el sujeto creía tal cosa y ahora no cree tal otra es porque no hizo tal inferencia.

De cualquier modo que sea, el sistema aún en su versión más compleja adolece de los dos siguientes defectos. El primero, ya mencionado múltiples veces, es que mientras no se restrinja la libertad en que estamos de elegir los conjuntos A como nos plazca, el problema de la omnisciencia lógica no se ha resuelto completamente. Esta es la misma objeción que he señalado para los sistemas que recurren a operadores de consciencia irrestrictos. El segundo punto concierne a la naturaleza sintáctica del sistema. Si un operador prescribe o describe la acción de llevar a cabo un modus ponens con premisas $(P \rightarrow Q)$ y P, nada indica qué hará o deba hacer el sujeto si tiene $(R \rightarrow R)$ y R. Esta dificultad la encontramos en el sistema de Jago. Aún si interpretamos estas reglas como operando con esquemas y no con enunciados concretos tenemos un problema. Yo pensaría que quien cree explícitamente una proposición P, tal vez no creerá o no tiene por qué creer (explícitamente) $\neg\neg\neg\neg\neg\neg P$, pero sí $\neg\neg P$. Sin embargo, el que lo haga, o lo deba hacer, dependerá de si cuenta con las reglas correspondientes. ¿Cómo determinar cuáles son las que debe de seguir? Parece que no hemos resuelto el problema completamente.

Capítulo 5

Conclusiones

En principio, el problema de la omnisciencia lógica consiste en que en los sistemas lógicos tradicionalmente empleados para modelar la creencia (o el conocimiento) el agente cree (o sabe) todas las consecuencias lógicas de lo que cree (o sabe), lo cual no ocurre ni lejanamente con los individuos de carne y hueso. Sin embargo, podría pensarse que esto es simplemente el resultado de una idealización como es frecuente encontrarla en todas las ciencias. Mi primer objetivo ha sido aclarar esta cuestión. Para ello recurrí a las caracterizaciones de Yap y Colyvan sobre la idealización y, en mi opinión, la omnisciencia lógica no es una idealización en ninguno de los sentidos que estos autores destacan y, por lo tanto, no es tolerable en un sistema de lógica doxástica. Yap y Colyvan se refieren a idealizaciones descriptivas y normativas o a combinaciones de ambas.
Una descripción de las creencias de un agente que lo hace lógicamente omnisciente no es adecuada, ni siquiera si se pretende captar lo esencial del fenómeno de la creencia. Sostuve que tampoco lo es si el objetivo es más bien de caracter normativo. Para ello esgrimí la idea de lo que no se puede tampoco se debe. Concluí que la omnisciencia lógica sí es un problema, pero observé enseguida que no es un problema que esté adecuadamente definido. Cuando se le formula como yo acabo de hacerlo todavía no se ha delimitado adecuadamente qué se espera que sea una solución del mismo. Lo constaté con las declaraciones de quienes han intentado resolverlo y en la diversidad muy grande de sistemas que he recorrido en las páginas anteriores. Además de evitar que el agente que modelan sea lógicamente omnisciente, he notado en algunos de ellos un afán de representar más adecuadamente la creencia. Sin embargo, aún esos dos objetivos delimitan muy poco el problema.

Mi estrategia ha consistido en esclarecer los términos mismos en que está planteada la cuestión. Aunque he dicho siempre "omnisciencia", no me he enfocado en el conocimiento sino en la creencia. Por ello, una de mis primeras tareas ha sido esbozar las principales concepciones filosóficas contemporáneas de la creencia. Aunque no parece haber algo que todas compartan, hay dos vertientes que he destacado y que combinadas subyacen a nuestra noción ordinaria de creencia. La creencia puede ser considerada como un estado mental privado de cada sujeto, algo a lo que solo él puede en última instancia, tener acceso, y con lo cual busca representarse el mundo y su situación en él. En el segundo sentido la creencia es un concepto de nuestra teoría psicológica intuitiva con la que explicamos y predecimos la conducta de los demás. Otras concepciones filosóficas de la creencia se encuentran en algún punto de la línea que une esos dos polos, aunque puedan agregar características propias. Ilustré con mayor detalle esas posturas extremas con la ideas de Dretske y Fodor, por un lado, y de Dennett y Cherniak, del otro.

El término "lógica" me ha remitido a la cuestión sobre el papel que la lógica juega en relación con nuestras creencias. La construcción de algunos de los sistemas vistos parece tener como objetivo más o menos explícito describir la competencia inferencial de un individuo normal, de la misma manera que la gramática de una lengua describe la competencia lingüística de un hablante promedio de esa lengua. Pero ese no puede ser un objetivo de la lógica, porque la competencia inferencial del razonador promedio puede ir cambiando con el tiempo, mientras que las verdades de la lógica suelen pensarse como inalterables. Ateniéndome a las concepciones filosóficas de la lógica, he destacado dos vertientes que aparecen (ambas) en autores clásicos como Frege y Kant, a saber, el rol constitutivo y el rol normativo. El primero puede ilustrarse con la frase del *Tractatus*: "No se puede pensar contra las leyes de la lógica". El segundo, en cambio, queda recogido por la cita de Kant: "En lógica se trata de las leyes necesarias, no de las contingentes, no de la manera en que pensamos, sino de la manera en que debemos pensar". El primer sentido aparece en la obra de Cherniak que revisé, pues según él algunas leyes lógicas (consideradas holísticamente) determinan el que una serie de proposiciones puedan adscribirse a un sujeto como creencias suyas. Sin tal mínina coherencia no podemos considerarlo un agente racional, es decir, un creyente. Asimismo para Dennett la incapacidad para predecir o explicar la conducta de un sujeto que parece actuar de manera ilógica nos impide adscribirle creencias. En cuanto al rol normativo, parecería que la lógica debe servirnos para pensar correctamente, es decir, para formarnos ideas adecuadas de nosotros mismos y de nuestras circunstancias. Esto concuerda con la noción de creencia como representación, tal y como fue ilustrada por las concepciones de Fodor y Dretske. De esta manera, propuse asimilar el papel constitutivo de la lógica con la creencia como atribución, y su rol normativo con la creencia como representación.

De la creencia como atribución no es verosímil que haya reglas que puedan aplicarse universalmente dado el caracter holístico de los elementos que tomamos en cuenta en la adscripción de actitudes proposicionales. Sin embargo, retuve de este enfoque que las creencias de un individuo tienen un entramado lógico y que solemos atribuir a un individuo ciertas creencias que son consecuencias obvias de creencias que le han sido ya adscritas. Así es que un sistema que permita evaluar el esfuerzo cognitivo que toma derivar la conclusión de las premisas de un argumemto puede ser útil en la determinación de una lógica atributiva.

En cuanto a la creencia como representación y el correspondiente papel normativo de la lógica, he referido el debate entre quienes no consideran que la lógica tenga por función guiar el cambio razonado de creencias en la vida diaria y los que piensan lo contrario. Ambos bandos admiten que la lógica es útil en la ampliación de nuestros conocimientos en ciertas disciplinas científicas, en las que se procede de manera deductiva de premisas a conclusiones. Lo que Harman y Goldman y, más recientemente, G. Russell pusieron en duda que la lógica tuviera que ver con la forma en que cambiamos (o debemos cambiar) nuestras creencias a la luz de nueva información. Para ello ofrecieron una serie de argumentos u objeciones a la normatividad de la lógica. Algunas más fueron propuestas por Cherniak y por MacFarlane. Recuerdo al lector las más significativas:

1) Constreñimiento. La lógica no puede exigir al agente que crea una consecuencia lógica de algunas de sus creencias, si esta le parece absurda.

2) Demandas excesivas. No puede obligarse a un sujeto a que crea una consecuencia lógica oculta y compleja del conjunto de sus creencias,

3) Trivialidades. No es conveniente que el agente crea una consecuencia lógica de sus creencias, cuando dicha consecuencia es una trivialidad.

4) Rigor. La normatividad de la lógica no puede ser laxa. El sujeto está obligado a creer ciertas consecuencias lógicas de sus creencias.

5) Inconsistencia. Algunas veces es racional mantener creencias de las que sabemos que son inconsistentes, pero solo provisionalmente. Es necesario que la lógica pueda representar esta situación, pero que recomiende abandonarla pronto.

6) Prioridad. No es correcto que la lógica exija más a quien es más consciente de hechos de consecuencia lógica que a quien lo es menos.

A mi juicio todas ellas deben ser tomadas en cuenta para resolver el problema de la omnisciencia lógica, si pensamos que la lógica es normativa. Al comentar este problema suele aducirse las limitaciones cognitivas de los agentes y se piensa en la objeción de demanas excesivas. Sin embargo, si la lógica debe normar el cambio razonado de creencias, las otras objeciones son también importantes.

En ese debate me situé en un punto intermedio, al igual que MacFarlane, Field y Steinberger. Este punto consiste en aceptar que la lógica no es normativa pero que de sus enunciados verdaderos pueden derivarse de manera sistemática principios normativos. Los llamados "principios puente" son condicionales cuyo antecedente es una verdad lógica y cuyo consecuente es un principio normativo. El consecuente se genera colocando en lugares adecuados del antecedente operadores deónticos. Una variante son los principios K, que parten de que el individuo sabe un hecho de consecuencia lógica para derivar un permiso u obligación que la lógica le otorga o le impone. En mi caso, yo subsituí la $\mathcal{K}$ por la $\mathcal{K}$ (de creencia). Mi propuesta al examinar un sistema de lógica doxástica ha sido, por una parte, considerar las observaciones 1) a 6) anteriores, y, por otro, interpretar algunas de las fórmulas de su lenguaje como principios puente. Mostré que algunas estrategias seguidas en la construcción de sistemas doxásticos responden a algunas de las objeciones mencionadas y a un intento de ofrecer una mejor representación de la creencia.

Los primeros 3 sistemas examinados están basados en la idea de que una semántica en que los mundos posible de Kripke son reemplazados por situaciones modela mejor las creencias explícitas del sujeto y atenúa la objeción de las modalidades excesivas. Por un lado lo que un individuo concibe como siendo posiblemente el mundo real no valida cada proposición o su negación. Muchas proposiciones no entran en consideración en la situación en que el sujeto cree hallarse. En la lógica de Hintikka una propsición P de la que el sujeto es completamente ignorante (por ejemplo, la existencia de planetas extrasolares para un niño que no ha oído hablar de ello) se trata de la misma manera que una proposición Q de la que el individuo duda (tembló ayer en Oaxaca), a saber, habrá mundos accesibles que validen P y otros que validen su negación. Sin embargo, tenemos la intuición de que esas proposiciones debían tratarse de diferentes maneras. El agente puede imaginarse circunstancias en las que Q ocurre y otras en las que no (como siendo reales), mientras que P no entre en su espacio cognitivo. Por ello, parece natural que los nodos que forman un modelo de Kripke no otorguen a cada proposición un valor de verdad. Aparentemente también un sujeto puede tener creencias contradictorias, aunque esto es más polémico. Por ello, Levesque reemplaza mundos posibles por situaciones. Estos son conjuntos máximamente consistentes en una lógica de 4 valores de verdad, pues en una situación una proposición puede ser verdadera, falsa, ambas cosas o ninguna. Sin embargo, la lógica del sistema es clásica. Por supuesto, el sujeto no es lógicamente omnisciente de acuerdo a esta lógica. La objeción del constreñimiento puede siempre evitarse adoptando la modalidad $\mathfrak{T}$ de los principios puente. En cambio, las creencias del agente son cerradas bajo la consecuencia de la lógica FDE, que describí en el segundo capítulo, pero la exigencia que esto genera sobre él es menor porque hay un algoritmo sencillo para determinar cuándo una proposición es consecuencia-FDE de otra. El sistema no consigue evitar otras objeciones pues, por ejemplo, el agente creerá (o deberá creer, según mi lectura normativa) muchas consecuencias lógicas triviales de sus creencias. Sin embargo, el sistema aportó recursos novedosos para tratar la omnisciencia lógica y por ello fue seguido pronto por otros que buscaron mejorarlo.

Uno de ellos fue el de Lakemeyer. Dos son las novedades que aporta este autor. La primera es que en el sistema de Levesque no hay fórmulas modales de segundo grado (es decir, ningún operador de creencia puede estar en el alcance de otro) por lo que el problema de la creencia de orden superior no se trata. El sistema de Lakemeyer resuelve este problema y para ello recurre a un expediente novedoso. Así como en la semántica de Routley la verdad de la negación de una fórmula α en un mundo w, no depende del valor de α en ese mundo, sino del valor de α en un mundo w^{*}, así en el sistema de Lakemeyer el valor de verdad de una fórmula doxástica (es decir, que comienza con un operador $\mathcal{B}$) no depende de los mundos accesible a w a través de la relación R, sino de los mundos visibles desde w por medio de otra relación $\bar{R}$. Con ello, por supuesto, una fórmula doxástica y su negación pueden ser verdaderas en una situación dada, pero eso solo afecta a la creencia de orden superior porque la validez depende de la evaluación en mundos (situaciones en las cuales R y $\bar{R}$ coinciden). Desde el punto de vista de la creencia de primer grado, no hay ninguna diferencia con respecto a la lógica de Levesque, así es que valen las mismas observaciones que antes hice. Sin embargo, Lakemeyer introduce una idea interesante en el trato de la negación de proposiciones doxásticas.

La lógica de la conciencia es un sistema diseñada por Fagin y Halpern que también se propone como una mejora o refinamiento del de Levesque. Contempla varios agentes y también representa la creencia de orden superior, pero se propone para enmendar un defecto que los autores encuentran en la lógica de Levesque. En esta el agente cree una contradicción si en todas las situaciones a las que tiene acceso aparece esta contradicción. Aún admitiendo que un individuo creyese explícitamente una contradicción, parece contra-intuitiva la forma en que Levesque representa este caso. Por ello, Fagin y Halpern eliminan las situaciones contradictorias. Además introducen una forma interesante de combinar los mundos posibles y situaciones doxásticas: solo hay mundos pero las evaluaciones de las fórmulas en ellos son relativas a conjuntos de letras proposicionales. De ese modo las verdades de la lógica se preservan en todo mundo posible, pero el agente no está obligado a creerlas, ni tampoco las consecuencias lógicas de lo que cree, excepto cuando admite la implicación correspondiente. El sistema no resuelve el problema de las trivialidades y, por supuesto, no tiene modo de representar que el individuo crea una proposición contradictoria, por lo que no que no podrá guiarlo en esa situación. Esto concuerda con la intención de sus autores.

La lógica de la conciencia generalpropuesta también por Halpern y Fagin es una simplificación del sistema anterior e introduce un expediente sumamente útil. La idea central parte de un diagnóstico tácito sobre qué causa la omnisciencia lógica: el sujeto puede no haberse percatado de la relación de consecuencia lógica entre sus creencias y una proposición de la que duda o, simplemente, no es consciente de esta. Fagin y Halpern proponen agregar a un sistema de Hintikka (digamos $\mathbf{K}$) un operador de consciencia, es decir, un operador unario que semánticamente se interpreta con una función que en cada mundo posible elije arbitrariamente un conjunto de fómulas. Estos son los enunciados de que el agente es "consciente". Ahora bien, el agente cree explícitamente una proposición P en un mundo w si P es verdadera en todos los mundos accesibles desde w, y si él es consciente de P en w. Una idea similar apareció en las propuestas de Field y Steinberger. Por ejemplo, el principio (d*) de Field es:
Si es obvio que A_1, A_2,... A_n juntas implican B, entonces uno debería imponer la restricción de que $P(B)$ debe ser al menos $P(A_1) + P(A_2) + \ldots P(A_n) - (n-1)$ en cualquier circunstancia en que sea cuestión de A_1, A_2,... A_n y B.
El fin de agregar "en cualquier circunstancia en que sea cuestión de...." las premisas y la conclusión es precisamente evitar que el individuo no sea consciente de algunas de ellas. Así elude Field la objeción de las trivialidades, mientras que previene las demandas excesivas con la noción de obviedad (que habría que precisar). En cambio, el agente de la lógica de la consciencia general no está obligado a creer ninguna de las consecuencias lógicas de lo que cree y con ello las dos objeciones anteriores son evitadas. Por otro lado, el agente no puede creer una contradicción y, por lo tanto, es vacuamente explosivo. Sin embargo, como hice notar el sección correspondiente, sería necesario restringir adecuadamente la función de consciencia para que el sistema representara adecuadamente las obligaciones que un sujeto contrae a través de sus creencias. Si no, el sujeto no estará obligado a creer una consecuencia de sus creencias, aunque fuese inmediata. Sin embargo, no hay modo natural de hacer dicha restricción. Me parece que lo más obvio sería que si el sujeto es consciente de un enunciado también sea consciente de sus subfórmulas, pero con ello lo haríamos omnicreyente bajo implicación creída, es decir, restauraríamos una forma, aunque atenuada, de omnisciencia lógica. Tal vez eso fuese aceptable, en la medida en que los principios K permiten lidiar con las demandas

excesivas. Sin embargo, aún el sujeto podrá creer dos proposiciones sin estar obligado a creer su conjunción. Ahora bien, en la literatura se ha empleado de otras formas la función de consciencia que pudiesen tener restricciones más acordes a nuestras intuiciones.

Enseguida vi un sistema propuesto por van der Hoek y Meyer y que puede considerarse, en cierto sentido, como el dual de la lógica de la conciencia general. Sin enbargo, la situación que intenta modelar es muy diferente: la idea es que el agente podría no ser omnisciente porque en cada situación el agente tiene ciertos dogmas en los que cree ciegamente. Con ello, no está obligado a creer las consecuencias lógicas de sus creencias, aunque sí todas las verdades lógicas. De nuevo, sin restricciones justificadas que se impongan a la función que, en cada mundo posible, colecta los prejuicios del creyente, el sistema no ofrece una solución satisfactoria al problema.

No me detendré mucho en el sistema NLP. Lo expuse porque constituye una propuesta con respecto a los parámetros que he propuesto. Por un lado, es un sistema que tiene una lógica no clásica y que, por otro, otorga esa misma lógica al agente. La elección de la lógica FDE obedece al *desideratum* de atenuar la omnisciencia lógica como en el caso de Levesque y de rechazar la falta de relevancia de la lógica clásica.

Los dos siguientes sistemas ofrecen un ejemplo sencillo de la llamada semántica de vecindades, es decir, en la que las relaciones de accesiblidad conectan a un mundo, no con otros, sino con conjuntos de mundos (o conglomerados). El primero contiene implícita la idea de que la falta de omnisciencia lógica puede provenir de que las creencias del sujeto son relativas a dominios o campos de conocimiento. Así el sujeto puede creer una proposición P dentro de un contexto teórico, y otra, Q en otro dominio pero no haber un contexto en que crea la conjunción de P y de Q. Con ello, se anulan algunas de las manifestaciones de la omnisciencia lógica, pero no todas. En particular el sujeto debe creer todas las verdades de la lógica y las consecuencias lógicas de todo lo que cree. En cambio, no se requiere que crea las que él cree que son consecuencias lógicas de sus creencias. Esto es un resultado contraintuitivo. La lógica estaría imponiendo obligaciones por consecuencia que el agente no reconoce pero no por las que él acepta. El siguiente es un sistema dual, que enfrenta problemas similares.

Enseguida examine un sistema de Rantala en que a los modelos de Kripke se agregan mundos imposibles en los cuales las fórmulas reciben arbitariamente valores de verdad independientemente del valor de verdad de sus subfórmulas. Sin embargo, la validez es verdad en todos los mundos posibles de todos los modelos. Este es un sistema muy flexible que permite modelar una gran cantidad de situaciones. Equivale a adoptar para el sistema la lógica clásica mientras que los agentes pueden tener cualquier lógica o ninguna. Mientras no limitemos de alguna manera la libertad con que se asignan valores de verdad a las fórmulas en mundos imposibles, las creencias del sujeto pueden no tener ningún vínculo lógico entre ellas y, por lo tanto, el sistema no puede modelar ni siquiera la creencia como atribución y, por supuesto, la lógica pierde todo su caracter normativo.

Vimos que las lógicas de la justificación proveen en primera instancia un manera de registrar cómo la justificación se transmite en una cadena deductiva. En ese sentido admiten una interpretación completamente externalista. La justificación podría no ser accesible al sujeto y, por lo tanto, no imponerle obligaciones u otorgarle derechos. También podrían leerse algunos de sus teoremas como principios puente del tipo: si el individuo cree P por la razón j y que P implica Q por la razón r, entonces debería creer Q por la razón $(r \cdot j)$. En un sistema que combine una lógica de la justificacíon con una lógica doxática (como $JDK4$), el operador doxástico $\mathfrak{B}$ puede ser interpretado como representando la creencia implícita, mientras que las fórmulas del tipo $j : \alpha$ se leen como afirmaciones de que el agente cree o debe creer explícitamente α por la razón j. Por supuesto que no sería correcta una lectura de estas fórmulas desde una perspectiva completamente descriptivista, pues de que el sujeto crea $\alpha \rightarrow \beta$ por la razón s y α por la justificación t no se sigue que el sujeto crea β en virtud de $s \cdot t$, pues bien pudiera ser que no creyera β o que creyera esta proposición por otras razones. En cambio, la interpretación propuesta daría pie a una lectura normativa con principios puente del tipo $Co+$ o $Bo+$. Si pensamos que las justificaciones son falibles, el problema del constreñimiento podría enfrentarse adoptando una versión W de ciertas fórmulas. Más importante es señalar que la lógica de la justificación tiene aquí una gran ventaja. Como vimos en el ejemplo del granero rojo un individuo puede tener dos justificaciones de desigual valor para aceptar una creencia. Una de las justificaciones es factual y la otra no, pero eso por razones externas al sujeto. Incluso el sujeto puede tener razones para ceptar una proposiciones y otras para rechazarla. La lógica no lo obliga o autoriza a aceptar una creencia sin más, sino solo relativamente a ciertas justificaciones. En cuanto a las demandas excesivas la lógica de la justificación la enfrenta con un recurso muy importante. Hay una medida del esfuerzo cognitivo que toma a un agente realizar una inferencia para tener una creencia justificada. Podemos preguntarnos si una vez que tiene justificaciones para creer ciertas proposiciones qué tanto esfuerzo le lleva sacar alguna consecuencia lógica concreta de ellas. Por lo tanto, podemos medir con precisión cuantitativa qué tan verosímil es que adquiera esta creencia justificada. Con ello podemos precisar qué obligaciones impone a un sujeto una consecuencia lógica de sus creencias o qué tan necesario es que crea una conclusión (de una inferencia) para que le podamos atribuir una creencia en las premisas. En lo que se refiere a la inconsistencia, en el sistema $JKD4$ el agente no puede tener una justificación para creer una proposición si ya tiene una justificación para creer su negación, pero ello se debe a la presencia del axioma D y a la correspondiente propiedad semántica (ninguna justificación es decuada en ningún mundo para una contradicción). Otros sistemas pueden ser diseñados para que toleren la contradicción. En general, podría otorgarse al sujeto una lóigca distinta de la que tiene el sistema.

La lógica de la justificación parece bien pertrechada para responder a las objeciones de Harman. Sin embargo, como apunte antes, parece más propicia para normar el cambio de creencias en ciertos contextos tales como los científicos o, en general, aquellos en que se obtienen conclusiones a través de cadenas de inferencias muy controladas. Para que la lógica de la justificación provea una guía al creyente este requiere tener un registro de las razones que lo han llevado a sostener ciertas creencias. Como Harman muestra, aludiendo a varios estudios empíricos, el individuo ordinario no tiene ese registro y procede más bien de acuedo a patrones coherentistas.

Enseguida expuse la teoría AGM de revisión de creencias y algunos refinamiemtos de que fue objeto. Estas teorías son especialmente relevantes a mi tema por dos razones. La primera es que, desde su nombre, parecen diseñadas para resolver el problema que Harman plantea, a saber, el de modelar el cambio razonado de creencias. Por otro lado, sus principios admiten fácilmente una lectura de caracter normativo. Los postulados de la racionalidad que definen las operaciones de retracción y revisión enmarcan las directrices para cambiar de creencias de manera inteligente. Es cierto que son insuficientes pues hay operaciones que cumplen los postulados de la revisión que llevan a resultados instatisfactorios, como lo son las revisiones *maxi-choice* y *full-meet*. Sin embargo, obedecen a un principio de mutilación mínima, el cual prescribe que, ante la necesidad de realizar una modificación, hay que conservar lo más posible el cuerpo de creencias original. Esa idea es la que subyace al postulado *Recovery*, que ha sido muy debatido en la literatura. La teoría distingue entre la lógica del agente y la del sistema. En efecto la teoría supone que el operador $\mathfrak{Cn}$ satisface ciertas propiedades muy generales, pero no tiene por qué corresponder a la consecuencia clásica[1]. En cambio, al demostrar propiedades de estos operadores hemos usado la lógica clásica. Sin embargo, la teoría no elude el problema de la omnisciencia lógica y, en ese sentido, no esta-

[1]Suele suponerse que extiende o preserva la consecuencia del cálculo proposicional clásico, pero la teoría deja abierta otras posibilidades.

blece una buena normatividad. Las creencias del agente forman en todo momento un conjunto cerrado bajo consecuencia lógica. Por ello, expuse desarrollos más recientes de esta teoría, en los cuales el conjunto de creencias del agente antes o después de la revisión no tiene por qué ser cerrado. Uno de ellos es la teoría de las contracciones base en las que podemos distinguir entre creencias explícitas (que pertenecen al conjunto de creencias) e implícitas (las consecuencias lógicas de las anteriores). De esta teoría presenté dos variantes, la de las contracciones parciales y las de las contracciones kernel. Para ambas vimos que al eliminar el postulado de cerradura, es necesario modificar en consecuencia otros postulados. Para la versión de contracción parcial, en lugar de extensionalidad, tenemos uniformidad: dos enunciados que, por todo lo que el sujeto cree, son equivalentes, producirán el mismo resultado cuando hacemos la contracción con uno o con el otro. En lugar de *recovery* habrá un postulado que prescribe que solo se elimine un enunciado β de un conjunto de creencias al hacer una revisión con α, si hay un subconjunto de las creencias originales que no implica α hasta que se le agrega β. Postulados similares se requieren para la contracción kernel, una operación que es, en cierto sentido, el dual de las contracción parcial. Ahora cada una de esas contracciones da lugar a dos revisiones según que se use la identidad de Levy o la de Harper. En general, estas teorías suponen un sujeto capaz de hacer inferencias tal vez muy complejas para determinar qué conjuntos implican un enunciado y cuáles no. Sin embargo, hay dos manera de atenuar o enfrentar la objeción de demandas excesivas. Una es adscribiendo al sujeto una lógica menos demandante. La otra es suponiendo que sus creencias forman un conjunto finito. Con ello, el número de funciones selección e incisión también es finito y, en algunos casos, es verosímil que el agente pueda operar con ellas para obtener los resultados prescritos por los postulados. Por otra parte, al distinguir entre creencias implíctas y explícitas estas revisiones responden a la objeción de las trivialidades. En la sección correspondiente subrayé una ventaja de las revisiones base sobre las propuestas por la teoría AGM, a saber, que una revisión base da mayor peso a las creencias implícitas que a las explícitas, cuando se trata de conservar o eliminar creencias. Asimismo observé las virtudes y defectos del postulado de inclusión para las retracciones-base. Dado que según este principio el resultado de la operación debe ser un subconjunto del original, las directrices que marcan los postulados pueden ser muy fáciles de seguir y, en ese sentido, son normas viables. Además, con ello la teoría de los conjuntos base distingue dos conjuntos de enunciados con las mismas consecuencias lógicas, lo que es imposible en la teoría AGM. Sin embargo, el axioma de inclusión también puede dar resultados extraños porque las posibilidades se reducen notablemente. Respecto a la objeción del rigor, si no elegimos una revisión concreta sino que nos quedamos con los postulados (de algún tipo de revisión), entonces la teoría no es muy exigente con el sujeto, pues solo lo compele a elegir entre varias posibilidades. A pesar de sus varias versiones se puede hacer a la teoría de revisiones base una objeción similar a la del constreñimiento, a sabe, el resultado de una revisión base con un enunciado debe contenerlo. Sin embargo, es frecuente que un creyente rechace la información que recibe, sea porque aceptarla lo obligaría a modificar o desechar una proposición que cree con mucha firmeza o por alguna otra razón. Por ello, al final presenté muy escuetamente la teoría de las semirevisiones. Como puede apreciarse a la luz de las objeciones de Harman la teoría de revisión de creencias ha ido refinándose y, sin perder el caracter normativo de la lógica, ha modelado mejor al creyente. En ese sentido, ofrece una muy buena perspectiva para resolver el problema de la omnisciencia lógica tal como lo he ido delimitando. Sin embargo, no ofrece una representación de lo que es la creencia. A diferencia de otros sistemas que ofrecen una semántica, aquí los teoremas de caracterización proveen una manera de determinar las operaciones que satisfacen los diversos postulado, pero esto no proporciona ningún indicio de ningún concepto epistemológico.

A continuacíon expuse el sistema de Duc. Como lo dije en la sección correspondiente, este sistema no admite una lectura normativa, sino únicamente descriptiva. Sus teoremas nos informan qué proposiciones podrá conocer un sujeto a través de la inferencia lógica dado el conjunto de sus creencias. Podemos otorgarle una lógica distinta a la del sistema. Por otro lado, es un sistema que podemos considerar como dinámico y eso se refleja en que algunos de sus axiomas están tomados de la lógica temporal. Desde luego elude la omnisciencia lógica porque el sujeto modelado no sabe todas las consecuencias lógica de lo que sabe, ni las sabrá en algún momento. Lo que podemos asegurar que, dada una de estas consecuencias, podrá llegar a saberla alguna vez. El hecho de que Duc comprima en un solo operador cualquier aplicación sucesiva de inferencias simplifica mucho el sistema, pero lo vuelve menos adecuado para modelar al sujeto real.

El sistema de Agotnes y Alechina presenta importantes innovaciones en el tratamiento de la creencia o del conocimiento. Separa con toda claridad el lenguaje y la lógica del sistema del lenguaje y la lógica del agente. Las fórmulas atómicas del sistema son de la forma $B\alpha$ donde α es una proposición del lenguaje del sujeto. Es decir, el sistema solo expresa atribuciones de creencias (o de conocimientos). Sus modelos contienen “mundos posibles” que son nodos que con la relación de accesibilidad forman una red. Sin embargo, los nodos solo validan un número finito de átomos y la relación de accesibilidad representan transiciones inferenciales. Además de los conectivos lógicos usuales (incluido el operador $\diamond$ correspondiente a la relación de accesibilidad) , los autores introducen otros dos que tienen la peculiaridad de aplicarse a (y generar) conjuntos finitos de enunciados. Solo uno de ellos es definible (en términos de la conjunción). Con ese lenguaje es posible expresar con una fórmula que el agente sabe ciertas reglas de inferencia. La fórmula en cuestión representa que si el individuo saber ciertas premisas entonces con un paso de inferencia puede llegar a saber tales otras. La regla puede aumentar o disminuir el conjunto de las premisas. Ahora bien, el sistema no resuelve qué lógica hay que atribuir al agente, sino deja abierta la posibilidad de aplicarlo de diferentes maneras. Al parecer Agotnes y Alechina conciben su sistema como un instrumento para modelar la competencia razonante de un agente. También podrían sus fórmulas ser leídas como principios puente de la modalidad $\mathfrak{C}\mathcal{R}+$ o $\mathfrak{C}\mathcal{O}+$. Dado que no se sugiere que lógica podría tener el razonador, el problema de la omnisciencia lógica, tal y como yo lo estoy entendiendo aquí, no se reuelve. Sin embargo, la propuesta de Agotnes y Alechina introduce artificios importantes y novedosos en el tratamiento del problema.

Enseguida expuse el sistema de Jago. Su autor se propuso modelar la manera en que las creencias de un individuo evolucionan por la aplicación sucesiva de reglas de inferencia. Como lo dice al inicio de su artículo, para él no se trata únicamente de evitar que las creencias del agente sean cerradas bajo consecuencia lógica, sino también de enfrentar el problema de las demandas excesivas (o, mejor, de las limitaciones cognitivas del agente) y de las trivialidades. En su forma general, el sistema tiene algunas semejanzas con el de Agotnes y Alechina. Contiene también los dos lenguajes ya mencionados y modelos como los del sistema anterior, excepto que los mundos pueden validar además de proposiciones atómicas (tal vez una infinidad), reglas de inferencia. Estas son reglas completamente sintácticas que aplicadas a un conjunto de proposiciones generan otro. Hay reglas que generan nuevas creencias (o conocimientos) y otras que eliminan algunas que tenía el agente. Más adelante di un ejemplo semejante al que presenta Jago en su artículo de un sistema con enunciados específicos y un conjunto fijo de reglas concretas (es decir, sin esquemas). Este sistema particular se asemeja a un programa de cómputo en el que en una situación particular puede haber algunas reglas aplicables y el sujeto tiene la libertad de elegir cuál emplear. Una vez escogida una regla, esta determina a qué estado pasar. Jago ofreció un sistema axiomático, pero no de las inferencias que puede hacer el sujeto, sino de la forma en que operan los modelos. Los problemas de demandas excesivas y de trivialidades se evitan fácilmente porque las reglas son algoritmos de transformación de expresiones que, a partir de un conjunto de premisas, generan a lo más un resultado. El problema del constreñimiento también puede resolverse fácilmente gracias a la presencia de reglas aumentativas y diminutivas. Una regla podría recomendar al sujeto adoptar una creencia y otra prescribiría que al llegar a ese resultado, se elimine y se revisen las premisas que llevaron a él. Claramente el sistema satisface la demanda de rigor. El mayor inconveniente es que los algoritmos que acepta el creyente de ninguna manera tienen que ser catalogados como principios lógicos ni como principios puente. Un sistema así puede ser diseñado para cierto tipo de actividades muy específicas como en el ejemplo que di. Aunque se trate de simplemente reglas para modificar un conjunto de creencias, no tienen por qué provenir de la lógica. Por otro lado, el sistema no ofece ninguna representación de la creencia.

A continuación presenté la propuesta de Rasmussen diseñada para representar cómo se modifica el conjunto de conocimientos de un individuo por la aplicación sucesiva de reglas de inferencia. Aunque el autor pretendía utilizar mundos posibles para modelar los conocimientos del sujeto en un momento dado, su sistema es solo sintáctico. Además de evitar la omnisciencia lógica, pretende otorgar al sujeto modelado algunas capacidades inferenciales. A mi parecer, retomó ideas de Duc que en el sistema de este finalmente no fueron desarrolladas (en aras de un simplificación). Podemos concederle al creyente el conocimiento de las reglas de inferencia que nos plazca, pero no será omnisciente con respecto a ellas. Una fórmula dirá que si conoce las premisas de una de esas reglas entonces sabrá la conclusión, si hace la inferencia correspondiente. Por ello, no podemos interpretar las fórmulas del sistema como siendo normativas. Con cada regla de inferencia hay un superíndice numérico que registra el esfuerzo cognitivo que toma realizarla. Rasmussen no dice cómo este índice podría determinarse, pero así encara la objeción de las demandas excesivas. Si una cadena de inferencias requiere un esfuerzo cognitivo muy grande, no atribuiremos a un sujeto un conocimiento de ella, por el hecho de conocer sus premisas, ni esperamos que lo obtenga pronto. Con ello, como dije, podemos darle al sistema una lectura normativa. Podría exigírsele a un sujeto el conocimiento de las consecuencias obvias (es decir, que el obtenerlas demanda muy poco esfuerzo cognitivo) de sus conocimientos manifiestos. Sin embargo, el sistema no afronta en esta lectura objeciones tales como el constreñimiento ni las trivialidades. Tampoco ofrece una representación adecuada de la creencia.

El sistema de Bjerring y Rasmussen pretende retomar las virtudes del sistema anterior y subsanar algunas de sus deficiencias. Para sus autores la tarea prescriptiva de la lógica es clara y sirve de marco al problema de la omnisciencia lógica (en el caso de la creencia). El objetivo es normar la actividad deductiva del agente tomando en cuenta sus limitaciones cognitivas sin hacerlo lógicamente incompetente. Es un enfoque que intenta conciliar dos muy diferentes usos de los mundos posibles. En efecto, según hemos visto, la relación de accesibilidad puede representar sea las alternativas epistémicas del agente, es decir, las situaciones que este concibe como siendo posiblemente reales, sea las transicionces de un conjunto de creencias a otro por medio de inferencias. Para combinar ambos enfoques, los autores utilizan artificios un tanto rebuscados. El resultado es que el agente creerá α después de n pasos de inferencia si α puede ser derivado en ese número de pasos (por medio de las reglas de inferencia concedidas al agente) en cada mundo que concibe como siendo posiblemente real, los cuales pueden incluir mundos imposibles. Se determinará un número natural n (que dependerá del conjunto de inferencias en cuestión) de tal manera que una fórmula que pueda ser obtenida en menos que n pasos a partir de las creencias del sujeto, será considerada una inferencia fácil. Esto recoje la idea de consecuencia obvia de que hablaban Field y Steinberger. De esta manera se evita que las creencias del sujeto sean cerradas bajo consecuencia según la lógica del agente. Solo tendremos que si α es consecuencia lógica de un conjunto Γ de proposiciones que el agente cree, entonces habrá un cierto número m tal que después de m pasos creerá α. Por otro lado, solo las consecuencias fáciles impondrán alguna obligación. Con ello podemos formar principios puente de la forma $\mathfrak{C}\mathcal{O}+$. Así se enfrentan las objeciones de las demandas excesivas y de las trivialidades. El problema de la inconsistencia también puede evitarse porque las creencias del individuo pueden ser contradictorias (hay mundos imposibles), pero eso dependerá de qué lógica se le conceda. El problema del rigor queda contemplado por el hecho de que las consecuencias fáciles pueden exigírsele al creyente. Por último, la objeción del constreñimiento pude ser solventada pasando de la lectura $\mathfrak{C}\mathcal{O}+$ a $\mathfrak{T}\mathcal{O}+$, como hemos visto en otros casos. Desde luego que todo está contemplando pero solo esquemáticamente, a comenzar por la determinación de la cota para que una cadena de inferencias sea sencilla. Queda también determinar qué infrencias verosímiles atribuirle al agente, entre otras cosas. Lo que me interesa subrayar es que los autores avanzan en la dirección que he defendido en este trabajo.

Por último revisé un tipo de sistemas que se forman por la superposición de tres recursos: los modelos de verosimilitud o plausibilidad de Baltag y Smets, las funciones de reconocimiento (o consciencia) y los modelos de acción. Aunque hay muchas variantes de este tipo, ilustré las principales ideas con una propuesta de Velázquez-Quesada. Me parece importante subrayar que los modelos de verosimilitud constituyen una manera novedosa de representar la creencia y, a mi juicio, más adecuada. Es usual suponer que el conocimiento es una creencia con ciertas características extrínsecas. Por ejemplo, el conocimiento es una creencia verdadera (entre otras cosas), pero esto le viene por su concordancia con la realidad, que es algo externo a la creencia. En la propuesta de Baltag y Smets la creencia es distinta del conocimiento. En estos modelos las creencias están escalonadas por la adhesión que el sujeto les otorga. Esta característica había aparecido en otros sistemas. Por ejemplo, en la teoría AGM se introdujeron funciones de atrincheramiento que señalaban aquellas creencias a las que el sujeto estaba menos dispuesto a renunciar. Vimos también un sistema de van der Hoek y Meyer en que ciertas creencias del sujeto son eximidas de cualquier revisión en un momento dado (pues son los dogmas del sujeto). Me parece más atractiva la forma en que Baltag y Semts modelan este fenómeno porque permiten gradaciones en las creencias. Para evitar la omnisciencia lógica en los modelos de verosimilitud se agrega una función de consciencia, que puede ser intepretada de diversas maneras. En el caso que vimos la función selecciona las creencias de que no solamente el individuo es consciente sino que reciben su asentimiento (sean o no realmente verdaderas). Con ello fue posible distinguir creencias implícitas y explícitas. Hay varias formas de modelar en los modelos de verosimilitud el cambio de creencias cuando el sujeto recibe nueva información que aún no puede tener por cierta. En la que presenté se introduce una operación análoga al anuncio público llamada "actualización radical" y que consiste en una reordenación de los mundos que el sujeto estima como siendo posiblemente reales. En una actualización con α el orden antiguo se conserva en lo posible, excepto que los mundos en que α es verdadera se volverán más verosímiles que aquellos en que α es falsa. Enseguida expuse cómo Velázquez-Quesada combina este expediente con la mencionada función de reconocimiento para modelar el cambio de las creencias explícitas de un sujeto a partir de una actualización. Por último mostré cómo dicho autor amalgama todos estos recursos con los modelos de acción de Baltag, Moss y Solecki para modelar, por ejemplo, cómo evolucionan las creencias de un individuo por aplicación de reglas de inferencia. Con todos estos recursos es posible representar el conjunto de creencias de un agente y los cambios que sufre sea por información que recibe y que es confiable, aunque no segura, sea por el efecto de inferencias que lleva a cabo. Sin embargo, para evaluar qué tan bien se enfrenta con ello el problema de la omnisciencia lógica tal y como yo lo he enmarcado aquí, es necesario desglosar los diversos componentes que entran en juego. como dije, creo que la representación de las creencias de un agente y la forma en que se modifican por efecto de una actualización me parece concuerdan mejor con nuestras intuiciones. Sin embargo, para eludir el problema de la omnisciencia lógica es necesario introducir un operador de reconocimiento y aquí reitero las observaciones que hice con el operador de consciencia de Halpern y Fagin: mientras no quede adecuadamente limitado su uso en la semántica, el problema de la omnisciencia lógica no se ha resuelto. Si no lo sometemos a ninguna restricción las creencia explícitas del agente no tiene suficiente estructura lógica para considerarlas como tales, es decir, como las creencias de un agente racional. Hasta aquí las fórmulas del sistema admiten una lectura normativa. No ocurre lo mismo con los modelos de acción. Decir que si el sujeto cree ciertas premisas y realiza tal regla de inferencia concreta entonces el conjunto de sus creencias se modificará de tal modo, es simplemente una descripción de la regla, no de lo que el sujeto debe realizar en un circunstancia determinada.

Creo haber mostrado en este trabajo cómo ciertas discusiones filosóficas pueden tener una incidencia en delimitar y resolver problemas técnicos que se plantean en la lógica epistémica. La forma en que he abordado la cuestión de la omnisciencia lógica explica cómo la lógica sin soslayar su objetivo de estudiar relaciones abstractas entre proposiciones juega, sin embargo, una papel fundamental en normar la forma en que tratamos y evaluamos la información que recibimos y transmitimos en nuestra interacción con lo otros.

En la última década ha habido una tendencia a acercar la lógica a agentes reales de la misma manera que la economía y la teoría de juegos ha dejado de concentrarse en agentes ideales. Van Benthem por ejemplo, concibe a la lógica no como un sistema abstracto de relaciones entre proposiciones sino como un componente que interviene de forma frecuente en las interacciones de los seres humanos con su entorno. El esfuerzo por evitar la omnisciencia lógica forma parte de esa tendencia. Sin embargo, puede pensarse que la lógica se encarga de describir la competencia razonante de los seres humanos reales. He mostrado como esta variante de la lógica no es incompatible con su caracter normativo. Me parece que la lógica no sólo debe desarrollarse en este sentido más sensible a nuestras capacidades cognitivas reales sino que debe hacerlo en estrecha cercanía con el pensamiento filosófico. A lo largo de este trabajo he sostenido que hay posturas filosóficas que contribuyen a delimitar mejor ciertos problemas lógicos y con ello se puede avanzar hacia su solución.

Índice alfabético

Bibliografía

[1] Agotnes, T. y Alechina, N.. "The dynamics of syntactic knowledge". *Journal of Logic and Computation*, 17(1): 83-116, 2007.

[2] Alchourrón, C. E., Gärdenfors, P. y Makinson, D. (1985)."On the logic of theory change: Partial meet contraction and revision functions". *The Journal of Symbolic Logic*, 50(2): 510-530.

[3] Aristóteles (1989), *Aristotle, Prior Analytics*, Smith Robin (Ed.), Hackett Pub Co Inc.

[4] Artemov, S. (2008),"The logic of justification", *Review of Symbolic Logic*, 1 (4): 477-513.

[5] Artemov,S. Y Kuznets, R. (2006) "Logical Omniscience via proof complexity". En *In Computer Science Logic 2006, Lecture Notes in Computer Science, Vol. 4207*. (pp.135-149) Springer.

[6] Artemov, S. y Nogina, E. (2005) "On epistemic logic with justification", Conference: *Proceedings of the 10th Conference on Theoretical Aspects of Rationality and Knowledge* (TARK-2005), Singapore, June 10-12.

[7] Artemov, S., Kuznets (2009) R.,"Logical omniscience as a computational complexity problem", en A. Heifetz (ed.), *Theoretical Aspects of Rationality and Knowledge, Proceedings of the Twelfth Conference* (TARK 2009), ACM Publishers: 14-23.

[8] Artemov, S. (2011),"Why do we need justification logic?", in *Semantic Scholar*.

[9] Baltag, L. S. Moss, and Solecki, S. (1998),"The logic of public announcements and common knowledge and private suspicions". En I. Gilboa, editor, TARK: 43-56, San Francisco, CA, USA, Morgan Kaufmann.

[10] Baltag, A, y Smets, S. (2008) “A qualitative theory of dynamic interactive belief revision". In G. Bonanno, W. van der Hoek, and M. Wooldridge, editors, *Logic and the Foundations of Game and Decision Theory* (LOFT7), volume 3 of Texts in Logic and Games: 13-60. Amsterdam University Press.

[11] van Benthem, J.(2007), “Dynamic logic for belief revision". *Journal of Applied Non-Classical Logics*, 17(2):129-155.

[12] Bjerring, J. C y Rasmussen, M. S (2019) “A dynamic solution to the problem of logical omniscience". *Journal of Philosophical Logic*, volumen 48, issue 3, junio 2019, pp 501-521.

[13] Broome, J., (1999), “Normative requirements”, *Ratio*, 12 (1999), pp. 398-419,

[14] de Bruin, B. (2008), “Epistemic Logic and Epistemology” en Vincent Hendricks y Duncan Pritchard (Eds.) *New Waves in Epistemology*, Palgrave-Macmillan.

[15] Chan, Th. (Ed.) (2014) *The Aim of Belief*, Oxford University Press.

[16] Cherniak, C. Minimal Rationality, Cambridge MA, The MIT Press/A Bradford Book, 1986.

[17] Carroll, L. (1895). “What the Tortoise Said to Achilles”. *Mind*, 104 (416): 691-693.

[18] Collins, J. (2014), “Epistemic Closure Principles”. En J. Fieser y B. Dowden (eds.), *The Internet Encyclopedia of Philosophy*.
URL: http://www.iep.utm.edu/epis-clo/.

[19] Colyvan, M. (2012). Idealisations in normative models. *Synthese*, 190, 1337-350.

[20] Davidson, D. (1980), *Essays in actions and events*, Oxford University Press, Oxford, 1980.

[21] Davidson, D. (1985), *Inquiries into Truth and Interpretation*, Oxford, Clarendon Press, 1985.

[22] Dennett, D. *Brainstorms: Philosophical Essays on Mind and Psychology*, The MIT Press/A Bradford Book, 1978.

[23] Dennett, D. *The Intencional Stance*, The Massachusetts Institute of Technology, 1987.

[24] Dennett, D. *Intuition Pumps and Other Tools for Thinking*, W. W. Norton and Company, 2013.

[25] van Ditmarsch, H., van der Hoek,W. y Kooi, B. *Dynamic Epistemic Logic*, Vol. 337 of Synthese Library Series. Springer, Dordrecht, The Netherlands, 2007.

[26] Dretske, F. (1970), "Epistemic operators", *The Journal of Philosophy*, Vol. 67, No. 24:1007-1030.

[27] Dretske, F. (1971), "Conclusive reasons", *Australasian Journal of Philosophy*, Vol. 49: 1-22.

[28] Dretske, F. (1988) *Explaining Behaviour*, Cambridge, MA, MIT Press.

[29] Duc, H. N. (1997)."Reasoning about rational, but not logically omniscient, agents". *Journal of Logic and Computation*, 7(5):633.

[30] Fagin, R. Halpern, J.Y. (1985) "A formal model of knowledge, action and communication in distributed systems. En Proceedings of the fourth anual ACM symposium on principles of distributed computing, pp. 224-236.

[31] Fagin, R., Halpern,J. Y. y Vardi, M., "A non-standard approach to the logical omniscience problem" in *Proceedings of the Third Conference on Theoretical Aspects of Reasoning About Knowledge* (TARK-90) 1990, pp. 41-55.

[32] Fagin, R. Halpern, J.Y. (1988),"Belief,awareness and limited reasoning" *Int.Joint Conf. Artif. Intell.* 34 : 39-76.

[33] Field, H. (2009). "What is the Normative role of Logic?", *Proceedings of the Aristotelian Society*, 83. pp. 251-268.

[34] Fodor, J. (1987) "Psycosemantics. The problem of Meaning in the Philosophy of Mind", Cambridge MA, MIT Press, Pag 1-26. Traducido como: "La persistencia de las actitudes", En: Rabossi, (comp), *Filosofía de la Mente y Ciencia Cognitiva*, Barcelona, Paidós, 1995, p.p 69-101.

[35] Foley, R. (1993). *Working Without a Net: A Study of Egocentric Epistemology, New York*, Oxford University Press.

[36] Frege, G., (1979) *Posthumous Writings*, Hermes, Kambartel, Kaulbach (Eds), University of Chicago Press.

[37] Frege, G. (1918) (1956). “The thought”, *Mind* 65 (259), pp. 289-311.

[38] Davidson, D. (1982) (1984). “Communication and Convention”, *Synthese*, 59, pp. 3-17. Publicado asimismo en Davidson, D. *Inquiries into Truth and Interpretation*, pp. 265-80. Oxford: Clarendon Press, 1984.

[39] García Campos, J. (2009), *Epistemología y Psicología cognitiva. Un Acercamiento al Estudio de la Justificación*, *Centro de Estudios Filosóficos, Políticos y Sociales Vicente Lombardo Toledano*.

[40] Gettier, E. L. (1963). “Is Justified True Belief Knowledge?” Analysis 2, pp. 121-3.

[41] Gillet, E., Gochet, P. *La Logique de la connaissance: le probleme de l’omniscience logique*, Université de Liege, Séminaire de Logique et Epistémologie, 1990.

[42] Goldman, A. (1967), “A causal theory of knowing”. *The Journal of Philosophy*, 64: 335-372.

[43] Goldman, A. (1986). *Epistemology and Cognition*, Cambridge, Mass.: Harvard University Press.

[44] Halpern, J. Y., “Reasoning about knowledge: An overview” Processings of the First Theoretical Aspects of Reasoning About Knowledge (TARK-86), Morgan Kaufmann, 1986, pp. 1-18.

[45] Hansson, S. O. (1993) “Reversing the Levi identity”. Journal of Philosophical Logic, 22, pp. 637-669.

[46] Hansson, S. O, (1994) “”Kernel contraction”. *Journal of Symbolic Logic*, 59(3), pp. 845-859.

[47] Hansson, S. O. (1997). “Semi-revision”, *Journal of Applied Non-Classical Logics*, 7(2).

[48] Harman. G. (1986) *Change in View*. The MIT Press.

[49] Hintikka, Jaakko, (1962) (2005), *Knowledge and Belief: An Introduction to the Logic of the Two Notions*, second edition, Vincent F. Hendriks and John Symons (eds.), (Texts in Philosophy, 1), London: College Publications.

[50] Hintikka, J. (1962) *Knowledge and Belief*, Cornell University Press, Ithaca, NY.

[51] Hintikka, J. (1975)"Impossible possible worlds vindicated" *Jornal of Philosophical Logic* 4: 475-484.

[52] Hocutt, M. (1972). Is epistemic logic possible? *Notre Dame Journal of Formal Logic*, XII, I (4), 433-453.

[53] van der Hoek W., Meyer Ch. (1989) "Possible Logics for Belief", in *Logique at Analyse* 127-128, 1989: 177-194.

[54] Jaspars, J. O. M (1991),"Fused Modal Logic and Inconsistent belief", in M. De Glas, Gabbay (Eds.) *Proceedings of the 1st World Conference on the Fundamentals of Artificial Intelligence* (WOCFAI?91), París: 267-275.

[55] Jaspars, J. O. M.(1993) "Logical Omniscience and Inconsistent Belief", in M. De Rijke (ed.) *Diamonds and Defaults*, Kluwer Academic Publishers, Dordrecht: pp. 129-146.

[56] Jago, M. (2009). "Epistemic logic for rule-based agents". J*ournal of Logic, Language and Information*, 18(1):131-158.

[57] Kant, I. (1800) (1885). *Kant's Introduction to Logic and His Essay on The Mistaken Subtetly of the Four Figures*, Longman, Green.

[58] Kneale, W. y Kneale, M. (1980), *El Desarrollo de la Lógica*, Técnos.

[59] Konolige, K. (1986), *A Deduction Model of Belief*, Pitman/Morgan Kaufmann, Londres, Los Altos, 1986.

[60] Kripke, S. A. (1963) "Semantical analysis of modal logic I: Normal model propositional calculi" *Zeitschrift fur mathematische Logik und Grundlagen der Mathermarik*, 9, 67-96.

[61] Kripke, S, A, (1989) *Wittgenstein: Reglas y Lenguaje Privado*, Universidad Nacional Autónoma de México.

[62] Kyburg, H., (1961), *Probability and the Logic of Rational Belief*, Wesleyan University Press.

[63] Lakemeyer, G. (1987)"Tractable meta-reasoning in propositional logics of beliefs" *International Joint Conference in Artificial Intelligence*: 402-408.

[64] Lehrer, K., Paxson, T. (1969). Knowledge: undefeated justified true belief. *The Journal of Philosophy*, 66, 1-22.

[65] Levesque, H. (1986), "Logic of implicit and explicit belief", in *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*: 389-393.

[66] Levi, I,(1991) The Fixation of Belief and Informational Value, *Cambridge University Press*, 1991

[67] MacFarlane, J. (2004), "In what sense (if any) is logic normative for thought?", Versión del 21 de abril de 2004 para presentación en the Central Division APA.

[68] Marian, D., Warfield, D. (2008), "Knowledge-Closure and Skepticism", en Quentin Smith (Ed.) *Epistemology: New Essays*, Oxford University Press. pp. 137-196.

[69] Milne, P. (2009) "What Is the Normative Role of Logic?", *Proceedings of the Aristotelian Society*, Supplementary Volumes , 2009, Vol. 83, pp. 269-298.

[70] Moore, R.C. y Hendrix, G. (1979) "Computational Models of Beliefs and the Semantics of Belief-Sentences", Technical Note 187, SRI International, Menlo Park.

[71] Parikh, R., (2008), "Sentences, Belief and Logical Omniscience", or What Does Deduction tell us? *The Review of Symbolic Logic*, Vol, 1 no. 4.

[72] Platón (1980), *Gorgias*, Universidad Nacional Autónoma de México.

[73] Plaza, J. (1989)."Logics of public communications". en M. L. Emrich , M. S. Pfeifer, M. Hadzokadic, y Z. W. Ras (editores),*Proceedings of the 4th International Symposium of Methodologies for Intelligent Systems*, pǵinas 201-216.

[74] Priest, G. (2006). *Doubt Truth to Be a Liar*, Oxford University Press.

[75] Quine, W V. (1950) (1972), *Methods of Logic*, cuarta edición, Harvard University Press.

[76] Rantala, V. (1982) “Impossible worlds semantics and logical omniscience”, *Acta Philosophica Fennica*, 35: 18-24.

[77] Rantala, V., (1982)“Quantified modal logics: Non-normal worlds and propositional atti- tudes”, *Studia Logica* 41: 41-65.

[78] Rasmussen, M. S., (2015).“Dynamic epistemic logic and logical omniscience". *Logic and Logical Philosophy*, 24:377?399.

[79] Rasmussen, M. S. y Bjerring, J. C, (2019) “A dynamic solution to the problem of logical omniscience". *Journal of Philosophical Logic*, volumen 48, issue 3, junio 2019, pp 501-521.

[80] Rott, H. (2000) “Two dogmas of belief revision”, *Journal of Philosophy*, 97 (9); 503-22.

[81] Russell, G. (2020) “Logic is not normative”. *Inquiry*, 63:34,371-388.

[82] Russell, Gillian and Christopher Blake-Turner,(2023)“Logical Pluralism”, The Stanford Encyclopedia of Philosophy (Fall 2023 Edition), Edward N. Zalta and Uri Nodelman (eds.) url: https://plato.stanford.edu/archives/fall2023/entries/logical-pluralism/

[83] Sainsbury, M. (2002), “What logic should we think with?”, *Royal Insitute of Philosoiphy Supplement*, marzo 2002.

[84] Schwitzgebel, E.,“Belief”, *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition), Edward N. Zalta (ed.)
URL = <https://plato.stanford.edu/archives/win2021/entries/belief/>.

[85] Samuels, R., Stich, S., Faucher, L. (2004). *Reason and Rationality. Handbook of Epistemology*, I. Niiniluoto; M. Sintonen and J. Wolenski (Eds.) (pp. 1-50), Kluwer.

[86] Singer, P., (1997), *Repensar la Vida y la Muerte. El Derrumbe de la Ética Tradicional*, Paidós.

[87] Smullyan, R., (1988). *Forever Undecided*, Oxford Paperbacks.

[88] Stein, Howard (2011) *Without Good Reasons. The Rationality Debate in Philosophy and Cognitive Science*, Oxford Scholarship (Online).

[89] Steinberger, F, (2020), "The Normative Status of Logic", *The Stanford Encyclopedia of Philosophy* (Edición invierno 2020), Edward N. Zalta (ed.), url: https://plato.stanford.edu/archives/win2020/entries/logic-normative/

[90] Tarski, A.(1936) (1983), "The Concept of Truth in Formalizad Languages", en *Logic, Semantics and Metamathematics*, editado por John Corcoran, Indianapolis, Hackett.

[91] Vardi, M., "On the complexity of epistemic logic" in *Logic in Computer Science* 1989, 1989. pp. 253-262.

[92] Vardi, M., "On epistemic logic and logical omniscience", in *Proceedings of the First Conference on Theoretical Aspects of Reasoning About Knowledge*, Halpern, Eds. 1986, pp. 293-306.

[93] Velázquez-Quesada, F. R. (2014), "Dynamic Epistemic Logic for Implicit and Explicit Beliefs", *Journal of Logic, Language, and Information*, Spring 2014, Vol. 23, No. 2, FORMAL MODELS OF AWARENESS, (spring 2014), pp. 107-140.

[94] Velázquez-Quesada, F. R. (2015) "Epistemic Dynamics for Non-ideal Agents", Lecture Notes.

[95] Velleman, J. D., (2000), "On the aim of belief", in J. D. Velleman, The Possibility of Practical Reason, Oxford University Press, pp. 244-82.

[96] Weisberg, M., "Three kinds of idealization", *Journal of Philosophy*, 2007, 104 (12), 639-659.

[97] Wittgenstein, L. (1921) (1980) *Tractatus Logico-Philosophicus*, Alianza Universidad.

[98] Wittgenstein, L. (1958) (1988) *Investigaciones Filosóficas*, UNAM-Crtíca.

[99] Yap, A., "Idealization, epistemic logic, and epistemology", *Synthese*, 2014, 191: 3351-3366.