TRABAJO DE FIN DE GRADO

Lógica Epistémica Dinámica

Propuestas para enfoques probabilísticos y aplicaciones

Realizado por

Alexander Romero Vinogradov

Para la obtención del título de
Doble Grado en Matemáticas y Estadística

Dirigido por
D. Andrés Cordón Franco

Realizado en el departamento de
Ciencias de la Computación e Inteligencia Artificial

Convocatoria de julio, curso 2022/23

Agradecimientos

Agradezco a:

- Mis padres, por haberme traído a este mundo, por haberme proporcionado siempre su cariño, su amor y la mejor educación posible, y por haberme mantenido económicamente mientras escribía textos absurdos sobre formalismos extraños.
- Mi novia, por haberme apoyado siempre con su cariño y su amor incondicional, por cuidar de mí cuando lo necesito, y por seguir apoyándome y cuidándome incluso mientras escribo textos absurdos sobre formalismos extraños.
- Mi tutor, Andrés Cordón Franco, por haberme guiado magníficamente en la composición de este trabajo, por haber dirigido mi atención a fuentes sugerentes y útiles que han inspirado muchas de mis ideas en el mismo, y por haber sido tan simpático y amable durante todo el proceso.
- El resto de los profesores que he tenido a lo largo de mi carrera, por habérme proporcionado una formación imprescindible para escribir un trabajo como este. En particular, a Luis Valencia Cabrera por haberme puesto en contacto con Andrés.
- Mi amigo Juan Luis, por haber sido el primero en mencionar la "lógica intensional de Kripke" en mi presencia, hecho que sin lugar a dudas plantó en mi cabeza una semilla de curiosidad sin la cual probablemente este trabajo hubiese contado con un entusiasmo mucho menor por mi parte.
- Douglass R. Hofstadter, célebre polímata, por haber escrito un magnífico libro que siempre constituirá una de mis inspiraciones fundamentales, aunque solo sea indirectamente, en esta clase de ámbitos (y también en otros).
- David Martínez Rubio, a quien pude considerar mi mentor durante un breve periodo (aunque probablemente él no me considere su aprendiz), por habérme recomendado un magnífico libro de Douglass R. Hofstadter, por habérme inculcado el hábito de contar en binario con los dedos de las manos, y por haber sido, en general, otra de mis grandes inspiraciones en esta clase de ámbitos (y también en otros).
- Y a muchas más personas a las que también debería agradecer, pero que harían de esta lista algo interminable, por haber contribuido conjuntamente a definir las actitudes hacia el mundo que me hacen ser la persona que soy.
- ¡Y, por último, a mí mismo! *(Y de nuevo a mi novia por darme la idea de acabar así mis agradecimientos.)*

Resumen

En este trabajo hacemos un repaso de algunas de las principales propuestas en el área de la lógica epistémica dinámica, rama de la lógica modal: lógica epistémica (interpretada sobre la clase semántica $\mathcal{S}5$), lógica epistémica con anuncios públicos y lógica epistémica probabilística. Se exponen propiedades básicas y conceptos de interés para cada uno de estos sistemas, y se discuten en el contexto de ejemplos, muchos de ellos clásicos, que ponen de manifiesto su relevancia en diversos ámbitos de la realidad práctica. En aras de integrar los sistemas anteriormente mencionados, y en partícular la lógica epistémica con anuncios públicos y la lógica epistémica probabilística, se propone una generalización de los anuncios públicos que es coherente con los aspectos probabilísticos de nuestro lenguaje, y se estudian también las propiedades teóricas básicas de esta propuesta. Se discuten varias conjeturas, de las que se espera obtener resultados adicionales en investigaciones futuras, sobre el comportamiento formal de esta nueva construcción.

Abstract

In this monograph we review some of the main proposals in the area of dynamic epistemic logic, a branch of modal logic: epistemic logic (interpreted on the semantic class $\mathcal{S}5$), epistemic logic with public announcements and probabilistic epistemic logic. We set forth some of the basic properties and relevant concepts for each of these systems, which are discussed in the context of examples (many of which are classical problems in the field) which showcase their relevance in several scopes of application. For the sake of integrating the previous systems, and in particular epistemic logic with public announcements and probabilistic epistemic logic, a generalisation of public announcements which is coherent with the probabilistic aspects of our language is proposed, and the basic theoretical properties of this proposal are studied. We discuss some conjectures, for which aditional developments through further research are expected, about the formal behaviour of this new construction.

Índice general

1. Introducción

El presente trabajo se enmarca dentro del campo de la lógica modal y, más concretamente, de la lógica epistémica multiagente. La lógica modal estudia, en líneas muy generales, el comportamiento deductivo de las expresiones "es necesario que" y "es posible que". En diversos contextos, dichas expresiones adquieren diversos significados y connotaciones, reflejándose estos en las propiedades que manifiestan como parte de los correspondientes sistemas formales, pero con ciertas propiedades comunes: típicamente son extensiones de la lógica proposicional mediante nuevos *operadores modales* de necesidad ($\Box$) y de posibilidad ($\Diamond$). En la literatura del campo aparece una gran variedad de interpretaciones distintas para estos operadores modales; si A es una fórmula arbitraria, por ejemplo, algunas de las interpretaciones de la fórmula $\Box A$ son:

- A es éticamente obligatorio (lógica deóntica)
- A es deseado por alguien (lógica volitiva)
- Siempre será el caso que A (lógica temporal en modalidad futura)
- Siempre ha sido el caso que A (lógica temporal en modalidad pasada)

En este sentido, la interpretación en la que se enmarca nuestro trabajo sería la siguiente: "$\Box A \equiv$ Un determinado agente conoce A".

En líneas muy generales, nuestro punto de partida será el estudio de la lógica epistémica multiagente $\mathcal{S}5$, sistema fundamental y bien establecido en el campo, y, a partir de aquí, consideraremos tres extensiones del mismo, con un especial énfasis en la interacción entre sistemas de lógica epistémica y sistemas probabilísticos. Más concretamente, en el presente trabajo se estudiará:

- la lógica epistémica dinámica con anuncios públicos $\mathcal{L}_{K[]}$,
- la lógica epistémica probabilística $\mathcal{LP}_{\mathcal{K}}$, y
- la lógica epistémica probabilística dinámica $\mathcal{LP}_{\mathcal{K}[]}$ (que integra los dos sistemas anteriores).

Aunque, por supuesto, estamos muy interesados en las propiedades matemáticas y meta-matemáticas de estos sistemas y mencionaremos sus propiedades básicas a lo largo del trabajo, nuestro enfoque será más conceptual y nuestro principal objetivo será describir con detalle la semántica de dichos sistemas, motivar su introducción y proporcionar al lector una serie de ejemplos para que pueda forjar una buena intuición sobre el uso de dichos sistemas y cómo podrían usarse para modelar situaciones donde el razonamiento epistémico y probabilístico pueda ser de interés.

Durante nuestro estudio de la literatura del campo, detectamos una leve incongruencia en una de las propuestas clásicas para integrar los razonamientos epistémicos con los probabilísticos: siendo la lógica epistémica probabilística dinámica con anuncios públicos una extensión de la lógica epistémica dinámica con anuncios públicos, la propuesta que se esperaría de una generalización de estos "anuncios públicos" debería

ser coherente con la semántica de los anuncios públicos en el sistema clásico PA; no obstante, como comprobará el lector cuando llegue a la parte pertinente del trabajo, este no es el caso.

A la vista de ello, surgió un segundo objetivo:

- Proponer una nueva formulación de la lógica epistémica probabilística dinámica con anuncios públicos que exstienda al sistema clásico $\mathcal{S}5$, y, en partícular, emplee la misma semántica para el tratamiento de los anuncios públicos.

Dicho esto, a continuación ofrecemos una descripción más pormenorizada del contenido de nuestro trabajo.

- En el capítulo 2, tratamos de proporcionar cierto contexto histórico y filosófico para los formalismos que desarrollamos durante el resto del trabajo, y también presentamos la versión "básica" de los mismos, junto con sus propiedades fundamentales y otros aspectos teóricos; en particular, presentamos un concepto tan fundamental como el de *bisimulación*. Como se verá de forma más concreta en el capítulo correspondiente, el lenguaje básico $\mathcal{L}_K$ (*Lógica Epistémica*) pretende proporcionar una lógica para razonar sobre situaciones en las que varios agentes con conocimiento limitado sobre distintos conjuntos de "hechos del mundo" tratan de razonar sobre estos hechos y sobre lo que los *demás* agentes razonan sobre estos hechos (y sobre lo que los demás agentes razonan que los demás agentes razonan sobre estos hechos, etcétera).

- En el capítulo 3 introducimos una nueva construcción para este lenguaje, el "operador de anuncio público", que constituye una primera incursión en aspectos dinámicos de un sistema epistémico (es decir, en situaciones que involucren *cambios en la información de la que disponen los agentes*). A lo largo del capítulo se discuten varias propiedades interesantes, y a menudo contraintuitivas, de esta nueva construcción; en particular, se observa que, "estrictamente hablando", los anuncios públicos "no son necesarios", dado que cualquier fórmula con anuncios públicos puede traducirse a una fórmula equivalente sin anuncios públicos.

- En el capítulo 4 añadimos una dimensión probabilística a nuestro lenguaje. Dicha adición permite razonar no solo en términos absolutos de *certeza* y *posibilidad* (las "categorías estándar" de la lógica epistémica), sino también en términos de *grados de certeza*; no obstante, como veremos, no elimina un tipo de razonamientos en favor de los otros, sino que permite la coexistencia de ambos. Se discuten las dificultades y los matices que hay que tener en cuenta a la hora de analizar una situación epistémica con aspectos probabilísticos; en este sentido, presentamos algunas de las propiedades deseables más típicas que suelen exigirse en estos sistemas (**CONS**, **OBJ**, **UNIF**, **SDP**, **MEAS**), así como las consecuencias teóricas de imponer dichas propiedades. También expandimos la noción clásica de bisimulación a este nuevo ámbito probabilístico.

- En el capítulo 5 realizamos nuestro aporte principal: tratamos de proporcionar un "lenguaje epistémico dinámico probabilístico" que unifique los enfoques proporcionados en los dos capítulos anteriores, junto con algunas de las propiedades teóricas básicas que satisface esta propuesta, y algunos ejemplos e ilustraciones de la potencia conceptual de la misma. En particular, el principal teorema de este

capítulo demuestra que nuestro concepto de *restricción de un modelo epistémico probabilístico* está bien definida, en el sentido de que el resultado sigue siendo un modelo epistémico probabilístico. Por otra parte, demostramos que las principales propiedades desarrolladas en el capítulo anterior (**CONS**, **OBJ**, **UNIF**, **SDP**) se conservan a través de la restricción. También presentamos brevemente la propuesta de Barteld Kooi, principal fuente de inspiración para nuestra propia propuesta.

- En el capítulo 6 introducimos un nuevo operador ("operador de conocimiento común)", que en un sentido riguroso dota de una mayor expresividad a nuestros lenguajes, y nos permite plantear algunos de los ejemplos clásicos más interesantes; tratamos también de introducir una dimensión probabilística en algunos de estos ejemplos, y acabamos el capítulo haciendo una propuesta semi-especulativa que va todavía más allá de los sistemas estudiados en este trabajo.
- En el capítulo 7 presentamos brevemente sistemas axiomáticos (ya conocidos) para cada uno de los sistemas que hemos expuesto en este trabajo, así como nuestra propuesta especulativa de un posible sistema axiomático para nuestro lenguaje epistémico dinámico probabilístico.
- Finalmente, en el capítulo 8 hacemos una síntesis de los resultados en este trabajo, y tratamos de identificar algunas de las principales líneas de investigación que plantea.

Para cada una de las partes de este trabajo, contamos con una serie de fuentes en las que nos hemos basado para obtener los principales resultados teóricos y ejemplos. Para los capítulos 1, 2 y 5 nos hemos basado principalmente en el manual *Dynamic Epistemic Logic* de van Ditmarsch y otros [1]; en el capítulo 1, sobre todo para la introducción histórica, también hemos hecho uso del manual *Modal Logic for Open Minds* de van Benthem [2], así como de algunos artículos de la Enciclopedia de Stanford. Para el capítulo 3, nuestra principal referencia ha sido el artículo *Reasoning about Knowledge and Probability* de Halpern y Fagin [3], así como en el artículo *Probabilistic Dynamic Epistemic Logic* de Kooi [4]. Las axiomáticas del capítulo 6 son una combinación de [1] y [3], y nuestra propuesta está parcialmente inspirada en la axiomática en [4]. Una excepción es el capítulo 4, donde, salvo por el artículo de Kooi [4] que nos ha servido como inspiración para desarrollar nuestro propia propuesta, los resultados teóricos son originales.

Para resultados o conceptos conocidos de la teoría de la probabilidad, hemos utilizado el manual *An Introduction to Probability and Statistical Inference* de Roussas [5], y para resultados o conceptos conocidos de la lógica proposicional hemos utilizado el manual *Mathematical Logic for Computer Science* de Ben-Ari [6].

2. Lógica Modal Multiagente, Lógica Epistémica

Este primer capítulo consistirá en una breve introducción a la lógica modal, tanto en la que podríamos considerar que es su formulación más clásica (Modelos de Kripke o de los mundos posibles) como en la familia particular de modelos en los que nos interesará centrarnos en nuestro trabajo (lógica epistémica multiagente, y principalmente $\mathcal{S}5$). Comenzaremos con una pequeña sinopsis de las motivaciones filosóficas que llevaron a desarrollar estos formalismos (siguiendo básicamente el primer capítulo de van Benthem [2] y el primer capítulo de van Ditmarsch [1], así como algunas ideas sueltas de los artículos sobre lógica intensional [7] y lógica modal [8] de la Enciclopedia Filosófica de Stanford), seguida ya de una exposición formal de los conceptos más fundamentales de los que haremos uso durante el resto del trabajo (inspirada en contenido y orden de exposición en el segundo capítulo de van Benthem [2] y en el segundo capítulo de van Ditmarsch [1]).

En nuestra exposición formal nos referiremos desde el primer momento al caso más general (a saber, modelos de Kripke multiagentes), dado que, sin añadir un enorme grado de complejidad conceptual (una interpretación natural de los "modelos clásicos" es que se tratan simplemente de "modelos multiagentes monoagentes"), generalizan y añaden una enorme riqueza al lenguaje. En aras de proporcionar al lector un contexto completo, hemos añadido cuando lo hemos considerado conveniente algunas anotaciones sobre convenciones notacionales utilizadas en los modelos "clásicos" que difieren ligeramente de la notación utilizada en este trabajo.

2.1. Breve sinopsis histórica y conceptual

¿Qué es la lógica modal? Una respuesta básica y preliminar que podría ofrecerse es que se trata simplemente de una familia de sistemas formales que expanden la lógica proposicional con dos símblos adicionales, "caja" ($\Box$) y "diamante" ($\Diamond$), junto con sus correspondientes semánticas, que procederían a explicarse a continuación en aras de comenzar a demostrar teoremas sobre estos formalismos lo antes posible. No obstante, tal respuesta no sería del todo satisfactoria, dado que nos dejaría desprovistos de una "brújula conceptual" que nos sirviese para dar algún sentido inicial a los formalismos en cuestión - incluso aunque, como ya ha ocurrido en tantas otras ocasiones en la historia de las matemáticas, estos pudiesen adquirir un sentido propio e interpretaciones nuevas más allá de los objetivos con los que inicialmente fueron planteados (que es, de hecho, lo que ha sucedido con la lógica modal, cuyas aplicaciones en la actualidad se expanden a una variedad de disciplinas que abarcan, sin ánimos de ser exhaustivos, desde la lingüística hasta la teoría de juegos). La importancia de tener una tal "brújula conceptual" reside, a pesar de todo, en que nos proporciona una perspectiva sobre el contexto en el que el formalismo fue desarrollado originalmente, y nos permite comprender la naturalidad de las diversas ramificaciones en las que fue desarrollándose a

partir de entonces.

Así pues, ¿qué es la lógica modal? O más bien, ¿para qué sirve, o qué fines tenían en mente quienes idearon originalmente sus principios? Si seguimos la exposición de van Benthem, o nos guiamos simplemente por los nombres más comunes que reciben los operadores modales, "operador necesidad" y "operador posibilidad" (además de los términos neutros "caja" y "diamante"), descubrimos rápidamente que lo que se pretende con ellos es, fundamentalmente, rescatar dos conceptos clásicos de la lógica anterior a la revolución formal que experimentó a finales del siglo XIX y principios del XX (a manos de figuras como Boole, Frege, etc.), a saber: los conceptos de "necesidad" y "contingencia". Parafraseando a un van Benthem que parafrasea a Frege, parecería que, desde que esta revolución tuvo lugar, "decir que una proposición es necesariamente cierta no es más que decir que es cierta, más una lista de datos autobiográficos sobre cómo de convencido estás de ella".

Es algo irónico que sea precisamente en una observación un tanto sarcástica como la anterior donde puede encontrarse la clave para redescubrir y recuperar el valor de estos conceptos clásicos. En efecto, si las aseveraciones sobre la "necesidad" o la "posibilidad" de una proposición no se interpretan como algo sustancial a lo que dicha proposición *dice sobre el mundo*, sino que más bien a lo que dicha proposición *revela sobre el carácter limitado del acceso que el sujeto articulador de la proposición tiene al mundo*, dichas aseveraciones adquieren un sentido totalmente novedoso. De manera resumida, pues, la observación clave del enfoque modal es que algo puede ser considerado "necesario" o "contingente" solo desde un determinado punto de vista o acceso al mundo; pero esto, lejos de restar valor al formalismo, lo hace más expresivo: a la capacidad de formalizar proposiciones sobre el mundo desde un punto de vista "objetivo" se añade ahora la de formalizarlas desde puntos de vista "limitados" o, más generalmente, dotados de un carácter "intensional"[1] que en muchos contextos ni siquiera debe ser concebido como limitante; en la medida en que dichos puntos de vista pueden considerarse ellos mismos como entidades en el mundo susceptibles a ser estudiadas, la posibilidad de razonar sobre ellos expande enormemente nuestro horizonte teórico.

De modo que, para poder hablar en términos más concretos de lo que significa que algo sea "necesario" o "contingente" para un agente, en un contexto, o desde un punto de vista determinado, hay que ponerse de acuerdo de antemano en la interpretación concreta que se le dará a los "operadores intensionales" $\Box$ y $\Diamond$. Llegados a este punto, se nos pone por delante una cantidad enorme, potencialmente ilimitada, de posibles interpretaciones de los mismos. Una exploración pormenorizada de las principales queda totalmente al margen de nuestros objetivos en este trabajo, pero, como simple anotación y para ilustrar la enorme versatilidad de los formalismos que se van a explorar, algunas de las interpretaciones de la fórmula $\Box A$, donde A sería una proposición arbitraria, son: "A es (éticamente) obligatorio" *(lógica deóntica)*, "A es deseado (por un determinado agente)" *(lógica volitiva)*, "siempre será el caso que A" *(lógica temporal en modalidad*

[1] Si el lector está interesado en una discusión profunda de carácter filosófico sobre lo que queremos decir con el término "intensional", le recomendamos la lectura del artículo correspondiente [8] en la Enciclopedia Filosófica de Stanford. Dicho sea de paso, este artículo expande enormemente sobre todo lo discutido en esta introducción, que no deja de ser breve en relación a la vastedad de todo lo que podría aspirar a cubrir, de modo que la recomendamos en general a cualquier lector cuya curiosidad se haya visto suscitada.

futura), "siempre ha sido el caso que A" *(lógica temporal en modalidad pasada)*.

En nuestro trabajo no nos interesa en particular ninguno de los enfoques anteriores: más bien, nos centraremos en el enfoque *epistémico* (del griego *episteme*, "conocimiento"), y quizá nos detengamos tangencialmente en el *doxástico* (del griego *doxa*, "opinión" o "creencia"); el lector podrá intuir ya más o menos lo que implica cada uno de estos enfoques. Si es que se incluye algún aspecto que pudiese interpretarse como "temporal" en nuestros formalismos lógicos, será en la componente *dinámica*, que tiene que ver más con las "transformaciones discretas de estados de conocimiento" que con un concepto más abstracto y "por sí mismo" de temporalidad; finalmente, también incluiremos elementos *probabilísticos* que van algo más allá del enfoque modal básico, y requieren por lo tanto de construcciones adicionales para su estudio.

Todos estos aspectos irán siendo propiamente introducidos a lo largo de los correspondientes capítulos del trabajo, empezando por el aspecto epistémico, que introduciremos a continuación; no obstante, antes de hacer esto, me gustaría ilustrar de manera concreta al menos en una ocasión el concepto de "necesidad" más general, con un ejemplo que, como el lector verá, enlaza de maravilla con toda la conversación filosófica sobre el conocimiento, la creencia, y la naturaleza limitada de nuestro acceso al y nuestra conceptualización del mundo.

2.1.1. Ilustración preliminar: El Lucero del Alba

El ejemplo a continuación se ha extraído del artículo de Stanford de Lógica Intensional [8].

Coloquialmente, el planeta Venus es conocido como "El Lucero del Alba", dado que a menudo puede observarse su brillo intenso en el cielo del amanecer. Similarmente, también se le llama a veces "El Lucero del Atardecer", aunque esto es mucho más común en la esfera angloparlante ("The Evening Star"). Si bien a día de hoy sabemos que las etiquetas "Venus", "El Lucero del Alba" y "El Lucero del Atardecer" se refieren en los tres casos al mismo objeto astronómico (a saber, el planeta Venus), en principio, para una persona que ignore el contexto que hemos expuesto aquí, esta equivalencia no es algo obvio. De esta forma, mientras que puede entenderse que las frases "El lucero del alba es el lucero del Atardecer" y "El lucero del alba es el planeta Venus" pueden aportar información nueva en determinadas situaciones, una frase como "El planeta Venus es el planeta Venus" nos choca como algo cómicamente obvio. ¿Cómo es esto posible, si, en los tres casos, estamos reafirmando la equivalencia de tres objetos que efectivamente son "el mismo"?

Como ya hemos observado más arriba, el estudio de este tipo de cuestiones no se limita al siglo pasado, sino que, de hecho, una parte importante de los esfuerzos en el desarrollo inicial de la lógica modal "moderna" buscan precisamente redescubrir unas categorías de modalidad que se pueden hallar ya en la filosofía, teología y escolástica medieval [9], y cuyos orígenes pueden trazarse al menos hasta Aristóteles[2]. Acercándonos algo más a nuestros tiempos, pueden también encontrarse ciertos para-

[2] ¿A quién le sorprende?

lelismos con el trabajo de Kant[3], y más concretamente con la distinción entre juicios "a priori analíticos" y "a priori sintéticos" que este plantea en su Crítica de la Razón Pura [10], donde reflexiona sobre en qué medida y por qué mecanismos afirmaciones como $5 + 7 = 12$ (sección tercera, capítulo 1: *axiomas de la intuición*) nos "aportan", "producen" o "descubren" un conocimiento nuevo, que no teníamos antes. No obstante, lo que sí nos proporciona la lógica modal, y más concretamente la semántica de Kripke "de los mundos posibles", es un marco formal en el que razonar sobre este tipo de cuestiones con el rigor propio de una forma de proceder matemática.

En particular, el ejemplo anterior plantea una situación en la que agentes hipotéticos pueden tener un *conocimiento* limitado del mundo, que después puede ser *actualizado* a través de afirmaciones que se les *anuncian* de alguna forma ("¿Recuerdas ese lucero del alba y ese lucero del atardecer de los que me has hablado alguna vez? ¡Pues resulta que son la misma cosa!"). Este es el tipo de situaciones, *grosso modo*, que se prestan naturalmente a ser estudiadas en Lógica Epistémica Dinámica.

2.1.2. Lógica Epistémica

Dejemos por ahora de lado la parte "dinámica", y centrémonos en tratar de caracterizar antes de nada qué es lo que *se supone* que estudia la lógica epistémica. Por la palabra *epistémica*, puede deducirse que se trata de algo que tiene que ver con "el conocimiento", pero, ¿qué es "el conocimiento"? Desde luego, hay muchos enfoques posibles a la hora de estudiar una cuestión tan amplia como esta (y muchos dirían que es inabarcable). La epistemología, por ejemplo, se hace preguntas sobre qué es lo que constituye un "conocimiento (científico) adecuado" en relación con los mecanismos que se emplean para su obtención / fabricación y los sesgos implícitos en cada uno de ellos, o incluso con el contexto social / institucional en el que debe ser producido. La gnoseología, por otra parte, se preocupa más bien por cuestiones fenomenológicas relativas a nuestro acceso a las experiencias del mundo y sobre cómo estas nos producen "certezas". También pueden considerarse enfoques más técnicos, como los desarrollados por Shannon o Kolmogorov en el contexto de una "teoría de la información", que tienen que ver con los aspectos cuantificables de la información (p.e., ¿Cuánta información contiene un mensaje? ¿Qué recursos computacionales son necesarios para describir algo?). Pueden considerarse muchas otras posibilidades: desde los diversos enfoques psicológicos hasta las interpretaciones de la probabilidad como "grado de certeza" o "certidumbre" sobre determinados hechos, pasando por los esfuerzos de diversos paradigmas de la informática para desarrollar inteligencias artificiales que, de alguna manera, tendrían que estar dotadas de una "capacidad para conocer" como requisito para poder calificarse realmente como inteligentes. Por supuesto, en ningún momento debe entenderse que estos enfoques están de alguna manera mutuamente aislados; y, en particular, aunque la lógica epistémica tiene su propio enfoque característico a la

[3] Aunque el trabajo de Kant no suele citarse como una de las principales influencias en el desarrollo de la lógica modal contemporánea, consideramos que los paralelismos que pueden trazarse con otros trabajos son demasiado claros, que su influencia en el pensamiento moderno en general es demasiado amplia, y, sobre todo, que la posición de su trabajo en la cronología del pensamiento moderno, así como la naturaleza de su contenido, se ajustan demasiado bien al postulado de su influencia sobre el desarrollo de los nuevos enfoques que caracterizan a la lógica modal moderna como para ignorarlo.

hora de aproximarse a esta cuestión, sería ingenuo decir que, en última instancia, está separada por completo del resto de la ciencia en su conjunto, como si de un diminuto archipiélago de formalismo abstracto se tratase.

¿De qué manera podemos sintetizar, entonces, en qué consiste el enfoque característico de la lógica epistémica? Trataremos de ofrecer una caracterización más o menos original. Lo esencial sería lo siguiente: la lógica epistémica no se pregunta en ningún momento cuestiones fundamentales como *qué es el conocimiento* o *cómo tenemos acceso a nuestras intuiciones más inmediatas*, al menos no como base de su metodología; por el contrario, la suposición básica de este enfoque es que se partirá en todos los casos de un concepto preestablecido de lo que constituye un "hecho sobre el mundo", o se operará siempre dentro de contextos donde la validez de la aplicación de este tipo de consideraciones será, al menos en un sentido funcional o pragmático, "evidente". Partiendo de esta base, se definirán formalismos que en cierto modo deberán capturar las relaciones que de manera "natural" debería satisfacer cualquier construcción de conocimiento a partir de estos hechos, teniendo también en cuenta las diversas estructuras de accesibilidad que los diversos agentes "susceptibles de conocer" tienen respecto a cada uno de estos hechos. En particular, estos formalismos hacen naturalmente atractiva la posibilidad de razonar sobre formulaciones epistémicas de orden superior, a saber, formulaciones sobre el conocimiento que un agente posee sobre el conocimiento que posee otro agente (con potencialmente infinitos grados de recursividad). La naturalidad con la que este tipo de razonamientos pueden aplicarse a áreas como la teoría de juegos es evidente, y se nos viene inmediatamente a la cabeza en el tipo de intrigas que caracterizan una parte significativa de los *thrillers* en la ficción contemporánea.

Un área que ha contribuido enormemente al desarrollo de la lógica epistémica es el análisis de sistemas (informáticos) distribuidos – como manifiestan, sin ir más lejos, Ronald Fagin y Joseph Y. Halpern en su artículo *Reasoning about Knowledge and Probability* [3], que citaremos profusamente en secciones posteriores de este trabajo. En principio, podría parecer algo extraño, dado que los componentes de estos sistemas no son "conscientes" en ningún sentido evidente de la palabra (o al menos no lo han sido hasta el momento), y por lo tanto podría ser extraño imaginar que pueden "conocer" algo. No obstante, en muchos aspectos funcionales, los formalismos de la lógica epistémica también resultan muy adecuados para describir los diversos estados en los que dichos sistemas distribuidos pueden encontrarse en distintos puntos del tiempo. Esto tiene sentido si consideramos la naturaleza asimétrica del acceso y el intercambio de la información para cada componente en un sistema de este tipo. Un ejemplo clásico que veremos más adelante (en el capítulo 6) es el problema de los generales bizantinos, entre otros.

Por último, mencionamos brevemente una familia muy cercana de modelos lógicos, que constituyen la *lógica doxástica* (de la creencia) dada la estrecha relación filosófica que existe entre "conocer algo" y "creer en algo". Efectivamente, uno de los debates centrales en la epistemología desde, posiblemente, hace siglos gira en torno a la siguiente reflexión: es bastante claro que entre "saber algo" y "creer en algo que es verdad" hay una diferencia importante, dado que, *grosso modo*, puede que "creas en algo correcto por las razones equivocadas" - por ejemplo, el jugador que tras desperdiciar una fortuna en la máquina tragaperras, razona que como ya ha invertido tanto

dinero en el juego, le queda muy poco para ganar el gran premio, y acaba ganándolo por casualidad a pesar de que su razonamiento es completamente falaz. En su célebre artículo de 1963 [11], Gettier plantea ejemplos todavía más interesantes de situaciones en las que, incluso cuando una creencia verdadera parece estar bien justificada, sigue sin ser del todo claro que debiera considerarse conocimiento. La cuestión es, entonces, tratar de caracterizar en qué consiste exactamente esa "brecha". El estudio comparado de las lógicas epistémica y doxástica proporciona un posible enfoque en este respecto.

2.2. Sintaxis y semántica de los modelos de Kripke multiagentes

La definición del lenguaje modal básico se construye sobre un conjunto finito de agentes A, que normalmente denotaremos con las primeras letras del alfabeto a, b, c ..., y en casos más generales a_1, a_2, ... , a_k; y un conjunto numerable de átomos At, que denotaremos con las letras p, q, r ..., y en casos más generales p_1, p_2, ... , p_k, e incluirán también a los símbolos $\top$ (tautología / "siempre verdadero") y $\bot$ (contradicción / "siempre falso").

Definición 2.1. ***Lenguaje modal multiagente básico.*** *Sea At un conjunto numerable de átomos y A un conjunto finito de agentes. Definimos el lenguaje modal (multiagente) básico $\mathcal{L}_K$ inductivamente con la notación de Backus-Naur (BNF):*

$$\varphi ::= p \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid K_a\varphi \tag{2.1}$$

donde $p \in At$ y $a \in A$. □

Nota 2.1. Dado que esta última notación no es del todo común, una breve aclaración sobre su significado: "todos los átomos son fórmulas"; "si φ es una fórmula, entonces $\neg\varphi$ también lo es"; "si φ_1 y φ_2 son fórmulas, entonces $\varphi_1 \wedge \varphi_2$ también lo es"; "si φ es una fórmula, entonces $K_a\varphi$ también lo es". □

Nota 2.2. Además de las conectivas ($\neg$, $\wedge$; "negación lógica" y "conjunción lógica", respectivamente) utilizadas en la definición anterior, utilizaremos varias notaciones convencionales típicas, cuyas definiciones, aunque probablemente sean conocidas más que de sobra por el lector, escribiremos a continuación en aras de la exhaustividad:

- $\varphi \vee \psi$ significa $\neg(\neg\varphi \wedge \neg\psi)$; el símbolo $\vee$ se llamará "disyunción lógica".
- $\varphi \to \psi$ significa $\neg\varphi \vee \psi$; el símbolo $\to$ se llamará "implicación lógica".
- $\varphi \leftrightarrow \psi$ significa $(\varphi \to \psi) \wedge (\psi \to \varphi)$; el símbolo $\leftrightarrow$ se llamará "equivalencia lógica".

Además, el lector habrá percibido que, para evitar un uso tedioso y excesivo de los paréntesis, estamos asumiendo implícitamente ciertas "reglas de prioridad" de unas conectivas sobre otras. La jerarquía estándar de prioridades será, de mayor a

menor prioridad: $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$. En cuanto a los operadores modales (como K_a o sus derivados), la prioridad de estos será mayor que la de cualquier conectiva proposicional binaria.

De la misma forma, utilizaremos también notaciones especiales para "abreviar" fórmulas en las que aparezca el operador modal, K_a. De todas estas, quizá la más importante sea $\hat{K}_a\varphi$, que significa $\neg K_a \neg \varphi$, y nos referiremos a este nuevo operador como el "dual" de K_a.

La definición del lenguaje modal básico podría haberse hecho perfectamente incluyendo todos estos símbolos (conectivas adicionales y operador modal dual) como parte de la misma. Si no se ha hecho esto, es para evitar tener que incluir sus respectivas semánticas en la definición de verdad que daremos más adelante. □

K_a y $\hat{K}_a$ son los llamados operadores epistémicos, o, circunscribiéndonos a un contexto más general, operadores modales. En este trabajo los notamos con la letra "K" en referencia a la palabra inglesa "knows" ("conoce"), pero en otros contextos se utiliza una gran diversidad de notaciones alternativas. En el contexto de la lógica epistémica, la fórmula $K_a\varphi$ suele leerse en lenguaje natural como "el agente a conoce φ / sabe que φ", o simplemente "a conoce φ / a sabe que φ"; por otra parte, la fórmula $\hat{K}_a\varphi$ suele leerse como "a considera que φ es posible". De ahí que nos referiremos a ellos, de aquí en adelante, como "operador conocimiento" y "operador posibilidad", respectivamente.

Para que el lector pueda hacerse una idea preliminar de la interpretación de estos operadores epistémicos, la ilustraremos con un ejemplo de cómo se leerían algunas fórmulas en lenguaje natural. Por ejemplo: digamos que a representa al agente "José Arcadio", b representa al agente "Mr. Brown", y "p" es la afirmación "José Arcadio sigue vivo". Entonces las fórmulas siguientes se podrían leer como:

- $K_a p$**:** José Arcadio sabe que (él mismo) está vivo.
- $\neg K_b p$**:** Mr. Brown no sabe que José Arcadio está vivo.
- $\hat{K}_a K_b p$**:** José Arcadio considera posible que Mr. Brown sepa que está vivo.
- $\neg K_b \hat{K}_a K_b p$: Mr. Brown no sabe que José Arcadio considera posible que Mr. Brown sepa que está vivo.

En particular, las últimas dos fórmulas son ejemplos de proposiciones epistémicas de orden superior.

Nota 2.3. Una notación más "clásica" y, en cierto modo, neutral, es la que utiliza los símbolos "caja" ($\Box_a$) y "diamante" ($\Diamond_a$), respectivamente. Como ya adelantamos en la introducción, es común referirse al primer símbolo como un "operador modal de necesidad", y al segundo como un "operador modal de posibilidad". Más adelante en nuestro trabajo, veremos que "K_a" y "$\hat{K}_a$" no son los únicos operadores modales que se ajustan a esta dicotomía de necesidad y posibilidad (como adelanto, algunos otros "operadores modales de tipo necesidad" serán C_a *(conocimiento común)* y $[\varphi]$ *(anuncio público)*, y algunos otros "de tipo posibilidad" serán $\hat{C}_a$ y $\langle\varphi\rangle$). Todos estos operadores irán por parejas, y se dirá que son "duales" dado que satisfarán las relaciones mutuas

$\Diamond\varphi \equiv \neg\Box\neg\varphi$ y $\Box\varphi \equiv \neg\Diamond\neg\varphi$ (la primera es por definición, la segunda se deducirá de la semántica). □

Ya tenemos el esquema básico para construir cualquier fórmula modal. Ahora tenemos que dotarlas de una semántica para poder razonar sobre su valor de verdad. Para ello tenemos que definir primero lo que es un modelo de Kripke.

Definición 2.2. ***Modelo de Kripke multiagente.*** *Dado un conjunto numerable de átomos At y un conjunto finito de agentes A, un modelo de Kripke (o de los mundos posibles) multiagente (o polimodal) es una estructura $M = \langle S, R^A, V^{At}\rangle$ donde:*

- *S es un conjunto de estados o "mundos". También nos referiremos a S como el dominio de M, $\mathcal{D}(M)$.*
- *R^A es una aplicación que, para cada $a \in A$, proporciona una "relación de accesibilidad" $R^A_a \subseteq S \times S$ (es decir, un subconjunto arbitrario de tuplas del producto cartesiano de S por sí mismo).*
- *$V^{At} : At \to \mathcal{P}(S)$ es una "aplicación evaluadora" que, para cada $p \in At$, proporciona un subconjunto $V^{At}(p) \subseteq S$ (como se verá a continuación cuando definamos la semántica, dicho subconjunto se interpretará como "el conjunto de los estados donde se satisface p").*

□

Nota 2.4. A menudo prescindiremos de referencias explícitas a los conjuntos At y A, y notaremos simplemente $M = \langle S, R, V\rangle$ (a veces utilizaremos paréntesis, (S, R, V), en lugar de los "corchetes diamante", simplemente porque nos resulta difícil recordar ser consistentes en un aspecto tan trivial). A veces también notaremos $\sim$ en lugar de R, y $\sim_a$ en lugar de R_a (principalmente utilizaremos esta notación para expresar que se trata de una *relación de equivalencia* 2.6). Otra notación que utilizaremos típicamente es R_ast en lugar de $s, t \in R_a$; la notación equivalente para $\sim_a$ será $s \sim_a t$. En lenguaje natural, acostumbraremos a leer R_ast como "t es accesible desde s (para el agente a)". □

Definición 2.3. ***Definición de verdad en un modelo de Kripke multiagente.*** *El valor de verdad de las fórmulas modales multiagentes se evalúa en tuplas (M, s) (modelo, estado), que llamaremos "modelos puntuados" o, abusando de la terminología, simplemente "estados", de la siguiente manera:*

$$\begin{array}{lll} M, s \vDash p & sii & s \in V(p) \\ M, s \vDash \neg\varphi & sii & no\ se\ tiene\ M, s \vDash \varphi \\ M, s \vDash \varphi \wedge \psi & sii & M, s \vDash \varphi\ y\ M, s \vDash \psi \\ M, s \vDash K_a\varphi & sii & Para\ todo\ t \in S:\ R_ast \Longrightarrow M, t \vDash \varphi \end{array} \tag{2.2}$$

Donde $a \in A$ y $p \in At$. □

Nota 2.5. A partir de la semántica anterior, el lector puede comprobar que la "cláusula semántica" para decidir si una proposición de la forma $\hat{K}_a\varphi$ es verdadera en un modelo puntuado puede escribirse de la siguiente manera:

$$M, s \vDash \hat{K}_a\varphi \qquad sii \qquad Existe\ t \in S:\ R_a st\ y\ M, s \vDash \varphi \tag{2.3}$$

□

Nota 2.6. A partir de ahora, notaremos "No se tiene $M, s \vDash \varphi$" como $M, s \nvDash \varphi$. □

En cualquier construcción teórica siempre es importante tratar de dar un "sentido informal" a los formalismos que vamos construyendo, y sobre todo cuando utilizan terminología tan sugerente como "mundos posibles" y "relaciones de accesibilidad". De modo que, ¿qué es todo esto de los mundos posibles y las relaciones de accesibilidad? *¿Universos paralelos? ¿Portales ultradimensionales?* Nada tan emocionante (aunque, por supuesto, los lectores más visionarios siempre son libres de dar cualquier interpretación alternativa que les parezca adecuada e interesante).

A efectos de nuestro planteamiento conceptual, un modelo de los mundos posibles es una forma de modelar la incertidumbre de los agentes en una situación epistémica, y la naturaleza limitada de su certeza sobre los hechos en el mundo. Más concretamente, cada estado o "mundo posible" representa un conjunto de hechos que se podrían dar de manera simultánea, y las relaciones de accesibilidad de cada agente representan los conjuntos de situaciones que resultan indistinguibles para los agentes en cuestión desde sus respectivos puntos de vista. En este sentido, que un agente conozca un hecho significa que *en todos los estados* que le resultan indistinguibles desde el "real" se satisface este hecho; por otra parte, que un agente considere posible un hecho significa que *en al menos uno de los estados* que le resultan indistinguibles desde el "real" se satisface este hecho.

Un lector sagaz podrá percibir que, en general, los operadores de tipo necesidad se comportan de manera análoga a un cuantificador universal $\forall$, y los operadores de tipo posibilidad se comportan de manera análoga a un cuantificador existencial $\exists$. En realidad, puede considerarse que la lógica modal constituye un fragmento de la lógica de primer orden; para más detalles, consultar Van Benthem [2] (capítulo 7).

Una forma cómoda de representar modelos de Kripke que no sean demasiado complejos es a través de esquemas gráficos. Ilustramos esto con un ejemplo clásico que aparece en Van Ditmarsch [1] (páginas 16-22).

Ejemplo 2.1. **Situación Groningen-Liverpool-Otago (GLO).** Consideremos la siguiente situación: Anna, Bradley y Charles (a, b y c) son tres amigos, residentes en Groningen, Liverpool y Otago, respectivamente. Supongamos que por alguna razón Anna decide desarrollar una teoría sobre el tiempo que hace en Groningen y en Liverpool: en Groningen, o bien hace sol (g) o bien está nublado ($\neg g$). Análogamente en Liverpool, l o $\neg l$. Si consideramos todos los estados contemplados en esta teoría, obtenemos cuatro posibilidades, que podemos representar como $\langle g, l\rangle$, $\langle g, \neg l\rangle$, $\langle \neg g, l\rangle$ y $\langle \neg g, \neg l\rangle$. Dado que a se encuentra en Groningen, sabe el tiempo que hace allí, pero desconoce si en Liverpool hace o no hace sol; por lo tanto, los estados de la forma $\langle x, l\rangle$ y $\langle x, \neg l\rangle$,

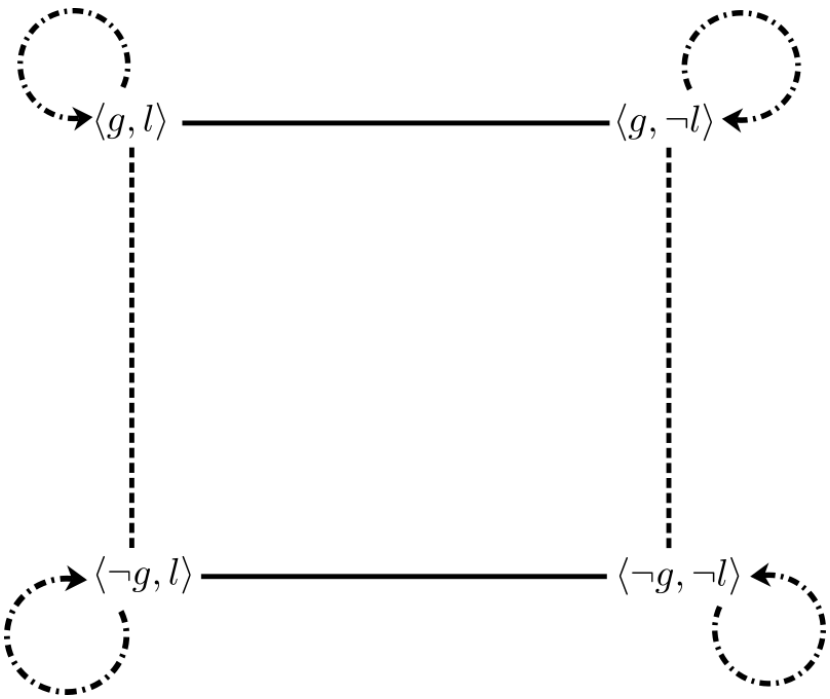

Figura 2.1: Ilustración del planteamiento inicial de la situación GLO.

con x fijo, son indistinguibles para a (esto lo escribiríamos como $R_a\langle x, l\rangle\langle x, \neg l\rangle$, o quizá como $\langle x, l\rangle \sim_a \langle x, \neg l\rangle$). Obviamente, b se encuentra en una situación análoga. Supongamos además, por último, que a y b son conscientes de sus respectivas circunstancias[4]. Esquemáticamente, esta situación podría representarse como en la figura 2.1. En esta figura, las líneas continuas representan las relaciones de accesibilidad del agente a, las líneas discontinuas representan las relaciones de accesibilidad del agente b, y las líneas mixtas representan que existen relaciones de accesibilidad de los dos agentes; en este caso, las líneas mixtas corresponden a las relaciones reflexivas, es decir, las que hay desde cada estado a sí mismo. Dado que, como veremos más adelante 2.6, estas relaciones de accesibilidad siempre se tienen en los modelos epistémicos, a menudo prescindiremos de ellas en las representaciones.

Concluyamos este ejemplo con un precursor del tipo de situaciones que se considerarán a partir de la parte *dinámica* de este trabajo. Supongamos que es de noche en Otago (por lo tanto es de día en Groningen y Liverpool), y c, que por alguna razón está despierto, decide llamar a sus amigos a y b por teléfono para hablar sobre sus negocios conjuntos. Empieza llamando a a, que se encuentra en Groningen, y a le comenta casualmente que en Groningen hace un día soleado. Antes de finalizar la conversación, c le dice (verazmente[5]) a a que a continuación va a informar a b, con quien también tiene negocios, sobre todo lo que ha hablado con a.

[4] Aunque ahora mismo estamos utilizando esta noción de "ser conscientes de sus respectivas circunstancias" de manera algo informal, en el capítulo 6 veremos que se puede formalizar rigurosamente a través del operador de conocimiento común C_B; por ahora, el lector puede simplemente considerar que esta noción simplemente hace referencia a que los agentes involucrados en el ejemplo tienen un mapa mental completo de la situación que hemos descrito.

[5] A menos que se indique explícitamente lo contrario, en todos los ejemplos de este libro donde algún agente "diga", "comente", o de cualquier otra forma comparta con todos o con algunos de los agentes una afirmación, se supondrá que la afirmación es veraz. Esta es otra de las limitaciones que

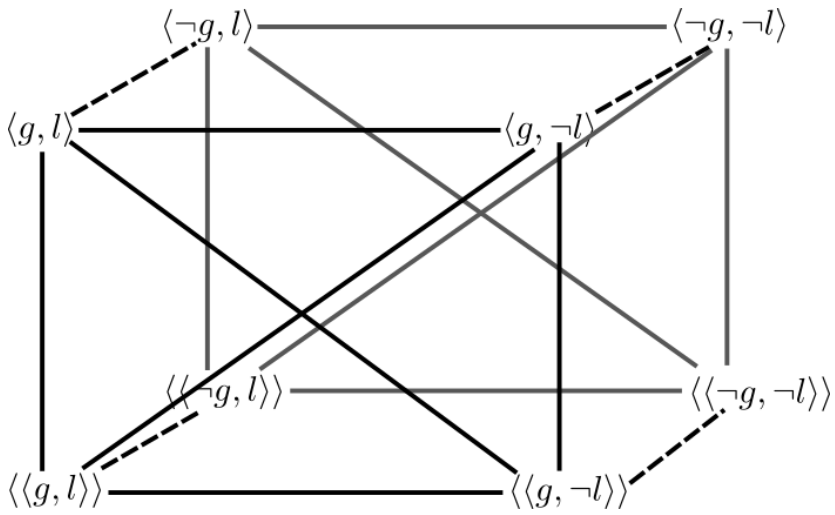

Figura 2.2: Ilustración de una situación algo más compleja.

Al finalizar la llamada, a se pregunta si c le hablará también a b sobre su comentario casual acerca del tiempo que hace en Groningen. De este modo, a se plantea dos (conjuntos de) estados posibles: el primero, que denotaremos como antes $\langle g, y\rangle$, donde y puede referirse tanto a l como a $\neg l$, representará la situación en la que se tenga $g \wedge y \wedge \neg K_b g$, es decir, el estado en el que c no le ha transmitido a b la información de a sobre el tiempo que hace en Groningen; el segundo, que denotaremos $\langle\langle g, y\rangle\rangle$, representará el estado en el que se tenga $g \wedge y \wedge K_b g$ (la otra posibilidad). Esta situación es bastante más compleja que la original, pero un esquema como el de 2.2, donde las líneas continuas representan las relaciones de accesibilidad del agente a y las discontinuas las del agente b (las reflexivas se han obviado), nos facilita enormemente su interpretación.

Alguien podría observar que, en realidad, este modelo no se ajusta a lo que nosotros hemos definido propiamente como *Modelo de Kripke*, dado que, además de los conjuntos $V(p)$, parecería que tenemos que considerar también un conjunto adicional $V(K_b g)$, que no se corresponde con el valor de verdad de ninguna fórmula atómica, sino de una fórmula más compleja. Esto es cierto, pero, por ahora, lo dejaremos de lado para regresar sobre ello cuando introduzcamos los aspectos dinámicos de nuestra lógica. Por el momento, quedémonos con el siguiente mensaje: que este tipo de diagramas puede ser realmente útil para resumir visualmente situaciones bastante complejas.

□

Nota 2.7. Además de las notaciones que ya hemos mencionado, otras abreviaturas importantes que utilizaremos en ocasiones serán $E_B\varphi$ y $\hat{E}_B\varphi$, donde $B \subseteq A$ es un subconjunto de los agentes. Dichas abreviaturas se definen, respectivamente, como:

- $E_B\varphi := \bigwedge_{a\in B} K_a\varphi$
- $\hat{E}_B\varphi := \neg E_B\neg\varphi$

algunos podrían percibir de la lógica epistémica: en principio, no plantea ninguna manera obvia de modelar situaciones, del todo comunes, en las que a alguno de los agentes le podría interesar mentir. En la conclusión de nuestro trabajo (8) trataremos de incluir algunas propuestas en este respecto.

Recordemos que el conjunto de agentes A siempre será finito en nuestras construcciones, lo que garantiza la buena definición de su semántica a partir de la semántica ya proporcionada.

La elección de la letra "E" hace referencia a la palabra "everybody" (todo el mundo) en inglés, dado que una forma natural de leer estas abreviaturas sería "Todos los agentes del conjunto B saben que φ" o "Alguno de los agentes del conjunto B consideran posible que φ". Esta abreviatura se tratará a efectos prácticos como su propio operador. Observemos el aspecto "asimétrico" de estos operadores: para que se tenga $E_B\varphi$, tiene que darse que, para todos los agentes $b \in B$, se tenga $K_b\varphi$; en cambio, para que se tenga $\hat{E}_B\varphi$, basta con que haya un solo agente $b \in B$ para el que $K_b\varphi$.

Por último, otra abreviatura de carácter más técnico es la correspondiente a la aplicación iterada de operadores modales, a saber: $K_a^k := K_a K_a^{k-1}$, donde k es un número natural mayor que 0 (y las respectivas para $\hat{K}$, E y $\hat{E}$), y $K_a^1 := K_a$. Retomando nuestro ejemplo de José Arcadio y Mr. Brown, la fórmula $\hat{E}_A^3\hat{K}_b p$, con $A = \{a, b\}$ podría leerse como "Alguno entre José Arcadio y Mr. Brown considera posible que alguno entre José Arcadio y Mr. Brown considere posible que alguno entre José Arcadio y Mr. Brown considere posible que Mr. Brown considere posible que José Arcadio esté vivo". Naturalmente, vemos por qué el lenguaje natural no es el más adecuado para trabajar con este tipo de razonamientos - aunque, sin lugar a dudas, podemos estar seguros de que cualquiera que haya trabajado durante un tiempo en lógica epistémica y haya obtenido algún resultado inverosímil en algún momento habrá tratado de convencerse o disuadirse a sí mismo de su corrección pronunciando para sí mismo su "significado" en voz alta, con las esperanzas de que esto le proporcione alguna claridad adicional. □

2.3. Resultados básicos y conceptos importantes

En esta sección, como su nombre indica, presentaremos algunos resultados básicos y conceptos importantes que, con sus respectivas generalizaciones, irán apareciendo de manera recurrente a lo largo del resto de nuestro trabajo.

A continuación plantearemos un primer resultado que nos resultará de gran utilidad a la hora de razonar sobre la validez de fórmulas en la medida en que dichos razonamientos no involucren la parte propiamente modal del lenguaje, y nos permitirá, en resumen, hacer el mismo tipo de sustituciones, equivalencias y deducciones que si estuviésemos operando simplemente con fórmulas de la lógica proposicional. Esto es sumamente conveniente, dado que el énfasis de nuestro trabajo estará casi exclusivamente en la parte modal, y un resultado así nos ahorra muchísimo tiempo que de otra forma tendríamos que invertir en cuestiones técnicas de interés escaso. La prueba de este resultado no se proporciona dado que puede considerarse "estándar" en el campo, y, en todo caso, se trata de un resultado sumamente intuitivo.

Definición 2.4. ***Átomo modal.*** *Sea φ una fórmula en $\mathcal{L}_K$. Definimos el conjunto de sus átomos modales como:*

$$At_{Mod}(\varphi) = \begin{cases} \varphi & si\ \varphi = K_a\psi\ o\ \varphi = p \in At \\ At_{Mod}(\psi) & si\ \varphi = \neg\psi \\ At_{Mod}(\varphi_1) \cup At_{Mod}(\varphi_2) & si\ \varphi = \varphi_1 \wedge \varphi_2 \end{cases} \tag{2.4}$$

En otras palabras, los "átomos modales" de una fórmula son sus subfórmulas "más simples hasta el nivel del operador modal más externo". □

Proposición 2.1. Sea M, s un modelo de Kripke puntuado, y sean $\varphi_1, ..., \varphi_k, \varphi \in \mathcal{L}_K$. Consideremos el lenguaje $P := P(\varphi_1, ..., \varphi_k, \varphi)$ cuyos átomos son $\bigcup_{i=1}^{k} At_{Mod}(\varphi_i)$, y cuyas conectivas lógicas son $\{\wedge, \neg\}$. Entonces, si $S := \{\varphi_1, ..., \varphi_k\}$,

$$S \vDash_P \varphi \Longrightarrow (M, s \vDash \varphi_1, ..., M, s \vDash \varphi_k \Longrightarrow M, s \vDash \varphi)$$

Donde la consecuencia lógica en P, representada como $\vDash_P$, tiene la semántica típica de la lógica proposicional [6] (definición 2.48).

□

Corolario 2.1. Las reglas de cálculo lógico para la lógica proposicional también son válidas para deducir el valor de verdad de fórmulas $\varphi \in \mathcal{L}_K$ en cualquier modelo puntuado M, s.

Demostración. Obvio teniendo en cuenta que

$$S \vdash_P \varphi \Longleftrightarrow S \vDash_P \varphi$$

Donde $\vdash_P$ representa la deducibilidad lógica a partir de reglas de cálculo lógico para la lógica proposicional en P (una axiomática estándar puede consultarse en [6]).

Q.E.D.

□

Observación 2.1. Obviamente, dado que la lógica modal es más expresiva que la lógica proposicional, las reglas del cálculo lógico de la lógica proposicional no proporcionan una axiomática *completa* de la misma. □

Nota 2.8. A partir de ahora, supondremos que el lector tiene cierta destreza en el manejo de las equivalencias, sustituciones y deducciones básicas en la lógica proposicional. □

Antes de proseguir con otras cuestiones, queremos señalar que los resultados que acabamos de presentar pueden extrapolarse sin dificultad alguna a cualquier otro operador que se plantee en este trabajo. En este sentido, evitaremos volver a presentar "el mismo resultado" en cada capítulo, pero el lector deberá tener en cuenta que podemos utilizar sin problemas las leyes del cálculo lógico de la lógica proposicional siempre y cuando sea "fuera" de los operadores no-proposicionales.

2.3.1. Familias de modelos interesantes

A continuación definiremos varias clases de modelos que son especialmente interesantes en el contexto de nuestro trabajo. Cada una de estas clases puede considerarse la "más natural" en diversos contextos; en particular, la clase de los modelos que llamaremos propiamente "epistémicos" (es decir, modelos que se ajustan a situaciones en las que diversos agentes pueden razonar sobre hechos que conocen con certeza, y no creencias o hechos con otros tipos de fenomenología) suele denotarse por $\mathcal{S}5$. La clase que más abajo denotaremos como $\mathcal{KD}45$, por ejemplo, correspondería a una familia menos restrictiva de modelos que a menudo se interpretan como "modelos para la lógica de las creencias" ("logic for belief"); filosóficamente, es natural y consistente que la clase $\mathcal{S}5$ esté incluida en $\mathcal{KD}45$, dado que, como comentamos en nuestra breve sinopsis, "conocer algo" implica "creer que ese algo es verdad", pero no necesariamente se tiene el recíproco.

Todas estas consideraciones de carácter más bien filosófico habrán dado lugar, sin lugar a dudas, a un gran número de conversaciones de enorme interés a lo largo de la historia del pensamiento. Desde un punto de vista más técnico, que es el que, no obstante, nos interesa en este trabajo, el estudio particularizado de los diferentes modelos también tienen un interés especial: cada uno de ellos satisface una serie de propiedades que se traducen en diversos conjuntos de axiomas para el cálculo proposicional. Curiosamente, dichas axiomáticas, en principio algo técnico, reflejan a menudo también aspectos filosóficamente significativos del tipo de fenómenos que se tratan de modelar, y al mismo tiempo plantean preguntas importantes sobre los mismos (por ejemplo, ¿es correcto asumir que, en lo que al conocimiento respecta, se da el principio de "introspección negativa"; es decir, que *sabemos que no sabemos algo*?) Dejamos estas preguntas para un capítulo posterior en el que presentamos los principios del cálculo proposicional para cada uno de los sistemas desarrollados (7).

Definición 2.5. *Cuando se tenga $M, s \vDash \varphi$ para todo $s \in \mathcal{D}(M)$, escribiremos $M \vDash \varphi$ y diremos que φ es verdadera en M. Si $M \vDash \varphi$ para todo modelo en una determinada familia χ, diremos que φ es válida en χ y escribiremos $\chi \vDash \varphi$. Si χ es $\mathcal{K}$, donde $\mathcal{K}$ es la familia de todos los modelos de Kripke, podremos escribir simplemente $\vDash \varphi$, y diremos que φ es válida. De manera análoga a la nota 2.6 de más arriba, indicaremos con el símbolo $\nvDash$ la negación de cualquiera de estos hechos; es decir, la existencia de algún contraejemplo para cualquiera de los casos.* □

Definición 2.6. ***Algunas familias de modelos interesantes.***

- *La clase de todos los modelos de Kripke se denotará* $\mathcal{K}$.
- *Se dirá que* R_a *es serial si para todo* s *existe* t *tal que* R_ast. *La clase de los modelos de Kripke* $\{M = \langle S, R, V\rangle \mid$ *Todo* R_a *es serial*$\}$ *se denotará* $\mathcal{KD}$.
- *Se dirá que* R_a *es reflexiva si para todo* s, R_ass. *Análogamente al punto anterior, la clase de los modelos cuyas relaciones son todas reflexivas se denotará* $\mathcal{T}$.
- *Se dirá que* R_a *es transitiva si para todos* s, t, u, *se satisface que si* R_ast *y* R_atu, *entonces* R_asu. *La clase de los modelos transitivos (es decir, aquellos cuyas relaciones son todas transitivas) se denotará* $\mathcal{K}4$. *La clase de los modelos reflexivos transitivos se denotará* $\mathcal{S}4$.
- R_a *se dice euclídea si para todos* s, t, u, *si* R_ast *y* R_asu *entonces* R_atu. *La clase de los modelos euclídeos transitivos se denotará* $\mathcal{K}45$. *La clase de los modelos euclídeos, transitivos y seriales se denotará* $\mathcal{KD}45$.
- R_a *se dirá una relación de equivalencia si es reflexiva, transitiva y simétrica (si* R_ast, *entonces* R_ats). *Esto es equivalente a decir que es reflexiva, transitiva y euclídea. La clase de los modelos cuyas relaciones son de equivalencia se denotará* $\mathcal{S}5$. *En el caso de que un modelo* M *pertenezca a* $\mathcal{S}5$, *diremos que se trata de un modelo epistémico.*

□

A continuación, presentamos algunas de las propiedades básicas que se satisfacen en las familias de modelos que vamos a estudiar:

Proposición 2.2. **Omnisciencia lógica.** Sean φ y ψ fórmulas en el lenguaje $\mathcal{L}_K$, y sea a un agente en algún modelo M. Entonces se tienen:

$$\mathcal{K} \vDash K_a\varphi \wedge K_a(\varphi \to \psi) \to K_a(\psi) \tag{LO1}$$

$$\mathcal{K} \vDash \varphi \Longrightarrow \mathcal{K} \vDash K_a\varphi \tag{LO2}$$

$$\mathcal{K} \vDash \varphi \to \psi \Longrightarrow \mathcal{K} \vDash K_a\varphi \to K_a\psi \tag{LO3}$$

$$\mathcal{K} \vDash \varphi \leftrightarrow \psi \Longrightarrow \mathcal{K} \vDash K_a\varphi \leftrightarrow K_a\psi \tag{LO4}$$

$$\mathcal{K} \vDash (K_a\varphi \wedge K_a\psi) \to K_a(\varphi \wedge \psi) \tag{LO5}$$

$$\mathcal{K} \vDash K_a\varphi \to K_a(\varphi \vee \psi) \tag{LO6}$$

$$\mathcal{S}5 \vDash \neg(K_a\varphi \wedge K_a\neg\varphi) \tag{LO7}$$

Demostración. **LO1.** Tenemos que probar que se tiene $K_a\varphi \wedge K_a(\varphi \to \psi) \to K_a\psi$ para un par M, s arbitrario. Sean pues M un modelo y s un estado arbitrario. Supongamos que $M, s \vDash K_a\varphi \wedge K_a(\varphi \to \psi)$. Tenemos que probar que, entonces, $M, s \vDash K_a\psi$. Veámoslo:

1. $M, s \vDash K_a\varphi \wedge K_a(\varphi \to \psi)$
2. $M, s \vDash K_a\varphi$ *y* $M, s \vDash K_a(\varphi \to \psi)$
3. *Para todo* $t \in S$: $R_ast \Longrightarrow M, t \vDash \varphi$
   Para todo $t \in S$: $R_ast \Longrightarrow M, t \vDash \varphi \to \psi$
4. *Para todo* $t \in S : R_ast \Longrightarrow M, t \vDash \varphi$ *y* $M, t \vDash \varphi \to \psi$
5. (*Aplicando Modus Ponens*) *Para todo* $t \in S : R_ast \Longrightarrow M, t \vDash \psi$
6. $M, s \vDash K_a\psi$

LO2. Sea M, s un modelo puntuado arbitrario. Tenemos que ver que para cualquier t con R_ast, $M, t \vDash \varphi$. Como $\mathcal{K} \vDash \varphi$, entonces, en particular, $M, t \vDash \varphi$ para cualquier t, y tenemos el resultado.

LO3. Sean φ y ψ tales que $\mathcal{K} \vDash \varphi \to \psi$. Sea M, s arbitrario, y supongamos que $M, s \vDash K_a\varphi$. Tenemos que probar que $M, s \vDash K_a\psi$. Por **LO2**, tenemos que $M, s \vDash K_a(\varphi \to \psi)$. Por **LO1**, tenemos el resultado.

LO4. Trivial a partir de LO3.

LO5 y LO6. Consideramos que las pruebas son lo suficientemente similares a la de **LO1** como para como para obviarlas.

LO7. Sea M, s un modelo epistémico puntuado ($M \in \mathcal{S}5$). Supongamos que se tiene $M, s \vDash K_a\varphi \wedge K_a\neg\varphi$. Entonces se tiene $M, s \vDash K_a\varphi$ y $M, s \vDash K_a\neg\varphi$. Ahora bien, como el modelo M es epistémico, en particular R_a es reflexiva, por lo que se tiene $M, s \vDash \varphi$ y también se tiene $M, s \vDash \neg\varphi$, con lo que llegamos a una contradicción, de lo que se sigue el resultado.

Q.E.D.

□

Observación 2.2. En realidad, la propiedad **LO7** se tiene en una clase más general de modelos, a saber, en $\mathcal{KD}$ (la clase de los modelos seriales). En efecto: sea M, s un modelo serial puntuado. Supongamos que se tiene $M, s \vDash K_a\varphi \wedge K_a\neg\varphi$. Dado que el modelo es serial, existe $t \in S$ con R_ast, y por la semántica de K_a, $M, t \vDash \varphi$ y $M, t \vDash \neg\varphi$, lo cual nos lleva a una contradicción y tenemos el resultado. □

El hecho de que se tengan las propiedades de la proposición anterior se conoce como el *problema de la omnisciencia lógica (**L**ogical **O**mniscience)*, dado que lo que expresan en conjunto es que los agentes son razonadores perfectos. Por ejemplo, $LO1$ expresa que el conocimiento es cerrado bajo consecuentes. $LO2$ expresa que los agentes conocen todas las propiedades que sean universalmente válidas. Las propiedades $LO3-$

$LO6$ expresan que el agente es capaz de realizar deducciones lógicas a partir de su propio conocimiento, y $LO7$ expresa que el conocimiento del agente es internamente consistente. A menudo se considera que esta es una de las mayores limitaciones del enfoque de la lógica epistémica a la hora de analizar "situaciones de conocimiento", dado que la mayoría de los seres inteligentes con la capacidad de albergar y producir conocimiento no son, a pesar de todo, razonadores perfectos. En nuestra conclusión (8) trataremos de ofrecer nuestras propias consideraciones sobre este aspecto.

2.3.2. Bisimulaciones

El concepto de *bisimulación* es una de las nociones más importantes en el estudio teórico de los modelos de Kripke. Sin adelantarnos demasiado al resultado que demostraremos a continuación, este concepto es básicamente una formalización de la siguiente idea: dado un modelo puntuado M, s, ¿hasta qué punto podemos "alterar el modelo" sin que ningún agente pueda distinguir el modelo original del modificado? Lo que se busca con este concepto es una noción de "equivalencia" entre modelos de Kripke, en un sentido muy concreto.

Definición 2.7. ***Bisimulación.*** *Sean dos modelos $M = \langle S, R, V\rangle$ y $M' = \langle S', R', V'\rangle$. Una relación no vacía $\mathfrak{R} \subseteq S \times S'$ se dice bisimulación sii para cualquier par $(s, s') \in \mathfrak{R}$ se tienen las siguientes propiedades:*

- ***átomos*** *$s \in V(p)$ si y solo si $s' \in V'(p)$, para todo $p \in At$.*
- ***ida*** *Para todo $a \in A$ y todo $t \in S$, si $R_a st$, entonces existe $t' \in S'$ tal que $R'_a s't'$ y $(t, t') \in \mathfrak{R}$.*
- ***vuelta*** *Para todo $a \in A$ y todo $t' \in S'$, si $R'_a s't'$, entonces existe $t \in S$ tal que $R_a st$ y $(t, t') \in \mathfrak{R}$.*

Denotamos que existe una bisimulación entre M y M' en los estados s y s' con la notación $(M, s) \underline{\leftrightarrow} (M', s')$; en tal caso, decimos que (M, s) y (M', s') son bisimilares. □

Obviamente, si M, s y M', s' son bisimilares, la condición **átomos** garantiza la existencia de un "acuerdo" en las fórmulas básicas en todos los estados relacionados a través de $\mathfrak{R}$; en un contexto epistémico, la condición **ida** "preserva la ignorancia" de fórmulas al pasar de M, s a M', s', y la condición **vuelta** "preserva su conocimiento". Esto se ve de forma clara en la demostración del siguiente teorema, que enuncia la indistinguibilidad de modelos bisimilares en el lenguaje $\mathcal{L}_K$.

Definición 2.8. *Escribimos $(M, s) \equiv_{\mathcal{L}_K} (M', s)$ para denotar que $M, s \vDash \varphi$ sii $M', s' \vDash \varphi$ para cualquier fórmula $\varphi \in \mathcal{L}_K$. En tal caso, decimos que los modelos puntuados son equivalentes en $\mathcal{L}_K$.* □

Teorema 2.1. *Dados dos modelos puntuados* (M, s), (M', s'), *si se tiene* $(M, s) \underline{\leftrightarrow} (M', s')$, *entonces también se tiene* $(M, s) \equiv_{\mathcal{L}_K} (M', s')$.

Demostración. Procedemos por inducción estructural sobre φ. Primero hacemos la prueba en un sentido: $M, s \vDash \varphi \implies M', s' \vDash \varphi$.

> **Caso base:** Dados $(M, s) \underline{\leftrightarrow} (M', s')$ y $p \in At$, por **átomos**, $M, s \vDash p$ si y solo si $M', s' \vDash p$.
>
> **Paso de inducción:** A continuación supondremos que las fórmulas φ y ψ satisfacen la siguiente propiedad (hipótesis de inducción): para cualesquiera $(M, s) \underline{\leftrightarrow} (M', s')$, $M, s \vDash \varphi$ si y solo si $M', s' \vDash \varphi$ (respectivamente para ψ).
>
> > **Negación:** Supongamos que se tiene $M, s \vDash \neg\varphi$. Esto es equivalente a $M, s \nvDash \varphi$, y por h.d.i. esto es equivalente a $M', s' \nvDash \varphi$, esto es, $M', s' \vDash \neg\varphi$.
> >
> > **Conjunción:** Supongamos que se tiene $M, s \vDash \varphi \wedge \psi$, esto es, $M, s \vDash \varphi$ y $M, s \vDash \psi$. Por hipótesis de inducción, esto es equivalente a $M', s' \vDash \varphi$ y $M', s' \vDash \psi$, es decir, $M', s' \vDash \varphi \wedge \psi$.
> >
> > **Operador modal:** Supongamos que se tiene $M, s \vDash K_a\varphi$. Sea t' arbitrario de forma que $R_a s't'$. Por **vuelta** sabemos que existe un $t \in S$ tal que $R_a st$ y $(t, t') \in \mathfrak{R}$. Por hipótesis de inducción, $M, t \vDash \varphi$ si y solo si $M', t' \vDash \varphi$. Dado que $M, s \vDash K_a\varphi$, por la semántica se tiene $M, t \vDash \varphi$, y por lo tanto $M', t' \vDash \varphi$. Dado que t' es arbitrario, esto se tiene para cualquier t' tal que $R_a s't'$; es decir, $M', s' \vDash K_a\varphi$.

La prueba en el otro sentido es análoga, solo que en el caso del operador modal se hace uso de la condición **ida** en lugar de **vuelta**.

Q.E.D.

□

Observación 2.3. Hemos visto que la existencia de una bisimulación es una condición suficiente para que dos estados verifiquen el mismo conjunto de fórmulas. Algún lector puede preguntarse si se trata también de una condición necesaria, y la respuesta es que no es así: en efecto, un contraejemplo son los dos modelos que describiremos en el siguiente ejemplo. Si el lector quiere indagar adicionalmente en las relaciones entre bisimulación y equivalencia, le referimos al capítulo 8 (Expresividad) de van Ditmarsch [1]. □

Ejemplo 2.2. (Van Ditmarsch [1], capítulo 8)

Sea P un conjunto numerable de átomos, numerados de tal forma que p_n es el n-ésimo de la numeración. Consideramos los modelos de un solo agente $M_1 = \langle S^1, R^1, V^1 \rangle$ y $M_2 = \langle S^2, R^2, V^2 \rangle$, con:

$$\begin{array}{ll} S^1 = \{s^1\} \cup \mathbb{N} & S^2 = \{s^2, \omega\} \cup \mathbb{N} \\ R^1 = \{s^1\} \times \mathbb{N} & R^2 = \{s^2\} \times (\mathbb{N} \cup \{\omega\}) \\ V^1(p_n) = \{n\} & V^2(p_n) = \{n\} \end{array} \tag{2.5}$$

El modelo M_2 tiene un estado adicional ω en el que ningún átomo es cierto. En este caso, no es difícil ver que los modelos M_1, s^1 y M_2, s^2 no son bisimilares, dado que no se puede dar la condición de **vuelta**: en M^2 el estado ω es accesible desde s^2, pero en M_1 no hay ningún estado en el que ningún átomo sea cierto que sea accesible desde s^1.

Veamos ahora que, sorprendentemente, dichos modelos sí son *equivalentes*.

Demostración. Procedemos por inducción estructural sobre las fórmulas.

Caso base. Obvio, dado que en s^1 y en s^2 no se satisface ningún átomo.

Hipótesis de inducción. Sea $\varphi \in \mathcal{L}_K$ de profundidad a lo sumo m. Entonces $M_1, s^1 \vDash \varphi \iff M_2, s^2 \vDash \varphi$.

Operadores negación y conjunción. Se comprueban fácilmente.

Operador epistémico. Sea $\varphi = K\psi$, con ψ de profundidad a lo sumo m. Supongamos que φ puede distinguir ambos modelos (es decir, se tiene en uno de ellos pero no en el otro). Observemos que cada estado $n \in \mathbb{N}$ satisface las mismas fórmulas en ambos modelos: efectivamente, cada uno satisface el átomo p_n correspondiente y no satisface ningún otro, y, además, desde estos estados ningún estado es accesible (ni siquiera él mismo), de modo que las fórmulas modales que se satisfacen para cada estado son las mismas en ambos modelos. Por lo tanto, la única forma de φ distinga ambos modelos es que se tenga $M_1, s^1 \vDash \varphi$ pero no se tenga $M_2, s^2 \vDash \varphi$, o sea, que se tenga $M_i, k \vDash \psi$ para todo $k \in \mathbb{N}$, $i = 1, 2$, pero que se tenga también $M_2, \omega \nvDash \psi$.

Ahora bien, la fórmula ψ es finita, por lo que solo aparece en ella a lo sumo una cantidad finita de átomos. Sea n el mayor natural tal que p_n aparece en ψ. Claramente, los estados $(n+1)$ y ω deben coincidir en φ ($M_2, n+1 \vDash \psi \iff M_2, \omega \vDash \psi$). En particular, como $M_2, \omega \nvDash \psi$, tenemos también que $M_2, n+1 \vDash \psi$. Pero esto contradice el hecho de que ψ se tiene en todos los estados $k \in \mathbb{N}$. Concluimos pues que, en efecto, se tiene $M_1, s^1 \equiv_{\mathcal{L}_K} M_2, s^2$.

Q.E.D.

□

3. Lógica epistémica dinámica con anuncios públicos

La exposición de este capítulo está basada principalmente en el capítulo 4 de Van Ditmarsch [1].

3.1. Idea y motivación

En el capítulo anterior hemos presentado una lógica para razonar sobre situaciones de conocimiento que involucran a varios agentes, y, si bien en un ámbito puramente teórico el estudio de esta lógica y de las posibles construcciones que se pueden realizar en torno a ella ya nos ofrece una enorme variedad de vías de investigación de gran interés, en un contexto práctico podría ser razonable pedir algo más de nuestras herramientas conceptuales. Más concretamente, el mundo – al margen de consideraciones filosóficas en las que tratemos de concebir la sucesión de todos los instantes del tiempo como una simultaneidad en algún plano de abstracción superior al de nuestra experiencia cotidiana – es un lugar dinámico, caracterizado por una infinidad de cambios que se dan de un momento a otro, y que es importante tener en cuenta si queremos tratar de modelar alguna situación real de manera práctica.

Conceptualmente, y de manera provisional, podemos considerar básicamente dos tipos de cambios que pueden tener lugar en una situación epistémica, donde varios agentes razonan sobre el conocimiento que ellos mismos y los demás tienen sobre un determinado conjunto de hechos del mundo. El primer tipo tiene que ver con los cambios que se pueden dar en el mundo como tal, y en particular, en el conjunto de hechos que estarían considerando los agentes. En este trabajo no se tratará este tipo de cambios, y tampoco son, por lo general, los que se están considerando cuando se habla de "lógica epistémica dinámica". Más bien, se trata del segundo tipo de cambios: aquellos que tienen que ver con la información de la que disponen los propios agentes sobre un mismo conjunto de hechos, y con lo que las acciones y el comportamiento de estos mismos agentes *revela* o *comunica* sobre la información de la que disponen.

Un primer enfoque para modelar este tipo de cambios es el de los *anuncios públicos*. Un modelo conceptual básico para pensar sobre los anuncios públicos es el siguiente: además de los agentes involucrados en la situación epistémica, existe también un "agente externo omnisciente", que conoce tanto la situación epistémica como el estado actual del mundo, y que comunica información veraz a todos los agentes involucrados de forma pública (es decir, de tal forma que cada uno de los agentes es consciente de que el resto de los agentes también han obtenido esta información), con la que dichos agentes *actualizan* sus modelos mentales de la situación epistémica. No obstante, existen situaciones más allá de este modelo conceptual básico que también podrían ser modeladas perfectamente con anuncios públicos: sin ir más lejos, en una situación en la que alguno de los agentes involucrados compartiese cierta información y el resto de los

agentes tuviesen garantías de que esta información es veraz también podría considerarse que lo que ha tenido lugar es un *anuncio público*; no obstante, lo que se anuncia en este caso no es simplemente la información como tal, sino también el hecho de que el agente que la ha anunciado la conocía anteriormente (en términos de nuestro lenguaje formal, lo que se anuncia no es φ, sino $K_a\varphi$).

Por supuesto, los modelos de anuncios públicos tienen grandes limitaciones, dado que no son capaces (por sí solos) de capturar la enorme complejidad y diversidad de situaciones comunicativas que pueden darse entre varios agentes epistémicos (de hecho, la segunda parte del ejemplo que planteamos sobre la situación GLO (2.1) ya no se ajustaría a este paradigma, dado que el agente a sabe que el agente c va a comunicar cierta información al agente b, pero no conoce exactamente el contenido de esta información); no obstante, los presentamos como "primera incursión" hacia la lógica epistémica dinámica dado que puede considerarse que se tratan del "caso más simple" dentro de este ámbito, y, con todas sus limitaciones, ya nos permiten atacar una cantidad considerable de problemas que de otra forma sería mucho más tedioso plantear y analizar con rigor. Sorprendentemente, no puede decirse que los anuncios públicos, al menos por sí solos, sean estrictamente hablando *imprescindibles* para plantear el tipo de situaciones para las que fueron diseñados: como veremos más adelante, el lenguaje lógico $\mathcal{L}_{K[]}$ (lógica epistémica con anuncios públicos) es *igual de expresivo*[1] que el lenguaje $\mathcal{L}_K$. Esto no significa que los anuncios públicos sean un despropósito y que no merezca la pena estudiarlos: como mínimo, disponer de anuncios públicos supone una ventaja enorme desde el punto de vista del "usuario", tanto por la claridad conceptual que proporcionan como por simplificar enormemente las fórmulas que se hace necesario manejar; y, como veremos en el capítulo 6, el lenguaje $\mathcal{L}_{KC[]}$ *sí* que es más expresivo que el lenguaje $\mathcal{L}_{KC}$.

Finalmente, y antes de proseguir con la presentación formal del nuevo lenguaje, es importante hacer un pequeño aviso: el modelo de anuncio público que vamos a presentar en este capítulo no es el único que se ha propuesto, y tampoco me atrevería a asegurar que es el "paradigmático" o el más ampliamente extendido y aceptado dentro del campo de la lógica epistémica, aunque posiblemente lo sea[2]. En el capítulo 5 trabajaremos brevemente con un modelo de anuncios públicos, utilizado en el artículo de Kooi [4] sobre Lógica Dinámica Probabilística, y tratado brevemente también en van

[1] En este trabajo mencionamos a veces el concepto de "expresividad", o decimos que "un lenguaje es más expresivo que otro", pero nunca estudiamos rigurosamente lo que esto quiere decir. La idea básica es la siguiente: dos lenguajes son igual de expresivos si para cualquier fórmula en uno de ellos, existe una fórmula equivalente (que toma los mismos valores de verdad en los mismos estados) en el otro, y viceversa; por lo tanto, un lenguaje A será más expresivo que otro lenguaje B si todas las fórmulas de B tienen una fórmula equivalente en A, pero existe alguna fórmula de A que no tiene ninguna fórmula equivalente en B. Si el lector está interesado en indagar más al respecto, lo referimos a van Ditmarsch [1] (capítulo 8).

[2] De paso, creo que con esto se hace apropiado mencionar algo importante sobre la lógica epistémica, y más generalmente la lógica modal, como campos de investigación, y es que se tratan de campos extremadamente heterogéneos y poco estandarizados, al menos en comparación con otros campos y en base a la minúscula experiencia que tengo trabajando en el mismo. De hecho, conforme he ido avanzando en este trabajo, uno de los objetivos que me he propuesto ha sido el de proporcionar herramientas que puedan facilitar, al cabo del tiempo, una posible estandarización; aunque, desde luego, entiendo también los beneficios de tener un campo heterogéneo y dinámico como el que tenemos ahora mismo.

Ditmarsch [1] (capítulo 4, sección 9), distinto al que presentaremos aquí. Hay varias razones por las que hemos considerado apropiado trabajar con *este* modelo de anuncios públicos y no con algún otro, que podrían resumirse en:

i) Por continuidad con el modelo proporcionado en van Ditmarsch [1], que es una de nuestras referencias principales.

ii) Porque consideramos que es el más apropiado para la lógica epistémica, dado que, en cierto modo, "asegura" que los únicos anuncios públicos que tenga sentido hacer sean precisamente aquellos que son veraces (el modelo de anuncios públicos utilizado en Kooi [4] podría considerarse, por otra parte, más apropiado para situaciones doxásticas).

iii) Porque consideramos que el estudio del operador dual del "operador anuncio público" proporcionado por este modelo (esto se verá más adelante) también es de interés especial.

iv) En definitiva, por preferencia personal.

v) Y también porque, en todo caso, había que tomar una decisión arbitraria sobre el modelo de anuncios públicos que se presentaría en este trabajo.

3.2. Sintaxis y semántica de la lógica epistémica dinámica

Definición 3.1. ***Sintaxis del lenguaje*** $\mathcal{L}_{K[]}$**.** *Dado un conjunto finito de agentes A y un conjunto numerable de proposiciones atómicas At, definimos el lenguaje epistémico con anuncios públicos $\mathcal{L}_{K[]}(A, At)$ (por lo general, como ya es costumbre, obviaremos A y At en la notación anterior) con la siguiente BNF:*

$$\varphi ::= p \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid K_a\varphi \mid [\varphi_1]\varphi_2 \tag{3.1}$$

Con $p \in At$ y $a \in A$. □

La nueva construcción que se ha introducido nos permite escribir fórmulas de la forma $[\varphi]\psi$. Como veremos más adelante, el operador $[\,]$ (operador de anuncio público o "corchetes caja") es de tipo necesidad, de modo que, estrictamente, en lenguaje natural la fórmula $[\varphi]\psi$ debería leerse "después de *cualquier* anuncio público y veraz de φ se tiene ψ"; no obstante, dado que, como también veremos más adelante, este operador tiene la propiedad de ser *funcional* respecto a su interpretación en el conjunto de los modelos de Kripke puntuados[3], podemos permitirnos leerla de forma más simplificada como "tras anuncio público y veraz de φ se tiene ψ". El dual de $[\,]$ se denotará $\langle\,\rangle$ (operador de anuncio público dual o "corchetes diamante"), y la fórmula $\langle\varphi\rangle\psi$ se leerá "tras *algún* anuncio público y veraz de φ se tiene ψ". Quizá ahora mismo la diferencia entre estos dos operadores pueda parecer demasiado sutil y esquiva, pero con la

[3]Es decir, dada una fórmula φ, el conjunto de las fórmulas $\{\psi_i\}_{i\in\Lambda}$ donde se satisface la fórmula $[\varphi]\psi_i$ solo depende del estado M, s donde esta se evalúe.

correspondiente semántica y algunos ejemplos ilustrativos quedará todo mucho más claro.

Nota 3.1. En contextos más generales, la palabra "anuncio" es a veces sustituida por "actualización" (mucho más apropiada, por ejemplo, en el contexto del análisis de sistemas distribuidos). □

Definición 3.2. ***Semántica de*** $\mathcal{L}_{K[]}$**.** *Dado un modelo de Kripke* $M = \langle S, R, V\rangle$ *con agentes* A *y átomos* At*,*

$$\begin{array}{lll} M, s \vDash p & sii & s \in V(p) \\ M, s \vDash \neg\varphi & sii & M, s \nvDash \varphi \\ M, s \vDash \varphi \wedge \psi & sii & M, s \vDash \varphi \text{ y } M, s \vDash \psi \\ M, s \vDash K_a\varphi & sii & \text{Para todo } t \in S: \ R_a st \Rightarrow M, t \vDash \varphi \\ M, s \vDash [\varphi]\psi & sii & (M, s \vDash \varphi) \implies (M_{|\varphi}, s \vDash \psi) \ (\%) \end{array} \tag{3.2}$$

Donde $M_{|\varphi} = \langle S', R', V'\rangle$ *se denomina "la restricción del modelo* M *a la proposición* φ*", y se define como sigue:*

- $S' = S(\varphi) := \{s \in S \mid M, s \vDash \varphi\}$
- $R'_a := R_a \cap (S' \times S')$
- $V'_p := V_p \cap S'$

Y donde la marca $(\%)$ *indica que la cláusula debe ser evaluada perezosamente*[4]*.* □

De nuevo, es importante tratar de comprender informalmente las construcciones que hemos definido de manera formal y rigurosa. En este caso, la intuición que hay tras el formalismo es bastante fácil de ver: tras anunciarse públicamente una determinada información, los agentes proceden a "actualizar" sus modelos mentales de la situación, descartando todos los estados que son "incompatibles" con la información que han recibido, y continuan razonando en estos nuevos modelos actualizados. Ilustremos esto con un ejemplo básico, extraído parcialmente de van Ditmarsch [1].

Ejemplo 3.1. Ana (a), Benito (b) y Catalina (c) han extraído respectivamente una carta de un montón de tres, 0, 1 y 2, y todos ellos conocen comúnmente estas circunstancias. Ana ha extraído la carta 0, Benito ha extraído la carta 1 y Catalina ha extraído la carta 2. Obviamente, cada "jugador" conoce la carta que le ha tocado, pero no conoce las de los demás. Una forma de representar esta situación es con el modelo de la figura 3.1, donde los estados vienen representados por ternas ABC, donde A representa

[4]La evaluación perezosa es una estrategia de evaluación de expresiones que evita calcular el valor de sub-expresiones hasta que estos sean necesarios. Por ejemplo: una forma no-perezosa (impaciente) de evaluar la expresión $0*(1+1)$ sería $0*(1+1) = 0*2 = 0$. Una forma perezosa sería $0*(1+1) = 0$, haciendo uso de una regla que afirme que el producto de 0 por cualquier suma de naturales es también nula. En el caso de nuestra definición, el hecho de que la evaluación sea perezosa significa que, si el antecedente es falso, no hay que evaluar el consecuente. La razón para especificar este detalle aparentemente irrelevante se expondrá más adelante, pero, básicamente, lo que se busca es evitar problemas de indefinición.

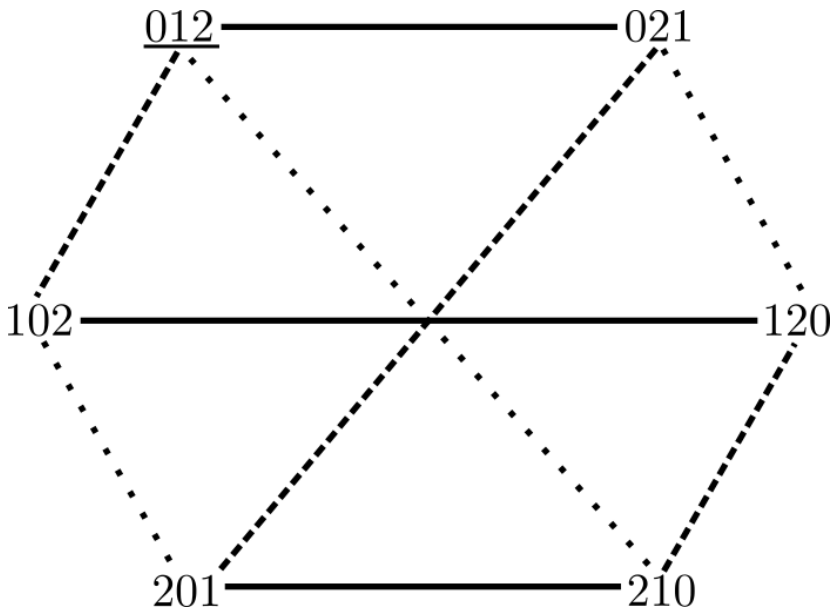

Figura 3.1: Estados posibles en el ejemplo 3.1.

la carta que le ha tocado a a, y así sucesivamente. Las líneas continuas representan las relaciones de accesibilidad de a, las líneas de puntos representan las relaciones de accesibilidad de b, y las líneas discontinuas repreesentan las relaciones de accesibilidad de c (en todos casos obviando las relaciones reflexivas). En nuestro ejemplo, el estado actual del mundo es el 012.

Digamos que los hechos se representan con átomos de la forma 0_a, "a tiene la carta 0", y llamemos $Hexa$ a nuestro modelo, por darle un nombre ingenioso. En este modelo y en el estado 012, se tiene que "Ana sabe que nadie más sabe cuál es su carta", que formalmente se escribiría:

$$Hexa, 012 \vDash K_a\neg\big((K_b0_a \vee K_b1_a \vee K_b2_a) \vee (K_c0_a \vee K_c1_a \vee K_c2_a)\big)$$

Supongamos ahora que un agente externo con información privilegiada, Donoso, anuncia públicamente la proposición $\neg 1_a$ ("Ana no tiene la carta 1"). Con esto, el modelo actualizado o restringido $Hexa_{|\neg 1_a}$ quedaría como en la figura 3.2, y en esta nueva situación Ana sabe que o bien Benito o bien Catalina (pero no ambos) saben cuál es su carta. Abreviando $(K_x0_y \vee K_x1_y \vee K_x2_y)$ como $knows_{x,y}$, esto se escribiría:

$$Hexa_{|\neg 1_a}, 123 \vDash K_a\big((knows_{b,a} \wedge \neg knows_{c,a}) \vee (knows_{c,a} \wedge \neg knows_{b,a})\big)$$

El operador de anuncio público nos permite expresar la "transición" de un modelo a otro en una sola fórmula:

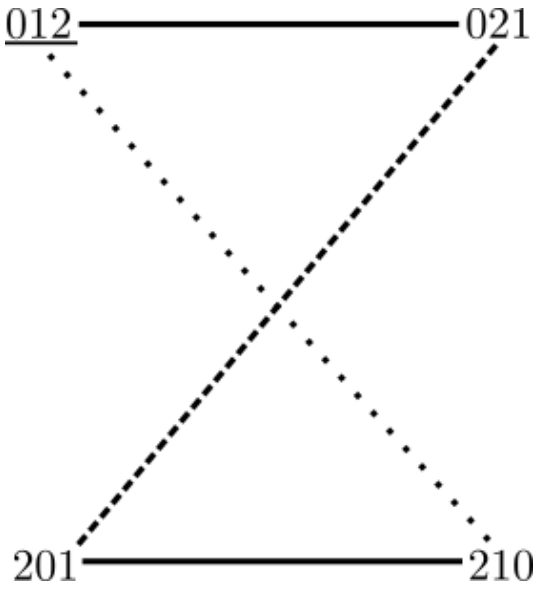

Figura 3.2: Modelo actualizado con la proposición $\neg 1_a$.

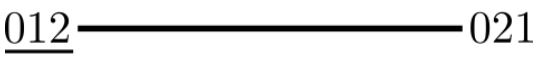

Figura 3.3: Modelo actualizado con la proposición $K_a \neg 0_b$.

$$Hexa, 012 \vDash K_a \neg (knows_{b,a} \vee knows_{c,a}) \wedge [\neg 1_a] K_a \big((knows_{b,a} \wedge \neg knows_{c,a}) \vee (knows_{c,a} \wedge \neg knows_{b,a}) \big)$$

De hecho, quizá sería todavía más apropiado utilizar en este caso el operador dual, pero dejaremos estas consideraciones para más adelante.

Ahora supongamos que no es un agente externo quien hace el anuncio público, sino que es Ana quien anuncia públicamente: "Benito no tiene la carta 0". Los únicos estados en los que Ana *sabe* que Benito no tiene la carta 0, es decir, los estados en los que se tiene $K_a \neg 0_b$, son precisamente aquellos donde es ella quien tiene esta carta. En esta nueva situación (figura 3.3), los agentes b y c conocen exactamente la distribución de las cartas entre los jugadores, pero el agente a no la conoce. Esto tiene sentido, dado que la información que se ha anunciado públicamente era algo que Ana *ya sabía*, mientras que, para Benito y Catalina, la información sí aporta algo nuevo. De hecho, observamos, comparando este anuncio público con el que planteamos antes, que los anuncios públicos de la forma $[K_a \varphi]$ parecen revelar "más información" que anuncios públicos "neutrales" $[\varphi]$. Esto *también* tiene sentido, dado que, al menos en modelos epistémicos, la validez de una proposición $K_a \varphi$ implica la validez de la proposición φ, pero el recíproco (obviamente) no es cierto.

□

Observación 3.1. **Evaluación perezosa de los operadores de anuncio público.** Antes de proseguir, es importante hacer un pequeño apunte técnico sobre este operador y su dual para evitar confusiones mayores en el futuro. Comenzamos observando que, de acuerdo a la definición general de los operadores duales, la fórmula $\langle\varphi\rangle\psi$, en realidad una abreviatura de $\neg[\varphi]\neg\psi$, se interpretaría mediante la cláusula semántica:

$$M, s \vDash \langle\varphi\rangle \quad sii \quad (M, s \vDash \varphi)\ y\ (M_{|\varphi}, s \vDash \psi)(\%) \tag{3.3}$$

Tanto en el caso del operador "primal" como del operador dual, es importante que la cláusula se interprete evaluando de forma perezosa la parte derecha de la misma. Quizá un lector sagaz ya haya visto por qué esto es así, pero, en cualquier caso, la razón no es compleja, y se puede ilustrar con un ejemplo extremadamente simple. Supongamos que M es un modelo con un solo estado s, en el que se satisfacen los átomos p y q. ¿Cómo se interpretaría entonces la validez de la fórmula $[\neg p]q$ en el estado M, s?

Tenemos que ver si se tiene $(M, s \vDash p) \Longrightarrow (M_{|p}, s \vDash q)$. No obstante, si evaluásemos esta expresión de forma impaciente, nos encontraríamos con un problema, dado que el conjunto de estados del modelo $M_{|p}$ es el vacío, y, en particular, $s \notin \emptyset$. Dado que no sabemos interpretar la validez de fórmulas sobre estados que no pertenecen al conjunto de estados del modelo, el resultado sería indefinido. En cambio, si evaluamos la expresión de forma perezosa y aplicamos la regla de que "si el antecedente de una implicación es falso, entonces, independientemente del consecuente, esta es verdadera", tendríamos que la fórmula se interpreta como verdadera – lo cual tiene sentido si consideramos que "del anuncio público de algo falso se puede deducir cualquier cosa".

El mismo problema lo tendríamos si quisiésemos interpretar la validez de $\langle\neg p\rangle q$, y, de nuevo, la evaluación perezosa nos lo resuelve si aplicamos la regla de que "si el primer término de una conjunción es falso, entonces, independientemente del segundo término, la conjunción es falsa", y consideramos que el "autómata" encargado de interpretar estas expresiones siempre evalúa en las conjunciones el primer término antes que el segundo. □

3.3. Conceptos importantes de la lógica epistémica con anuncios públicos

La primera propiedad que enunciaremos es de especial importancia en el contexto de la lógica epistémica, ya que nos dice que si el modelo del que partimos es epistémico, su restricción tras un anuncio público nos proporciona un modelo que sigue siendo epistémico; esta propiedad, por lo tanto, es indispensable para poder hacer uso de los anuncios públicos en este contexto.

Proposición 3.1. Sea $M \in \mathcal{S}5$ y $\varphi \in \mathcal{L}_{K[]}$. Entonces $M_{|\varphi} \in \mathcal{S}5$.

Demostración. Sea $a \in A$ un agente arbitrario. Si R_a es la relación de accesibilidad asociada a a en el modelo M y R'_a es la relación de accesibilidad asociada a a en el modelo restringido, tenemos que comprobar tres cosas:

- R'_a es reflexiva.
- R'_a es simétrica.
- R'_a es transitiva.

Reflexiva. Trivial.

Simétrica. Sean $s, t \in S(\varphi)$ con $(s,t) \in R'_a$, esto es, $(s,t) \in R_a \cap S(\varphi)^2$. Por simetría de R_a, tenemos:

$$(s,t) \in R_a \Longrightarrow (t,s) \in R_a$$

Por otra parte, es obvio que $(s,t) \in S(\varphi)^2 \Longrightarrow (t,s) \in S(\varphi)^2$. Usando ambas implicaciones, obtenemos finalmente:

$$(t,s) \in R'_a$$

Transitiva. Sean $s, t, u \in S(\varphi)$ con $(s,t), (t,u) \in R'_a = R_a \cap S(\varphi)^2$. Por transitividad de R_a, tenemos:

$$(s,t), (t,u) \in R_a \Longrightarrow (s,u) \in R_a$$

Y, obviamente, $(s,t), (t,u) \in S(\varphi)^2 \Longrightarrow (s,u) \in S(\varphi)^2$. Usando ambas implicaciones, obtenemos finalmente:

$$(s,u) \in R'_a$$

Q.E.D.

□

A continuación presentamos varias propiedades y conceptos importantes, y a menudo sorprendentes, relacionados con el comportamiento del operador de anuncio público y su dual. Dado que se trata de muchas propiedades y que son bien conocidas en el campo, presentaremos la mayoría de ellas sin demostración. La primera tiene que ver con la relación entre el operador de anuncio público y la negación.

Proposición 3.2. **Anuncios públicos y negación.** La fórmula $[\varphi]\neg\psi \leftrightarrow (\varphi \rightarrow \neg[\varphi]\psi)$ es válida (en cualquier modelo de Kripke).

□

Lo que nos dice este resultado es que, dado un estado M, s, la fórmula $[\varphi]\neg\psi$ puede ser verdad por dos razones. La primera razón es que, *tras llevarse a cabo* un anuncio público y veraz de la fórmula φ en el estado M, s, no se tenga ψ. La segunda razón es que un anuncio público y veraz de la fórmula φ *no pueda hacerse* en el estado M, s (porque φ no es cierta en M, s).

Obsérvese que $\neg[\varphi]\psi$ es equivalente a $\langle\varphi\rangle\neg\psi$. Este lema nos permite hacernos una idea mucho más concreta de la diferencia entre lo que expresan las fórmulas $[\varphi]\psi$ y $\langle\varphi\rangle\psi$: en el primer caso, lo que se está expresando es que ***si pudiese realizarse*** el anuncio público y veraz de φ en el estado M, s, ***entonces*** en el modelo restringido resultante se ***tendría*** ψ. En el segundo caso, lo que se expresa es que ***puede realizarse*** el anuncio público y veraz de φ, ***y*** en el modelo restringido resultante se ***tiene*** ψ. Regresando a la introducción que hicimos inicialmente de estos operadores, quizá ahora tengan más sentido las "lecturas en lenguaje natural" que se hacen de cada uno de ellos: en el caso de la segunda fórmula, tiene sentido que la leamos como "tras *algún* anuncio público de φ se tiene ψ", dado que esto da a entender que, efectivamente, *algún* anuncio público (y veraz) de φ puede realizarse. En el caso de la primera fórmula, que se lea como "tras *cualquier* anuncio público de φ se tiene ψ" daría a entender que *si es que* puede anunciarse públicamente φ, entonces se tiene ψ, y, en otro caso, puede tenerse cualquier cosa.

Las dos proposiciones que presentamos a continuación nos permiten expandir adicionalmente nuestra interpretación de estos operadores.

Proposición 3.3. **Los anuncios son funcionales.** La fórmula

$$\langle\varphi\rangle\psi \to [\varphi]\psi$$

es válida.

□

Proposición 3.4. **Los anuncios son parciales.** La fórmula

$$\langle\varphi\rangle\top$$

no es válida.

Demostración. Basta considerar un estado epistémico M, s donde no se tenga φ.

Q.E.D.

□

Lo que nos dice este par de proposiciones es lo siguiente: por una parte, si un anuncio público puede ejecutarse, entonces solo puede ejecutarse de una forma ($M, s \vDash \varphi$ *y* $M_{|\varphi}, s \vDash \psi$ implica que $M, s \vDash \varphi \Longrightarrow M_{|\varphi}, s \vDash \psi$); es decir, los anuncios públicos son *funcionales*. Por otra parte, los anuncios públicos no siempre pueden ejecutarse; en otras palabras, son *funciones parciales* (su dominio no es el total).

Planteamos ahora un conjunto de situaciones interesantes que puede llegar a parecer contraintuitiva: ¿es posible que el hecho de anunciar algo públicamente haga que lo que se anuncia pase a ser falso? La respuesta es que sí. Consideremos la situación del siguiente ejemplo:

Ejemplo 3.2. Ana (a) y Benito (b) son un matrimonio de opositores. Ana está a la espera de sus resultados en las oposiciones más recientes para profesora de inglés, que deberían estar a punto de salir. Finalmente, los resultados aparecen en la correspondiente página web, y Ana ve que ha obtenido una plaza. Entonces le dice a Benito, que sabe que Ana está atenta a los resultados de las oposiciones, pero que no lo está él mismo: "Supongo que todavía no lo sabes, pero he conseguido una plaza en las oposiciones más recientes".

Planteemos formalmente la situación. Representamos la afirmación "Ana ha obtenido una plaza en las oposiciones" con el átomo op. En el modelo epistémico inicial, hay dos estados posibles del mundo: s_1, donde Ana obtiene la plaza (el estado real del mundo) y s_2, donde Ana no obtiene la plaza ($S = \{s_1, s_2\}$). Ana conoce los resultados de las oposiciones, de modo que su relación de accesibilidad es $R_a = \{(s_1, s_1), (s_2, s_2)\}$. Benito, en cambio, no conoce los resultados, de modo que su relación de accesibilidad es $R_b = S \times S$. Entonces se anuncia públicamente la siguiente proposición: $\neg K_b op \wedge op$[5].

Evidentemente, el anuncio público se puede realizar, dado que $M, s_1 \vDash \neg K_b op \wedge op$. Escribamos $\varphi := \neg K_b op \wedge op$. Ahora bien, en el modelo restringido $M_{|\varphi}$ desaparece el estado s_2, de modo que el único estado accesible a b desde el estado s_1 es el propio s_1 y, por lo tanto, la proposición $K_b op$ pasa a ser cierta ($\neg K_b op$ pasa a ser falsa). En particular, esto significa que $M_{|\varphi} \vDash \neg\varphi$, o lo que es lo mismo,

$$M, s_1 \vDash \langle\varphi\rangle\neg\varphi$$

□

Este tipo de situaciones son de interés especial en el campo de la lógica epistémica dinámica, y dan lugar a los conceptos de *actualización exitosa y no exitosa*. Estos conceptos admiten un estudio en una profundidad mucho mayor de lo que se verá en este trabajo, haciendo uso de las herramientas presentadas en el capítulo 6, entre otras. El lector interesado puede consultar van Ditmarsch [1] (capítulo 4, sección 4.7).

A continuación presentamos otros varios resultados clásicos y bien conocidos. Las demostraciones de algunos de ellos pueden encontrarse también en el capítulo de anuncios públicos de van Ditmarsch.

Proposición 3.5. Las siguientes fórmulas son todas equivalentes:

- $\varphi \to [\varphi]\psi$
- $\varphi \to \langle\varphi\rangle\psi$

[5]Podría considerarse que, dado que la anuncia Ana, en realidad lo que se anuncia es $\neg K_b op \wedge K_a op$. En este caso, da igual, dado que, como el lector podrá comprobar, la restricción a ambas proposiciones del modelo inicial proporciona el mismo modelo.

- $[\varphi]\psi$

□

Proposición 3.6. Las siguientes fórmulas son todas equivalentes:

- $\langle\varphi\rangle\psi$
- $\varphi \wedge \langle\varphi\rangle\psi$
- $\varphi \wedge [\varphi]\psi$

□

Proposición 3.7. Las siguientes fórmulas son válidas:

- $\langle\varphi\rangle K_a\psi \leftrightarrow (\varphi \wedge K_a(\varphi \rightarrow \langle\varphi\rangle\psi))$
- $\langle\varphi\rangle \hat{K}_a\psi \leftrightarrow (\varphi \wedge \hat{K}_a\langle\varphi\rangle\psi)$

□

Concluimos esta sección con otro resultado que puede parecer bastante sorprendente (aunque ya lo adelantamos en la sección inicial de este capítulo). Lo que se deriva de este resultado en última instancia es, básicamente, que podemos reescribir cualquier fórmula en $\mathcal{L}_{K[]}$ como una fórmula equivalente de $\mathcal{L}_K$, con lo que, "estrictamente hablando", los anuncios públicos "no son realmente necesarios", al menos en $\mathcal{L}_K$. Por supuesto, todo aquel que quiera sugerirnos que podemos prescindir de los anuncios públicos a cambio de construcciones formales equivalentes es igualmente libre de escribir fórmulas proposicionales usando solo la barra de Sheffer ($\uparrow$)[6].

Teorema 3.1. ***Reescrituras de $\mathcal{L}_{K[]}$ a $\mathcal{L}_K$.*** *Se tienen las siguientes equivalencias:*

$$[\varphi]p \leftrightarrow \varphi \rightarrow p \tag{3.4a}$$
$$[\varphi]\,(\psi \wedge \chi) \leftrightarrow ([\varphi]\psi \wedge [\varphi]\chi) \tag{3.4b}$$
$$[\varphi]\,\neg\psi \leftrightarrow (\varphi \rightarrow \neg[\varphi]\psi) \tag{3.4c}$$
$$[\varphi]\,K_a\psi \leftrightarrow (\varphi \rightarrow K_a[\varphi]\psi) \tag{3.4d}$$
$$[\varphi]\,[\psi]\chi \leftrightarrow [\varphi \wedge [\varphi]\psi]\chi \tag{3.4e}$$

□

En efecto, es fácil intuir que con este conjunto de escrituras es posible tomar cualquier fórmula en $\mathcal{L}_{K[]}$ e ir reescribiéndola sucesivamente hasta obtener una fórmula equivalente en $\mathcal{L}_K$. Si el lector no está convencido de este hecho y desea una prueba formal del mismo, lo referimos a van Ditmarsch [1] (teorema 8.44).

[6] El operador NAND, que por sí solo forma un sistema funcionalmente completo de conectivas en la lógica proposicional; por ejemplo, la fórmula $P \wedge Q$ se escribiría $(P \uparrow Q) \uparrow (P \uparrow Q)$. Alguien todavía más purista podría expresar estas fórmulas sin hacer uso de paréntesis.

Observación 3.2. En particular, de lo anterior se sigue que el concepto de bisimulación que hemos dado para $\mathcal{L}_K$ es "coherente" con $\mathcal{L}_{K[]}$; es decir, si dos modelos puntuados M, s y M', s' son bisimilares, entonces se satisfacen las mismas fórmulas de $\mathcal{L}_{K[]}$ en ambos.

□

4. Lógica epistémica probabilística

4.1. Idea y motivación

Desde su concepción, el concepto de probabilidad ha tenido interpretaciones muy diversas, y tratar de hacer una enumeración exhaustiva deteniéndonos pormenorizadamente sobre cada una de ellas, si bien puede ser una tarea digna de su propio esfuerzo, no nos parece conveniente en el contexto de nuestro trabajo. No obstante, una de las divisiones más importantes que pueden establecerse entre todas estas interpretaciones, y que es de especial interés en cuestiones epistémicas, es la que distingue, por una parte, la probabilidad como una valoración interna del sujeto ("probabilidad como grado de certeza"), y, por otra, como un aspecto intrínseco a la realidad objetiva con entidad propia (podrían citarse en este sentido, por ejemplo, las teorías contemporáneas con las que tratarían de explicarse algunos de los fenómenos más extraños que se observan a escala subatómica). Tampoco entraremos en los aspectos más teóricamente complejos y filosóficamente sugerentes de las implicaciones de cada una de estas posibilidades (o incluso de una combinación de ambas), pero el lector podrá imaginar de sobra que cada una de estas requiere de un tratamiento particular y matizado.

Las herramientas conceptuales que introduciremos en este capítulo pueden prestarse tanto a unas interpretaciones como a las otras; no obstante, sí es cierto que, en un principio, quizá las que más naturalmente se ajustan a ellas son las del primer tipo ("probabilidad subjetiva"). Esto tiene que ver en gran parte con el contexto inmediato en el que se desarrolló este tipo de formalismos, a saber, en el análisis de sistemas distribuidos – los propios Ronald Fagin y Joseph Y. Halpern, cuyo artículo *Reasoning about Knowledge and Probability* [3] es nuestra principal referencia en este capítulo, pueden considerarse sin temor a exagerar eminencias en este campo de investigación[1]. En este ámbito, es importante que los formalismos utilizados permitan distinguir claramente entre, por una parte, un conjunto de "hechos" de naturaleza aleatoria o probabilística – por ejemplo, el output de algún programa pseudo-aleatorio – y, por otra parte, una noción de conocimiento de estos hechos o estimación de sus probabilidades por parte de los nodos en el sistema – podría ocurrir, como veremos en los ejemplos, que un mismo "nodo" pudiese considerar varias distribuciones de probabilidad posibles sobre un mismo conjunto de hechos.

Esto no impide, por supuesto, que se puedan representar situaciones en las que las nociones de probabilidad reflejen una propiedad "objetiva" del mundo externo: generalmente, esto podría interpretarse como la condición de que todos los agentes en el sistema están de acuerdo en la distribución de probabilidad que gobierna un conjunto de hechos bajo un determinado conjunto de circunstancias; de hecho, esta es una de las "propiedades interesantes que puede cumplir un modelo" que estudiaremos más adelante.

[1] Ambos han recibido el premio Gödel, otorgado por la European Association for Theoretical Computer Science (EATCS), Halpern en 1997 y Fagin en 2014, entre otros reconocimientos.

4.2. Sintaxis y semántica del lenguaje epistémico probabilístico

Definición 4.1. ***(Sintaxis del lenguaje $\mathcal{LP}_{\mathcal{K}}$).*** *Dado un conjunto finito de agentes A y un conjunto numerable de proposiciones atómicas At, definimos el lenguaje epistémico probabilístico $\mathcal{LP}_{\mathcal{K}}(A, At)$ (como de costumbre, generalmente obviaremos A y At en la notación anterior) con la siguiente BNF:*

$$\varphi ::= p \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid K_a\varphi \mid \sum_{k=1}^{n} Q_k Pr_a(\varphi_k) \geq Q \tag{4.1}$$

Con $p \in At$, $a \in A$, $k \in \{1, 2, 3, ...\}$ y los términos Q_k, $k = 1, ..., n$ y Q son racionales arbitrarios.

□

Esta nueva variante de nuestro lenguaje nos permite formalizar razonamientos sobre hechos probabilísticos, y, más concretamente, sobre sus probabilidades o combinaciones lineales de las mismas[2]. El operador n-ario $\sum_{k=1}^{n} Q_k Pr_a(\ ) \geq Q$ se llamará en general "operador probabilístico", y a menudo lo expresaremos con la más ligera notación $Q_1 Pr_a(\ ) + ... + Q_k Pr_a(\ ) \geq Q$. A los términos Q_k los llamaremos "coeficientes", y Q se denominará el "término libre" o "término independiente".

Antes de proseguir con una definición rigurosa de cómo se interpretaría una fórmula que hace uso de la nueva construcción, tratemos de imaginarlo de manera informal. Consideremos la siguiente situación más o menos abstracta: sean a y b dos agentes, y p un hecho que puede darse o no con cierta probabilidad P. Entonces:

- La fórmula $Pr_a(p) \geq Q$ se leería "el agente a asigna una probabilidad mayor o igual que Q al hecho p".
- La fórmula $K_a(Pr_a(p) \geq Q)$ se leería "el agente a sabe que el agente a asigna una probabilidad mayor o igual que Q al hecho p", o simplemente "el agente a sabe que la probabilidad del hecho p es mayor o igual que q".
- La fórmula $Pr_b(K_a(Pr_a(p) \geq Q)) \geq Q'$ se leería "el agente b asigna una probabilidad mayor o igual que Q' al hecho de que el agente a sepa que la probabilidad del hecho p es mayor o igual que Q".
- La fórmula $K_b(Pr_b K_a(Pr_a(p) \geq Q) \geq Q')$ se leería "el agente b sabe que el agente b asigna una probabilidad mayor o igual que Q' al hecho de que el agente a sepa que la probabilidad del hecho p es mayor o igual que Q", o simplemente "el agente

[2]La razón por la que nos limitamos a combinaciones lineales es de naturaleza técnica, y tiene que ver con la posibilidad de obtener una axiomática completa para el cálculo proposicional en este lenguaje – de hecho, la razón por la que nos limitamos a términos Q_k racionales y no, por ejemplo, reales, o la razón por la que no permitimos "mezclar" varios agentes en una misma instanciación del operador probabilístico (es decir, no se permiten fórmulas como $Pr_a(\varphi) + Pr_b(\psi) \geq Q$) son similares. Estas cuestiones se discutirán por encima en el capítulo 7.

b sabe que la probabilidad de que el agente a sepa que la probabilidad del hecho p sea mayor o igual que Q es mayor o igual que Q'".

Más allá de ver, una vez más, lo impracticable que sería hacer este tipo de razonamientos sin disponer de un formalismo adecuado, el lector podrá percibir también que existe una dualidad entre que un agente "asigne" una probabilidad a un hecho y que el agente "sepa" cuál es la probabilidad de este hecho, que en un principio podría resultar un tanto extraña. *Grosso modo*, el motivo de introducir esta dualidad es que, en determinadas situaciones, tiene sentido considerar que un agente asigne simultáneamente varias probabilidades a un mismo hecho (lo veremos más claramente en los ejemplos). Aunque esta distinción pueda parecer inicialmente confusa, si se comprende y se usa adecuadamente puede dotar al formalismo de una enorme versatilidad respecto a otros que podrían parecer "más naturales" a priori.

Dado que, según parece, este lenguaje introduce una nueva dimensión probabilística totalmente ajena a los conceptos "clásicos" de modalidad, resulta necesario definir una nueva familia de modelos para interpretarlo.

Definición 4.2. ***Modelos de Kripke probabilísticos.*** *Dado un conjunto numerable de átomos At y un conjunto finito de agentes A, un modelo de Kripke probabilístico es una estructura $M = \langle S, R^A, V^{At}, \Pi^{A,At} \rangle$, donde:*

- *La estructura $\langle S, R^A, V^{At} \rangle$ define un modelo de Kripke.*
- $\Pi^{A,At} : A \times S \to Probs(S)$*, donde $Probs(S)$ es la notación que usaremos para representar el conjunto de los espacios probabilísticos sobre subconjuntos de S. Más concretamente,*

$$\Pi^{A,At} : (a, s) \mapsto \langle S_{(a,s)}, \mathfrak{A}_{(a,s)}, Pr_{(a,s)} \rangle$$

donde:

  - $S_{(a,s)} \subseteq S$ *es el espacio muestral para a en s.*
  - $\mathfrak{A}_{(a,s)}$ *es un σ-álgebra sobre $S_{(a,s)}$.*
  - $Pr_{(a,s)}$ *es una medida de probabilidad sobre $\mathfrak{A}_{(a,s)}$ $(Pr_{(a,s)} : \mathfrak{A}_{(a,s)} \to [0, 1])$.*

Como recordatorio, en cualidad de medida de probabilidad, $Pr_{(a,s)}$ debe satisfacer las siguientes propiedades [5] [3]:

- $Pr_{(a,s)}(S_{(a,s)}) = 1$
- $Pr_{(a,s)}(E) \geq 0 \quad \forall E \in \mathfrak{A}_{(a,s)}$
- ***(Aditividad numerable)*** *Sean $\{E_k\}_{k=0}^{\infty} \subseteq \mathfrak{A}_{(a,s)}$ una familia numerable de conjuntos* ***disjuntos*** *en el σ-álgebra. En tal caso, se tiene:*

[3] También se ha propuesto, en este mismo campo, el uso de nociones no estándar de probabilidad para resolver ciertas limitaciones de la noción "clásica"; el propio Halpern hace propuestas en esta dirección en su artículo *Lexicographic probability, conditional probability, and nonstandard probability* [12].

$$Pr_{(a,s)}(\bigcup_{k=0}^{\infty} E_k) = \sum_{k=0}^{\infty} Pr_{(a,s)}(E_k)$$

Como de costumbre, evitaremos hacer referencia explícita a A y At a menos que las ambigüedades del contexto lo ameriten.

□

Definición 4.3. ***Semántica de $\mathcal{LP}_{\mathcal{K}}$.*** *Dado un modelo de Kripke probabilístico $M = \langle S, R, V, \Pi \rangle$ con agentes A y átomos At,*

$$\begin{array}{lll} M, s \vDash p & sii & s \in V(p),\ p \in At \\ M, s \vDash \neg\varphi & sii & M, s \nvDash \varphi \\ M, s \vDash \varphi \wedge \psi & sii & M, s \vDash \varphi \text{ y } M, s \vDash \psi \\ M, s \vDash K_a\varphi & sii & \text{Para todo } t \in S:\ R_a st \Rightarrow M, t \vDash \varphi \\ M, s \vDash \sum_{(k=0)}^{n} Q_k Pr_a(\varphi_k) \geq Q & sii & \sum_{k=0}^{n} Q_k Pr_{*(a,s)}(S_{(a,s)}(\varphi_k)) \geq Q \end{array} \tag{4.2}$$

Donde:

- *El conjunto $S_{(a,s)}(\varphi)$ se define como*

$$S_{(a,s)}(\varphi) := \{s \in S_{(a,s)} \mid M, s \vDash \varphi\}$$

- *El asterisco ($_*$) en la función de probabilidad denota que estamos utilizando la medida interior, es decir,*

$$Pr_{*(a,s)}(G) := sup\{Pr_{(a,s)}(E) \mid E \in \mathfrak{A}_{(a,s)}, E \subseteq G\}$$

(La mayor de las medidas de subconjuntos de S en el σ-álgebra.)

□

Nota 4.1. Abusando de la notación, escribiremos $Pr_{*(a,s)}(\varphi_k)$ en lugar de $Pr_{*(a,s)}(S_{(a,s)}(\varphi_k))$

□

Observación 4.1. La razón por la que tenemos que recurrir a utilizar la medida interior para la semántica del operador probabilístico es que, en principio, no siempre podemos asegurar que todos los conjuntos de la forma $S_{(a,s)}(\varphi)$ sean medibles (que estén en el σ-álgebra). También podría haberse hecho uso del concepto análogo de "medida exterior" (la menor de las medidas de subconjuntos de S en el σ-álgebra), pero quizá resulte más natural hacerlo de esta manera, dado que de la otra pueden darse algunas situaciones un tanto extrañas (por ejemplo, podemos tener dos conjuntos disjuntos con medida exterior 1, lo cual traducido a los términos de nuestra lógica implicaría

que podríamos tener fórmulas mutuamente contradictorias de forma que todas ellas evaluasen a probabilidad 1 para una misma asignación).

Más adelante veremos ciertas condiciones suficientes para que todos los conjuntos considerados sean medibles. La principal ventaja de esto es que nos permitirá definir una axiomática completa más sencilla y natural, aunque, como también veremos, el caso "no medible" también admite una axiomática completa.

□

Como ya hemos hecho en capítulos anteriores, tratemos de dar cierto sentido informal a los formalismos que acabamos de definir. Básicamente, la situación que tendríamos es la siguiente: además de las relaciones de accesibilidad que ya teníamos como parte de la semántica de Kripke, que configurarían la parte modal de nuestro lenguaje y nos permitirían razonar en términos de "conocer" y "considerar posible", añadimos ahora también que, para cada agente a en cada estado s, se considera una distribución de probabilidad sobre algún subconjunto (no vacío) de S, sobre la cual se interpretaría la parte probabilística del lenguaje. Obsérvese que los conjuntos $S_{(a,s)}$ (espacio muestral de a en s) y $R_a(s) := \{t \in S \mid R_a st\}$ (conjunto de estados accesibles para a desde s) no tienen por qué coincidir; de hecho, ni siquiera tiene por qué ocurrir que el primero sea un subconjunto del segundo. Esto tiene varias "consecuencias imprevistas de interpretación no trivial", como que un agente pueda considerar posibles (utilizando la noción modal de posibilidad) varios estados del mundo asignando probabilidades solo a algunos de ellos; que considere posible un estado al que asigna una probabilidad nula; o, incluso, que asigne probabilidades positivas a estados que ni siquiera considera posibles. Todo esto, no obstante, debe verse como una ventaja por parte del formalismo, y no como una limitación, dado que significa que el conjunto de situaciones que podemos representar es mucho más general de lo que podríamos haber pensado en un principio; y, si queremos estudiar situaciones más concretas, basta con añadir restricciones adicionales a la versión más general del formalismo.

De hecho, estas consideraciones podrían dar pie a que nos preguntásemos: ¿son los modelos epistémicos (considerándolos sin la parte probabilística) los más apropiados para plantear este tipo de situaciones? (¿O bajo qué circunstancias lo son?) De hecho, en el próximo capítulo veremos que hay autores que consideran conveniente trabajar en una clase *todavía más general* de modelos (a saber, la clase más general posible: el conjunto $\mathcal{K}$ de todos los modelos de Kripke). En todo caso, aunque los conceptos y resultados más generales que presentamos en este trabajo no se limitan en su aplicación y relevancia a los modelos en $\mathcal{S}5$, en aras de la concreción trataremos de enmarcarlos siempre en el contexto de alguna situación epistémica.

Nota 4.2. El lector habrá observado que, en principio, el lenguaje que hemos definido solo incluye el constructor para fórmulas de la forma $Q_1 Pr_a(\ ) + ... + Q_k Pr_a(\ ) \geq Q$, y no para otras desigualdades e igualdades que podría resultar natural querer expresar. Por suerte, la mayoría de estas puede expresarse a partir de la desigualdad que ya tenemos. Escribimos a continuación algunos ejemplos:

- $Q_1 Pr_a(\varphi_1) + ... + Q_k Pr_a(\varphi_k) < Q \iff \neg(Q_1 Pr_a(\varphi_1) + ... + Q_k Pr_a(\varphi_k) \geq Q)$
- $Q_1 Pr_a(\varphi_1) + ... + Q_k Pr_a(\varphi_k) \leq Q \iff (-Q_1) Pr_a(\varphi_1) + ... + (-Q_k) Pr_a(\varphi_k) \geq -Q$

- $Q_1 Pr_a(\varphi_1) + ... + Q_k Pr_a(\varphi_k) = Q \Longleftrightarrow ((-Q_1)Pr_a(\varphi_1) + ... + (-Q_k)Pr_a(\varphi_k) \geq -Q)$
  $\wedge (Q_1 Pr_a(\varphi_1) + ... + Q_k Pr_a(\varphi_k) \geq Q)$
- $Q_1 Pr_a(\varphi_1) \geq Q_2 Pr_a(\varphi_2) \Longleftrightarrow Q_1 Pr_a(\varphi_1) - Q_2 Pr_a(\varphi_2) \geq 0$

Otra notación de la que haremos uso habitualmente será $KP_a^{\geq Q}\varphi$, que será una abreviatura de la expresión $K_a(P_a(\varphi) \geq Q)$. Esta abreviatura nos serviría para expresar de manera concisa la afirmación "el agente a sabe que la probabilidad de φ es mayor o igual que Q", dado que, como vimos en nuestros ejemplos iniciales, debe distinguirse en nuestra semántica entre la *asignación* de una probabilidad y el *conocimiento* de dicha asignación. Finalmente, de manera análoga podríamos definir otras notaciones como $KP_a^{=Q}$, $KP_a^{<Q}$, etcétera.

□

4.3. Propiedades y resultados básicos de la lógica epistémica probabilística

Introducimos esta sección con un ejemplo de Halpern y Fagin [3] que nos permite observar la gran pluralidad de matices que es posible expresar con el lenguaje $\mathcal{LP}_{\mathcal{K}}$, y, por otra parte, la importancia de tener en cuenta estos matices a la hora de modelar una situación.

Ejemplo 4.1. **Incertidumbre probabilística e incertidumbre esencial.** Consideremos la siguiente situación: un sistema distribuido está constituido por dos nodos a y b. El nodo b admite un bit de input con valor 0 o 1. Una vez introducido el input, el mismo nodo simula el lanzamiento de una moneda justa, y realiza una acción *Act* si el input coincide con el resultado del lanzamiento (es decir, si el input es 1 y en el resultado del lanzamiento es cara, o el input es 0 y en el resultado del lanzamiento es cruz). El nodo a desconoce el valor del input introducido en b.

Dado que el valor del input es conocido para b, es fácil razonar que este nodo sabe que la probabilidad de realizar la acción *Act* (antes de haberse simulado el lanzamiento de la moneda) es de 1/2. Se podría razonar de forma similar para el nodo a, que no conoce el valor del input: por una parte, si el input de b es 0, entonces realizará la acción en el caso en que el resultado del lanzamiento sea cruz, hecho cuya probabilidad es 1/2; análogamente si el input de b es 1, entonces la acción se realizará si el resultado del lanzamiento es cara (probabilidad 1/2). Por lo tanto, independientemente del input de b, la probabilidad de que se realice *Act* es 1/2. Obsérvese que este razonamiento no requiere de tener que especificar una distribución de probabilidad sobre el valor del input introducido; en todo caso, si tal distribución de probabilidad existiese, esto no afectaría a nuestro argumento.

Tratemos ahora de formalizar el argumento anterior en nuestro sistema. Tenemos cuatro estados posibles del mundo: $(0, X)$, $(0, C)$, $(1, X)$, $(1, C)$ (la primera componente representa el valor del input de b, y la segunda representa el resultado del lanzamiento de la moneda). Llamémoslos respectivamente s_1, s_2, s_3 y s_4, y sea S el conjunto de estos cuatro estados. Consideremos las siguientes proposiciones atómicas:

- Con Act denotamos el hecho de que se lleva a cabo Act
- Con C, X denotamos cada uno de los resultados del lanzamiento (cara y cruz, respectivamente).
- Con B_0, B_1 denotamos cada uno de los posibles valores del input de b.

Los nombres de los estados son autodescriptivos en cuanto a si se tiene o no en cada caso C, X, B_0 o B_1. Los estados en los que se tiene Act son $(0, X)$ y $(1, C)$. Las relaciones de accesibilidad también están claras: para el nodo a, todos los estados son accesibles desde cualquier otro; para el nodo b, un estado es accesible desde otro si y solo si sus primeras componentes (es decir, el valor del input) coinciden. Nos queda una cuestión por resolver: la asignación de probabilidades. A continuación presentamos varias opciones para el nodo a (también habría que definir la asignación de probabilidades del nodo b, pero en este ejemplo nos bastará con la del nodo a para ilustrar lo que nos interesa), todas ellas en principio razonables:

I) **Modelo M0.** Podemos asociar a cada estado s el espacio muestral S formado por todos los estados posibles. En principio esta podría parecer la opción más natural, dado que estamos considerando que el espacio muestral coincide en cada estado s con el conjunto de los estados accesibles (para a) desde s. Dado que asumimos que los eventos de introducir un input u otro no son probabilísticos, los únicos conjuntos medibles que podemos considerar, además del vacío y el total, son $\{(0, C), (1, C)\}$ y $\{(0, X), (1, X)\}$ (es decir, $\{s_2, s_4\}$ y $\{s_1, s_3\}$), cada uno de ellos con probabilidad $1/2$.

II) **Modelo M1.** Otra posibilidad es asociar a cada estado s el espacio muestral formado por aquellos estados cuyo input coincida con el del propio estado s. A los estados $(0, X)$ y $(0, C)$ se les asociaría el espacio muestral $\{(0, X), (0, C)\}$, y los conjuntos medibles no triviales serían los unitarios con probabilidad $1/2$ en cada caso; para los estados $(1, X)$ y $(1, C)$ tendríamos lo análogo.

III) **Modelo M2.** Por último, tenemos la opción trivial de asignar a cada estado s el espacio muestral formado por el propio estado s, con probabilidad 1 en cada caso.

De estas tres opciones, la única en la que el razonamiento informal que hemos expresado más arriba es correcto es en el modelo **M1**. En efecto, no es difícil comprobar que en cualquiera de los estados de este modelo se tiene $KP_a^{=1/2}Act$. Como podemos observar en la figura 4.2, en cualquiera de los estados s el nodo considera posibles dos asignaciones de probabilidad, y en ambas se tiene que el conjunto $S_{(a,s)}(Act)$ (que en unos casos es el conjunto unitario $\{(0, X)\}$ y en otros es el conjunto unitario $\{(1, C)\}$) tiene una probabilidad de exactamente $1/2$.

No es este el caso en el modelo **M0** (figura 4.1), donde para todo estado s se tiene una misma asignación de probabilidad, y en dicha asignación el conjunto $S_{(a,s)}(Act) := \{(0, X), (1, C)\}$ no es medible, y su medida interior es nula. Finalmente, en el modelo **M2** (figura 4.3) tenemos una situación un tanto curiosa: en cualquier estado s, es posible o bien que $P_a(Act) = 0$ o bien que $P_a(Act) = 1$; esto es, que en todos los estados se tiene $M_2, s \vDash K_a(P_a(Act) = 1 \vee P_a(Act) = 0)$.

Un comentario final sobre este ejemplo es que las consideraciones anteriores no

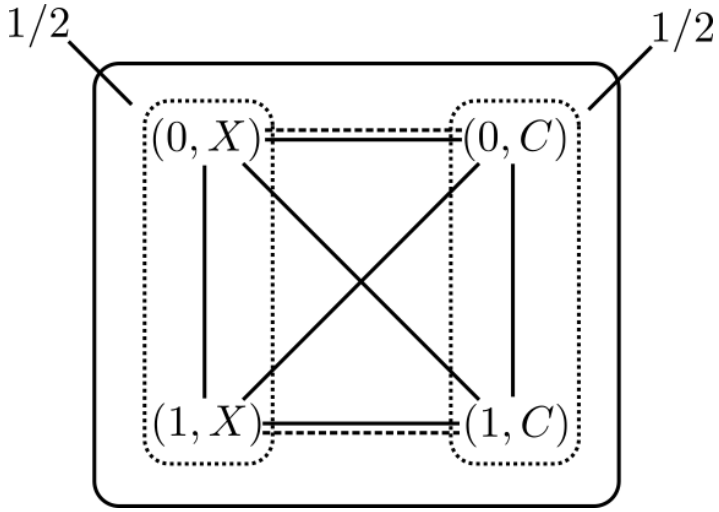

Figura 4.1: Representación del modelo **M0**. Las "cajas punteadas" representan los subconjuntos medibles de cada espacio muestral. En este caso, el único espacio muestral está representado con la caja de líneas continuas. Las rectas continuas representan la relación de accesibilidad del nodo a, y las rectas discontinuas representan la relación de accesibilidad del nodo b.

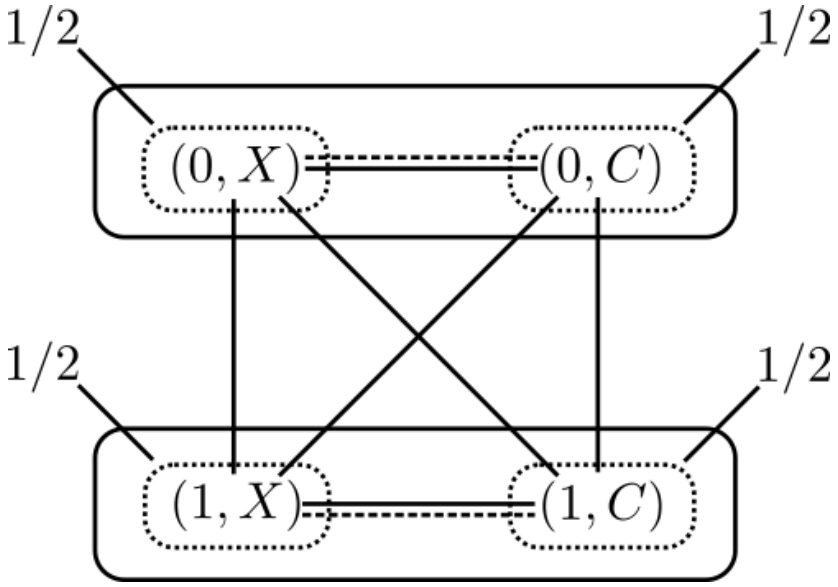

Figura 4.2: Representación del modelo **M1**.

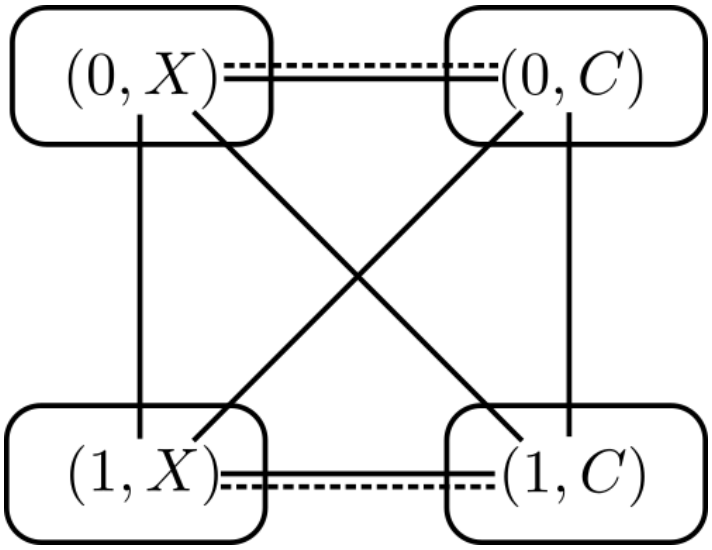

Figura 4.3: Representación del modelo **M2**. En este caso, cada estado se identifica con un espacio muestral.

deben dar pie a la interpretación de que el modelo $M1$ es "el correcto" en esta situación. Una interpretación más completa podría obtenerse considerando que los modelos $M2$ y $M1$ representan dos instantes sucesivos en la evolución de un sistema (una consecuencia un tanto confusa de la forma de la que hemos decidido nombrar a nuestros modelos es que el instante correspondiente al modelo $M2$ sería cronológicamente anterior al que corresponde al modelo $M1$). En el primer instante (modelo $M2$), el nodo b ya ha obtenido el input pero todavía no ha realizado el lanzamiento de la moneda; por lo tanto, cada uno de los estados correspondientes se encuentra todavía "indeterminado", y se les pueden asignar distribuciones de probabilidad. En el segundo instante (modelo $M1$), el lanzamiento ya se ha realizado, y puede considerarse que no tiene ya sentido asignar a cada uno de los estados una probabilidad de $1/2$, aunque, obviamente, el nodo a no conoce el resultado (de ahí que en todos los estados se tenga $M_2, s \vDash K_a(P_a(Act) = 1 \vee P_a(Act) = 0)$). Una forma de distinguir entre ambos tipos de incertidumbre es asignando nombres distintos a cada una de ellas: el primer tipo de incertidumbre (el nodo considera que existe una distribución de probabilidad que gobierna un determinado subconjunto de estados) puede llamarse "incertidumbre probabilística", mientras que el segundo tipo (el nodo no considera que exista ninguna distribución de probabilidad gobernando un determinado subconjunto de estados) puede llamarse "incertidumbre esencial"; obviamente, en muchas ocasiones puede transformarse una situación "esencialmente incierta" en otra "probabilísticamente incierta", simplemente añadiendo una distribución de probabilidad que se adecúe a las relaciones de accesibilidad de la situación inicial, pero no está claro que esto pueda hacerse siempre de manera satisfactoria[4].

[4]Un ejemplo sencillo sería un modelo cuyos estados fueran los números naturales, y cuya relación de accesibilidad fuese la total (cualquier estado es accesible desde cualquier otro). Sabemos que no existen medidas de probabilidad sobre los números naturales tales que a cada conjunto unitario se le asigne una medida no nula de forma equiprobable; por lo tanto, tendríamos que prescindir de esta probabilidad y "privilegiar" a unos estados sobre otros, lo cual podría no ser deseable dependiendo de la situación que estuviesemos tratando.

□

El ejemplo anterior nos muestra cómo una situación que en principio puede parecer sencilla ya se presta a bastantes consideraciones en cuanto a cómo se deben asignar los espacios probabilísticos. En particular, el ejemplo nos permite ver que hay muchas situaciones en las que no es deseable tomar $S_{(a,s)} = R_a(s)$ (de hecho, el único de los modelos anteriores en el que se satisfacía esta propiedad, **M0**, es el único para el que no hemos podido encontrar una interpretación adecuada en el contexto de nuestro ejemplo, a pesar de que, inicialmente, lo planteamos como el candidato más "natural"). No obstante, podemos observar que una condición que sí satisface cada uno de los modelos anteriores es que $S_{(a,s)} \subseteq R_a(s)$. Esta es una propiedad que en la mayoría de los contextos podríamos considerar "bastante natural", dado que, en ausencia de esta, un agente podría asignar probabilidades positivas sobre hechos que no considerase posibles. Podría decirse que un agente que asigna probabilidades positivas a hechos que considera imposibles es *inconsistente*; por lo tanto, bautizaremos esta propiedad como **CONS** (de *consistencia*).

Definición 4.4. ***CONS.*** *Decimos que un modelo de Kripke probabilístico satisface la propiedad* ***CONS*** *si para todo $a \in A$ y $s \in S$, se tiene $S_{(a,s)} \subseteq R_a(s)$.*

□

Obsérvese que esta propiedad no implica $s \in S_{(a,s)}$, es decir, un agente no tiene por qué considerar que el estado en el que se encuentra forme parte del espacio muestral en su asignación de probabilidad. Si bien esto puede parecer también un tanto extraño, hay situaciones en las que puede resultar apropiado. Para una discusión más pormenorizada de este tipo de situaciones, referimos al lector al artículo de Halpern y Fagin [3] y a las fuentes que estos citan en el mismo en relación con esta cuestión.

Otra propiedad que podríamos considerar deseable en ciertas situaciones es que las distribuciones de probabilidad sean "objetivas", es decir, que sean compartidas por todos los agentes; esto es apropiado en muchas de las situaciones típicas que se estudian en teoría de juegos, donde los "jugadores" deben conocer comúnmente un conjunto de hechos para que el "juego" tenga sentido. Regresando al ejemplo anterior, un motivo posible para considerar que el modelo **M2** es "más adecuado" que el modelo **M1** para representar la situación tras el lanzamiento de la moneda sería precisamente este: si, además de las asignaciones de probabilidad del nodo a, consideramos también las del b, no tendría ningún sentido que este siguiese asignando probabilidad $1/2$ al evento "se ha realizado la acción *Act*"; por lo tanto, si queremos hacer coincidir las asignaciones de probabilidad de ambos nodos, en principio parecería que este es el modelo que deberíamos elegir para representar la situación.

Definición 4.5. ***OBJ.*** *Decimos que un modelo de Kripke probabilístico satisface la propiedad* ***OBJ*** *si se tiene $\Pi_{(a,s)} = \Pi_{(b,s)}$ para todo $s \in S$ y $a, b \in A$.*

□

Otras propiedades que pueden ser naturales en ciertos contextos tienen que ver con la relación entre estados y asignaciones de probabilidad. Por ejemplo, podría interesarnos imponer que la asignación de probabilidad de un agente a sea la misma en

todos los estados que el agente considera posibles desde un estado dado. Llamemos **SDP** (*Space Determined Probability*) a esta propiedad.

Definición 4.6. ***SDP.*** *Decimos que un modelo de Kripke probabilístico satisface la propiedad* ***SDP*** *si para todo $a \in A$ y $s, t \in S$, si $R_a st$ entonces $\Pi_{(a,s)} = \Pi_{(a,t)}$.*

□

De los tres modelos de nuestro ejemplo anterior, solo **M0** satisface **SDP**. No obstante, **M1** y **M2** satisfacen una llamada *uniformidad* que, como veremos ahora, puede considerarse en cierto sentido como una relajación de **SDP**.

Definición 4.7. ***UNIF.*** *Decimos que un modelo de Kripke probabilístico satisface* ***UNIF*** *si para todo $a \in A$, $s, t \in S$, si $t \in S_{(a,s)}$ entonces $\Pi_{(a,s)} = \Pi_{(a,t)}$.*

□

Una forma de interpretar esta propiedad es que, si bien la asignación de probabilidades para un agente a no es la misma en todos los estados que considera posible, al menos es posible particionar $R_a(s)$ de forma que, para cada conjunto T en la partición, todos los estados s en T tengan una misma asignación de probabilidades de tal forma que el espacio muestral $S_{(a,s)}$ asociado a dicha asignación es el propio conjunto T. No obstante, esta interpretación solo puede hacerse si, además, también se satisface **CONS** (de lo contrario, no podemos asegurar que $S_{(a,s)} \subseteq R_a(s)$).

Proposición 4.1. Sea M un modelo de Kripke probabilístico satisfaciendo **CONS** y **SDP**; entonces, M satisface **UNIF**.

Demostración. Sean $a \in A$ y $s, t \in S$ arbitrarios tales que $t \in S_{(a,s)}$. Por **CONS**, $S_{(a,s)} \subseteq R_a(s)$, y, por lo tanto, $t \in R_a(s)$; por **SDP**, esto implica que $\Pi_{(a,s)} = \Pi_{(a,t)}$, y, en particular, $S_{(a,s)} = S_{(a,t)}$.

Q.E.D.

□

Una última propiedad, de carácter más técnico, que podríamos pedir a un modelo es que todos sus conjuntos sean medibles; esta propiedad, como ya mencionamos en la observación 4.1, simplifica de diversas formas los razonamientos probabilísticos.

Definición 4.8. ***MEAS.*** *Decimos que un modelo de Kripke probabilístico satisface* ***MEAS*** *si para todo $a \in A$, $s \in S$ y $\varphi \in \mathcal{LP}_{\mathcal{K}}$, el conjunto $S_{(a,s)}(\varphi)$ es medible ($S_{(a,s)}(\varphi) \in \mathfrak{A}_{(a,s)}$)*

□

Definición 4.9. ***PMEAS.*** *Decimos que un modelo de Kripke probabilístico satisface* ***PMEAS*** *si para todo $a \in A$, $s \in S$ y $p \in At$, el conjunto $S_{(a,s)}(p)$ es medible.*

□

No es difícil comprobar que si un modelo satisface **PMEAS**, entonces todas las fórmulas puramente proposicionales definen conjuntos medibles. No obstante, no está claro que esto sea así para proposiciones con componente modal o probabilística. A continuación proporcionamos una condición suficiente para **MEAS**.

Proposición 4.2. Sea M un modelo satisfaciendo **PMEAS**, **CONS**, **OBJ** y **UNIF**. Entonces M satisface **MEAS**.

Demostración. Procedemos por inducción estructural sobre las fórmulas de $\mathcal{LP}_{\mathcal{K}}$. Queremos probar que $S_{(a,s)}(\varphi)$ es medible, para todos $a \in A$, $s \in S$ y $\varphi \in \mathcal{LP}_{\mathcal{K}}$. El caso base ($p \in At$) se tiene por hipótesis, y los casos para los operadores proposicionales se comprueban fácilmente.

Operador modal. Sea φ tal que $S_{(a,s)}(\varphi)$ es medible para todo $a \in A$ y $s \in S$. Las propiedades **CONS** y **OBJ** juntas implican que para cualquier par de agentes a, b, $S_{(a,s)} = S_{(b,s)} \subseteq R_b(s)$. De aquí se sigue que $S_{(a,s)}(K_b(\varphi))$ o bien es $S_{(a,s)}$ o bien es el vacío (si $K_b\varphi$ se tiene en algún estado de $S_{(a,s)} = S_{(b,s)} \subseteq R_b(s)$, entonces por definición de $R_b(s)$ se tiene en todos); en ambos casos es medible.

Operador probabilístico. Sean $\varphi_1, ..., \varphi_k$ tales que en todos los casos el conjunto $S_{(a,s)}(\varphi_i)$ es medible, para todo $a \in A$ y $s \in S$. Dado $b \in A$ arbitrario, tenemos que probar que la fórmula $\psi := Q_1 Pr_b(\varphi_1) + ... + Q_k Pr_b(\varphi_k) \geq Q$ define un conjunto medible $S_{(a,s)}(\psi)$ para $a \in A$ y $s \in S$ arbitrarios. Semánticamente, $M, t \vDash \psi \iff Q_1 Pr_{*(b,t)}(\varphi_1) + ... + Q_k Pr_{*(b,t)}(\varphi_k) \geq Q$. El lado izquierdo de la desigualdad anterior es un valor numérico que depende de la distribución de probabilidad para el agente b en el estado t; denominémoslo $P_{(b,t)}$. Entonces $t \in S_{(a,s)}(\psi)$ si y solo si $t \in S_{(a,s)}$ y $P_{(b,t)} \geq Q$. Ahora bien, por **UNIF** tenemos que si $t \in S_{(a,s)}$, entonces $\Pi_{(a,s)} = \Pi_{(a,t)}$; en particular, esto implica que a todos los estados $t \in S_{(a,s)}$ les corresponde la misma asignación de probabilidades. Por lo tanto, $P_{(b,t)} \geq Q$ se tiene en un estado t si y solo si se tiene en todos, o lo que es lo mismo, o bien $S_{(a,s)}(\varphi) = S_{(a,s)}$ o bien $S_{(a,s)}(\varphi) = \emptyset$.

Q.E.D.

□

Antes de pasar a otro orden de cosas, presentamos una última propiedad de interpretación interesante que no se satisface en general, pero sí bajo ciertas condiciones. Dicha propiedad suele conocerse como "principio de Miller", y representa una limitación que se podría considerar natural sobre qué fórmulas con probabilidades de orden superior (probabilidades sobre probabilidades) son ciertas. En particular, el principio de Miller puede expresarse a través de la siguiente desigualdad:

$$Pr_a(\varphi) \geq b \, Pr_a(Pr_a(\varphi) \geq b) \tag{4.3}$$

Proposición 4.3. **UNIF y principio de Miller.** Sea M un modelo satisfaciendo **UNIF**. Entonces M satisface el principio de Miller.

Demostración. Sean $s \in S$, $a \in A$, $\varphi \in \mathcal{LP}_{\mathcal{K}}$ y $b \in \mathbb{Q}$ arbitrarios. Queremos probar:

$$M, s \vDash Pr_a(\varphi) \geq b \, Pr_a(Pr_a(\varphi) \geq b)$$

O lo que es lo mismo,

$$Pr_{(a,s)}(\varphi) \geq b \, Pr_{(a,s)}(Pr_a(\varphi) \geq b) \tag{4.4}$$

(Recordemos que las expresiones de la forma $Pr_{(a,s)}(\varphi)$ son un abuso de notación)

Tratemos de caracterizar el conjunto $S_{(a,s)}(Pr_a(\varphi) \geq b)$. Este es el conjunto de los estados $t \in S_{(a,s)}$ tales que $M, t \vDash Pr_a(\varphi) \geq b$. Ahora bien, por **UNIF**, $t \in S_{(a,s)} \Longrightarrow \Pi_{(a,s)} = \Pi_{(a,t)}$, por lo que si $Pr_a(\varphi) \geq b$ se tiene en algún estado $t \in S_{(a,s)}$, entonces se tiene en todos. Es decir:

$$Pr_{(a,s)}(\varphi) \geq b \Longrightarrow Pr_{(a,s)}(Pr_a(\varphi) \geq b) = 1$$

$$Pr_{(a,s)}(\varphi) < b \Longrightarrow Pr_{(a,s)}(Pr_a(\varphi) \geq b) = 0$$

En ambos casos, la desigualdad 4.4 se sigue de forma inmediata.

Q.E.D.

□

Concluimos esta sección, y este capítulo, con la generalización de una noción importante que presentamos en el capítulo 2, a saber, el concepto de bisimulación. Se trata en este caso de una generalización del concepto presentado en el artículo de Kooi [4].

Definición 4.10. ***Bisimulación en $\mathcal{LP}_{\mathcal{K}}$.*** *Sean M y M' dos modelos de Kripke probabilísticos, cuyos componentes se notarán respectivamente sin y con apóstrofe (por ejemplo: S y S' serán los respectivos conjuntos de estados). Una relación $\mathfrak{R} \subseteq S \times S'$ (no vacía) se dirá bisimulación si, dado $(s, s') \in S \times S'$ y $a \in A$, $\mathfrak{R}ss'$ implica:*

- ***Átomos, Ida y Vuelta.*** *Mismas propiedades que en 2.7.*
- ***P-Ida.*** *Dado un conjunto $E \subseteq S_{(a,s)}$, existe $E' \subseteq S'_{(a,s')}$ tal que*

$$Pr_{*(a,s)}(E) \leq Pr'_{*(a,s')}(E')$$

y

$$\forall t' \in E', \exists t \in E : (t, t') \in \mathfrak{R}$$

- ***P-Vuelta.*** *Dado un conjunto $E' \subseteq S'_{(a,s')}$, existe $E \subseteq S_{(a,s)}$ tal que*

$$Pr'_{*(a,s')}(E') \leq Pr_{*(a,s)}(E)$$

y

$$\forall t \in E, \exists t' \in E' : (t, t') \in \mathfrak{R}$$

Al igual que en la versión no probabilística, indicamos que existe una bisimulación entre los estados M, s y M', s' con la notación $(M, s) \underline{\longleftrightarrow} (M', s')$.

□

Comprobemos que el concepto anterior de bisimulación es adecuado, en el sentido de dos estados bisimilares satisfacen fórmulas equivalentes.

Proposición 4.4. **Equivalencia de modelos bisimilares en $\mathcal{LP}_{\mathcal{K}}$.** Sean $(M, s) \underline{\longleftrightarrow} (M', s')$. Entonces, para cualquier $\varphi \in \mathcal{LP}_{\mathcal{K}}$, se tiene $M, s \vDash \varphi$ si y solo si $M', s' \vDash \varphi$ (es decir, $M, s \equiv_{\mathcal{LP}_{\mathcal{K}}} M', s'$).

Demostración. Procedemos por inducción estructural sobre las fórmulas de $\mathcal{LP}_{\mathcal{K}}$. El caso base y los casos no probabilísticos se demuestran como en 2.1. Solo tenemos que demostrar, pues, el paso de inducción para el operador probabilístico.

Sean $\varphi_1, ..., \varphi_k$ satisfaciendo que, si M, t y M', t' son bisimilares, entonces $M, t \vDash \varphi_i$ si y solo si $M', t' \vDash \varphi$, $\forall i \in \{1, ..., k\}$. Queremos probar que $M, s \vDash Q_1 Pr_a(\varphi_1) + ... + Q_k Pr_a(\varphi_k) \geq Q$ si y solo si $M', s' \vDash Q_1 Pr_a(\varphi_1) + ... + Q_k Pr_a(\varphi_k) \geq Q$. Para aliviar la notación en esta prueba, denotaremos $E_i := S_{(a,s)}(\varphi_i)$ y $E'_i := S'_{(a,s)}(\varphi_i)$. En particular, si demostramos que $Pr_{*(a,s)}(E_i) = Pr'_{*(a,s')}(E'_i)$ para todo i, es fácil ver que se sigue el resultado. Establecemos dos subobjetivos:

I) $Pr_{*(a,s)}(E_i) \geq Pr'_{*(a,s')}(E'_i)$

II) $Pr_{*(a,s)}(E_i) \leq Pr'_{*(a,s')}(E'_i)$

Demostramos solo el primero, que hace uso de la cláusula **P-Ida**; el segundo es análogo pero con la cláusula **P-Vuelta**.

$E_i \subseteq S_{(a,s)}$. Por bisimilitud de M, s y M', s', y por la cláusula **P-Ida**, sabemos que existe $G'_i \subseteq S'_{(a,s')}$ satisfaciendo:

$$Pr_{*(a,s)}(E_i) \leq Pr'_{*(a,s')}(G'_i)$$

y

$$\forall t' \in G'_i, \exists t \in G_i : (t, t') \in \mathfrak{R}$$

Para probar nuestro subobjetivo nos bastaría ver que $G'_i \subseteq E'_i$, es decir, $\forall t' \in G'_i, M', t' \vDash \varphi_i$. Ahora bien, dado que nuestra φ_i satisface la inducción, y que existe $t \in E_i$ tal que los estados M, t y M', t' son bisimilares (por la segunda parte de la cláusula anterior), esto equivale a ver que $M, t \vDash \varphi_i$. Pero $M, t \vDash \varphi_i$ porque $t \in E_i = S_{(a,s)}(\varphi_i)$.

Q.E.D.

□

5. Lógica epistémica probabilística con anuncios públicos

5.1. Idea y motivación

Las principales motivaciones filosóficas para introducir tanto elementos dinámicos (anuncios públicos) como probabilísticos en nuestro lenguaje ya se han cubierto por separado en los capítulos anteriores, y no es difícil concebir cómo una combinación de ambos proporciona un sistema todavía más rico, tanto teóricamente como de cara al objetivo práctico de "modelar situaciones de la realidad". Una posible forma de ampliar nuestra apreciación de estas motivaciones sería planteando ejemplos más concretos de aplicaciones del formalismo. En este sentido, las herramientas de la lógica epistémica dinámica se prestan de manera especialmente conveniente, como veremos en algunas ilustraciones básicas de este capítulo y el siguiente, al desarrollo y análisis de protocolos de comunicación en sistemas distribuidos; un subconjunto interesante de estos cae bajo la sugerente etiqueta de "pruebas de cero conocimiento" (*zero-knowledge proofs*). La idea informal de este tipo de protocolos es, básicamente, que permitirían a un agente D (demostrador) probar a un agente V (verificador) que conoce un valor satisfaciendo una determinada propiedad sin tener revelar a V este valor o información adicional sobre el mismo (de la que no dispusiera ya V de antemano)[1]; un estudio más en profundidad de esta idea puede encontrarse en el artículo *Unifying Zero-Knowledge Proofs of Knowledge* de Maurer [13]. El lector familiarizado, por ejemplo, con los principales protocolos criptográficos de clave pública podrá reconocer en lo que acabamos de describir la idea básica que hay tras cualquiera de estos; asímismo, reconocerá también que existe cierto trasfondo probabilístico[2] que permite justificar teóricamente la *solidez* de estos protocolos.

Por otra parte, debido a la heterogeneidad de la lógica modal (y sus diversas vertientes e hibridaciones con otros tipos de lógicas) como campo de investigación, es común encontrar tal diversidad de propuestas a la hora de unificar varios conceptos que, no sin cierta ironía, en conjunto contribuyen todavía más a mantener la heterogeneidad del campo. En este capítulo hacemos la principal propuesta "original" de nuestro trabajo (hasta donde nosotros sabemos, no ha habido trabajos anteriores proponiendo exactamente el mismo enfoque): se trataría de nuestra propia propuesta para unificar

[1] De hecho, va más allá que esto: para que una prueba sea *realmente* de cero-conocimiento, tiene que satisfacer también la propiedad de que, si un observador externo O presenciase la interacción completa entre V y D, no quedaría convencido de nada; esto suele formularse como "el historial de transacciones del protocolo puede ser simulado por un único agente fraudulento", e, intuitivamente, esto puede comprenderse mejor imaginando una situación hipotética en la que el verificador y el demostrador en realidad estarían "compinchados" para tratar de probar su posesión de una información, que en realidad ninguno de ellos tiene, a un tercero.

[2] En este caso, tiene que ver con la inviabilidad computacional de adivinar el valor de la clave privada a partir de la clave pública. Otra forma de expresar esto es que la probabilidad de calcular la clave privada a partir de la clave pública por métodos computacionales conocidos y en un tiempo "razonable" es ínfima.

la lógica epistémica probabilística de Halpern y Fagin [3], expuesta en el capítulo anterior, con la lógica epistémica dinámica de anuncios públicos que hemos expuesto en el capítulo 3, generalizando la propuesta de Kooi [4] (como veremos, esta tiene grandes limitaciones en cuanto a los espacios probabilísticos que se pueden considerar).

5.2. Preámbulo: la propuesta de Kooi

Antes de introducir nuestra propuesta, presentaremos brevemente la de Kooi para que el lector pueda observar claramente los paralelismos entre ambas, así como los conceptos más importantes que hemos tomado de esta, y también los que hemos decidido descartar. Como siempre, empezamos definiendo el nuevo lenguaje y la semántica sobre la que se interpreta.

Definición 5.1. ***El lenguaje*** $\mathcal{LP}_{\mathcal{K}[]_{Kooi}}$**.** *Dado un conjunto finito de agentes A y un conjunto numerable de proposiciones atómicas At, definimos el lenguaje epistémico probabilístico con anuncios públicos $\mathcal{LP}_{\mathcal{K}[]_{Kooi}}(A, At)$ con la siguiente BNF:*

$$\varphi ::= p \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid K_a\varphi \mid [\varphi_1]\varphi_2 \mid \sum_{k=1}^{n} Q_k Pr_a(\varphi_k) \geq Q \tag{5.1}$$

Con $p \in At$ y $a \in A$.

□

Para interpretar las fórmulas de $\mathcal{LP}_{\mathcal{K}[]_{Kooi}}$, Kooi introduce una noción simplificada de los modelos de Kripke probabilísticos; nosotros lidiaremos con esto de una forma ligeramente distinta, aunque totalmente equivalente, en aras de la consistencia con el resto de nuestro trabajo.

Definición 5.2. ***Semántica de*** $\mathcal{LP}_{\mathcal{K}[]_{Kooi}}$**.** *Sea $M = \langle S, R, V, \Pi\rangle$ un modelo de Kripke probabilístico con agentes A y átomos At, y cuya asignación de probabilidades satisface, además, que para cada par agente-estado, el σ-álgebra asociada es $\mathfrak{A}_{(a,s)} = \mathcal{P}(S_{(a,s)})$ (el "σ-álgebra discreta"); la verdad de una fórmula en $\mathcal{LP}_{\mathcal{K}[]_{Kooi}}$ se interpreta según las siguientes cláusulas semánticas:*

$$\begin{array}{lll}
M, s \vDash p & sii & s \in V(p),\ p \in At \\
M, s \vDash \neg\varphi & sii & M, s \nvDash \varphi \\
M, s \vDash \varphi \wedge \psi & sii & M, s \vDash \varphi \text{ y } M, s \vDash \psi \\
M, s \vDash K_a\varphi & sii & \text{Para todo } t \in S:\ R_a st \Rightarrow M, t \vDash \varphi \\
M, s \vDash [\varphi]\psi & sii & M_{|\varphi}, s \vDash \psi \\
M, s \vDash \sum_{(k=0)}^{n} Q_k Pr_a(\varphi_k) \geq Q & sii & \sum_{k=0}^{n} Q_k Pr_{(a,s)}(S_{(a,s)}(\varphi_k)) \geq Q
\end{array} \tag{5.2}$$

Donde:

- *La restricción por una fórmula φ de un modelo de Kripke probabilístico M se define como $M_{|\varphi} := \langle S', R', V', \Pi' \rangle$, con $\Pi'_{(a,s)} := \langle S'_{(a,s)}, \mathcal{P}(S'_{(a,s)}), Pr'_{(a,s)} \rangle$, donde*

$$
\begin{aligned}
S' &:= S \\
R'_a &:= \{(s,t) \in R_a \mid M, t \vDash \varphi\} \\
V' &:= V \\
S'_{(a,s)} &:= \begin{cases} S_{(a,s)} & si\ P_{(a,s)}(\varphi) = 0 \\ S_{(a,s)}(\varphi) & en\ otro\ caso \end{cases} \\
P'_{(a,s)}(E) &:= \begin{cases} P_{(a,s)}(E) & si\ P_{(a,s)}(\varphi) = 0 \\ \frac{P_{(a,s)}(E(\varphi))}{P_{(a,s)}(\varphi)} & en\ otro\ caso \end{cases}
\end{aligned}
$$

- *Cualquier otra notación que se haya definido en otro capítulo se define aquí de la misma forma.*

□

Observación 5.1. El lector podrá apreciar de inmediato bastantes diferencias entre las filosofías de la propuesta anterior y las que hemos estado haciendo a lo largo del resto de nuestro trabajo. Para empezar, no es difícil percatarse de que la familia $\mathcal{S}5$ no es cerrada bajo la operación de restricción tal y como aparece en esta propuesta; en efecto, no es difícil encontrar ejemplos de modelos en $\mathcal{S}5$ cuyas restricciones según esta definición ya no están en $\mathcal{S}5$. En particular, esto hace que el enfoque de Kooi no sea del todo adecuado para tratar situaciones epistémicas; de hecho, si nos fijamos solo en la parte no probabilística de esta definición de restricción, vemos que coincide con la que propone van Ditmarsch [1] en la sección 4.9 de su manual, donde discute precisamente sobre cómo se podría relajar su concepto de anuncio público para adecuarlo mejor a modelos doxásticos.

Un aspecto de esta propuesta que podría verse como positivo es que nos permite prescindir de la medida interior para la cláusula semántica del operador probabilístico, que algunos lectores podrían considerar un tanto poco natural. No obstante, esto ocurre a costa de una gran pérdida de generalidad, dado que lo que asegura que todos nuestros conjuntos sean medibles es que en todas las asignaciones de probabilidad se usa el "σ-álgebra discreta".

Finalmente, otro de los aspectos que podrían considerarse problemáticos en esta propuesta es que la actualización de la asignación de probabilidades en el caso de que la probabilidad de la fórmula anunciada sea 0 no es del todo natural; de hecho, el propio Kooi discute esta cuestión brevemente en su artículo. En la propuesta que ofrecemos en este capítulo hemos tratado de dar un sentido algo más natural a lo que ocurre en este caso.

□

Kooi demuestra que su concepto de restricción está bien definido en el siguiente sentido:

Proposición 5.1. Si M, s es un modelo de Kripke probabilístico en las condiciones de la definición 5.2, entonces $M_{|\varphi}, s$ también lo es.

□

Nuestro reto, por tanto, es ofrecer una propuesta para la restricción que generalice a la anterior en el ámbito probabilístico, que mantenga el carácter epistémico de los modelos en el ámbito modal, y que esté "bien definida" en el sentido de la proposición anterior. En particular, un aspecto que sí nos parece fundamental preservar en nuestra propia propuesta es la forma en la que se actualiza la función de probabilidad, dado que se corresponde con la definición clásica de probabilidad condicionada [5]:

$$P(X \mid Y) := \frac{P(X \cap Y)}{P(Y)}$$

5.3. Sintaxis y semántica (de nuestra propuesta)

Las fórmulas de nuestro lenguaje $\mathcal{LP}_{\mathcal{K}[]}$ se definen de la misma forma que en 5.1; la principal diferencia está en su dimensión semántica.

Definición 5.3. ***Semántica de*** $\mathcal{LP}_{\mathcal{K}[]}$**.** *Sea $M = \langle S, R, V, \Pi\rangle$ un modelo de Kripke probabilístico*[3] *con agentes A y átomos At. La verdad de una fórmula en $\mathcal{LP}_{\mathcal{K}[]}$ se interpreta según las siguientes cláusulas semánticas:*

$$\begin{array}{lll}
M, s \vDash p & sii & s \in V(p),\ p \in At \\
M, s \vDash \neg\varphi & sii & M, s \nvDash \varphi \\
M, s \vDash \varphi \wedge \psi & sii & M, s \vDash \varphi \text{ y } M, s \vDash \psi \\
M, s \vDash K_a\varphi & sii & Para\ todo\ t \in S:\ R_a st \Rightarrow M, t \vDash \varphi \\
M, s \vDash [\varphi]\psi & sii & M, s \vDash \varphi \Longrightarrow M_{|\varphi}, s \vDash \psi\ (\%) \\
M, s \vDash \sum_{(k=0)}^{n} Q_k Pr_a(\varphi_k) \geq Q & sii & \sum_{k=0}^{n} Q_k Pr_{*(a,s)}(S_{(a,s)}(\varphi_k)) \geq Q
\end{array} \tag{5.3}$$

Donde:

- *La restricción por una fórmula φ tal que de un modelo de Kripke probabilístico M se define como $M_{|\varphi} := \langle S', R', V', \Pi'\rangle$, con $\Pi'_{(a,s)} := \langle S'_{(a,s)}, \mathfrak{A}'_{(a,s)}, Pr'_{(a,s)}\rangle$, donde*

[3] Como discutiremos más adelante, en realidad permitiremos también que la parte probabilística sea el "espacio cuasi-probabilístico" resultante de considerar un "espacio muestral vacío".

$$
\begin{aligned}
S' &:= S(\varphi) \\
R'_a &:= R_a \cap (S' \times S') \\
V' &:= V \cap S' \\
S'_{(a,s)} &:= \begin{cases} S_{(a,s)}(\varphi) & si\ Pr^*_{(a,s)}(\varphi) > 0 \\ \emptyset & e.o.c. \end{cases} \\
\mathfrak{A}'_{(a,s)} &:= \begin{cases} \{E(\varphi) \mid E \in \mathfrak{A}_{(a,s)}\} & si\ Pr^*_{(a,s)}(\varphi) > 0 \\ \{\emptyset\} & e.o.c \end{cases} \\
Pr'_{(a,s)}(E) &:= \begin{cases} 1 & si\ E = S_{(a,s)}(\varphi) \\ 0 & si\ E = \emptyset \\ \frac{Pr^*_{(a,s)}(E)}{Pr^*_{(a,s)}(\varphi)} & e.o.c. \end{cases}
\end{aligned}
$$

- *La notación (*) en el punto anterior indica que se trata de la medida exterior (no confundir con la medida interior, indicada con la notación ($_*$)). Esta se define análogamente a la medida interior:*

$$Pr^*_{(a,s)}(G) := inf\{Pr_{(a,s)}(E) \mid E \in \mathfrak{A}_{(a,s)}, E \supseteq G\}$$

- *Cualquier otra notación que se haya definido en otro capítulo se define aquí de la misma forma.*

□

Nota 5.1. Con la definición de restricción anterior, pueden producirse casos degenerados en los que $S'_{(a,s)} = \emptyset$. En estos casos, se entenderá que $Pr'_{(a,s)}(\emptyset) = 0$, y no 1. Si bien esto va en contra de una de las propiedades típicas de las funciones de probabilidad, a saber, que $Pr'_{(a,s)}(S'_{(a,s)}) = 1$, nos parece un "parche" preferible a la otra posibilidad, dado que esta produciría problemas todavía más extraños en su interacción con la propiedad de aditividad numerable (un ejemplo: $1 = Pr'_{(a,s)}(\emptyset) = Pr'_{(a,s)}(\emptyset \sqcup \emptyset) = 1+1 = 2$). Además, como discutiremos más abajo, también nos parece una solución mucho más filosóficamente coherente.

En particular, a partir de ahora nos tomaremos la licencia de decir que la terna $\langle \emptyset, \{\emptyset\}, P^\emptyset \rangle =: \Pi_\emptyset$, donde $P^\emptyset$ es la función que asigna al conjunto vacío el valor 0, es un "espacio probabilístico".

□

Antes de alabar las virtudes de nuestra propuesta, comprobemos que, efectivamente, satisface la propiedad de buena definición deseada:

Teorema 5.1. ***La restricción de un modelo probabilístico es un modelo probabilístico.*** *Dado $M = \langle S, R, V, \Pi\rangle$ modelo de Kripke probabilístico y $\varphi \in \mathcal{LP}_{\mathcal{K}[]}$, la restricción $M_{|\varphi}$ tal y como aparece en la definición 5.3 sigue siendo un modelo de Kripke probabilístico.*

Demostración. En los casos en que $Pr^*_{(a,s)}(\varphi) = 0$ (y, en particular, en los casos en que $\Pi_{(a,s)} = \Pi_\emptyset$), esto se tiene trivialmente. Para el caso general:

Hay que comprobar tres puntos:

I) $S'_{(a,s)} \subseteq S'$

II) $\mathfrak{A}'_{(a,s)}$ es un σ-álgebra sobre $S'_{(a,s)}$.

III) $Pr'_{(a,s)}$ es una función de probabilidad sobre $\langle \mathfrak{A}'_{(a,s)}, S'_{(a,s)}\rangle$.

El primer punto es obvio, solo lo hemos incluido en la enumeración para que figure que no lo hemos olvidado. Probemos los dos puntos restantes.

Prueba del punto II. Tenemos que comprobar tres condiciones:

- **Conjunto unidad.** $S'_{(a,s)} \in \mathfrak{A}'_{(a,s)}$
- **Cerrada bajo complementario.** Si $E \in \mathfrak{A}'_{(a,s)}$, entonces $S'_{(a,s)} \setminus E \in \mathfrak{A}'_{(a,s)}$
- **Cerrada bajo unión numerable.** Si $\{E_i\}_{i=1}^{\infty}$ es una sucesión de conjuntos en $\mathfrak{A}'_{(a,s)}$, entonces $\bigcup_{i=1}^{\infty} E_i \in \mathfrak{A}'_{(a,s)}$

Conjunto unidad. Dado que $S_{(a,s)} \in \mathfrak{A}_{(a,s)}$ en todos los casos y $S'_{(a,s)} = S_{(a,s)}(\varphi)$, tenemos el resultado.

Cerrada bajo complementario. Sea $E \in \mathfrak{A}'_{(a,s)}$. Entonces $E = G(\varphi)$ para algún $G \in \mathfrak{A}_{(a,s)}$. Afirmamos que $(S_{(a,s)} \setminus G)(\varphi) = S_{(a,s)}(\varphi) \setminus G(\varphi)$, y con esta afirmación tenemos el resultado, pues $S_{(a,s)} \setminus G \in \mathfrak{A}_{(a,s)}$ por las propiedades del σ-álgebra. Justificamos esta afirmación en la forma de un lema:

Lema 5.1. Sea $\mathfrak{A}$ un σ-álgebra (no trivial[4]) asociada a algún par agente-estado en un modelo de Kripke probabilístico, sean $A, B \in \mathfrak{A}$, y sea $\varphi \in \mathcal{LP}_{\mathcal{K}[]}$. Entonces $(A \setminus B)(\varphi) = A(\varphi) \setminus B(\varphi)$.

Demostración. Escribimos la definición de ambos conjuntos:

$$(A \setminus B)(\varphi) = \{s \in A \mid s \notin B \ y \ M, s \vDash \varphi\} =: (1)$$

$$A(\varphi) \setminus B(\varphi) = \{s \in A \mid s \notin B(\varphi) \ y \ M, s \vDash \varphi\} =: (2)$$

Claramente la condición $s \notin B$ es más restrictiva que la condición $s \notin B(\varphi)$, de modo que $(1) \subseteq (2)$. Para ver la contención contraria, consideremos $s \in (2)$; para ver que $s \in (1)$, tenemos que comprobar que $s \notin B$. Sabemos que $s \notin B(\varphi)$, de modo que la única forma de que $s \in B$ es que $s \in B \setminus B(\varphi)$. Pero $s \in B \setminus B(\varphi) \Rightarrow M, s \nvDash \varphi$, lo

[4] Con esto queremos decir que es algo más que el conjunto $\{\emptyset\}$.

cual contradice el hecho de que $s \in (2)$; por lo tanto, concluimos que $s \notin B$, y por lo tanto $(2) \subseteq (1)$.

Q.E.D.

□

Unión numerable. Sea $\{E_i\}_{i=1}^{\infty}$ una sucesión de conjuntos en $\mathfrak{A}'_{(a,s)}$. Entonces, para cada uno de ellos, existe $G_i \in \mathfrak{A}_{(a,s)}$ tal que $G_i(\varphi) = E_i$. Veamos que $(\bigcup_{i=1}^{\infty} G_i)(\varphi) = \bigcup_{i=1}^{\infty} G_i(\varphi)$:

$$s \in (\bigcup_{i=1}^{\infty} G_i)(\varphi) \Leftrightarrow s \in \bigcup_{i=1}^{\infty} G_i \ y \ M, s \vDash \varphi \Leftrightarrow \exists i \ t.q. \ s \in G_i \ y \ M, s \vDash \varphi$$

$$\Leftrightarrow \exists i \ t.q. \ s \in G_i(\varphi) \Leftrightarrow s \in \bigcup_{i=1}^{\infty} G_i(\varphi)$$

(En las líneas anteriores cuando hemos escrito $\exists i$ se ha sobreentendido que $i \in \{1, 2, 3, ...\}$.)

Con esto tenemos el resultado, y con esto acabamos de probar también que $\mathfrak{A}'_{(a,s)}$ es un σ-álgebra.

Prueba del punto III. Tenemos que comprobar tres condiciones:

- $Pr'_{(a,s)}(S'_{(a,s)}) = 1$
- $Pr'_{(a,s)}(E) \geq 0$ para todo $E \in \mathfrak{A}'_{(a,s)}$
- **Aditividad numerable.** Sea $\{E_i\}_{i=1}^{\infty}$ una familia de conjuntos **disjuntos** en $\mathfrak{A}'_{(a,s)}$, entonces $Pr'_{(a,s)}(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} Pr'_{(a,s)}(E_i)$

Las dos primeras propiedades son obvias a partir de las definiciones. Para ver la "propiedad difícil" (aditividad numerable) basta comprobar que para $\{E_i\}_{i=1}^{\infty}$ disjuntos 2 a 2 se tiene

$$Pr^*_{(a,s)}(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} Pr^*_{(a,s)}(E_i)$$

Para probar la igualdad anterior, haremos uso de una serie de lemas auxiliares que probaremos a continuación:

Lema 5.2. Sean $\mathfrak{A}$ un σ-álgebra (no trivial) asociada a un par agente-estado en un modelo de Kripke probabilístico, $\varphi \in \mathcal{LP}_{\mathcal{K}[]}$, $\mathfrak{A}'$ la restricción (no trivial) de $\mathfrak{A}$ a φ, $E \in \mathfrak{A}'$ y $\bar{G} \in \mathfrak{A}'$ tal que $E \subseteq \bar{G}$. Entonces existe $G \in \mathfrak{A}$ tal que $G \subseteq \bar{G}$ y $G(\varphi) = E$.

Demostración. Si $\bar{G}(\varphi) = E$, tenemos el resultado tomando $G = \bar{G}$. Supongamos ahora que $\bar{G}(\varphi) \supsetneq E$. Entonces $\bar{G}(\varphi) = E \sqcup K$ con $K := \bar{G}(\varphi) \setminus E \in \mathfrak{A}'_{(a,s)}$. Por lo

tanto, existe $L \in \mathfrak{A}_{(a,s)}$ con $L(\varphi) = K$. Entonces, utilizando el lema 5.1, tenemos que $G = \bar{G} \setminus L$ satisface las condiciones del lema.

Q.E.D.

□

Lema 5.3. Sea $\mathfrak{A}$ un σ-álgebra (no trivial) asociada a un par agente-estado en un modelo de Kripke probabilístico, $\varphi \in \mathcal{LP}_{\mathcal{K}[]}$, $\mathfrak{A}'$ la restricción (no trivial) de $\mathfrak{A}$ a φ, sean $\{E_i\}_{i=1}^{\infty}$ conjuntos en $\mathfrak{A}'$ **disjuntos 2 a 2**, y sea $G \in \mathfrak{A}$ tal que $G(\varphi) = \bigcup_{i=1}^{\infty}$. Entonces existen $G_i \in \mathfrak{A}$ tales que $G_i(\varphi) = E_i$, los G_i son disjuntos 2 a 2, y $\bigcup_{i=1}^{\infty} G_i \subseteq G$.

Demostración. Empezamos tomando $G_i^0 \in \mathfrak{A}$ para cada i de forma que $G_i^0(\varphi) = E_i$ (esto se puede hacer por la definición de $\mathfrak{A}'$). A partir de estos construiremos otros conjuntos que satisfagan las condiciones del lema.

Primero, para asegurar que los conjuntos finales estarán todos contenidos en G, tomamos $G_i^1 = G_i^0 \cap G$. No es difícil comprobar que $(A \cap B)(\varphi) = A(\varphi) \cap B(\varphi)$, y por tanto $G_i^1(\varphi) = E_i$.

Ahora, para asegurar que son disjuntos 2 a 2, tomamos:

$$\begin{aligned} G_1 &= G_1^1 \\ G_2 &= G_2^1 \setminus G_1^1 \\ &\vdots \\ G_k &= G_k^1 \setminus (\bigcup_{i=1}^{k-1} G_i^1) \end{aligned}$$

Dichos conjuntos son claramente disjuntos, siguen estando contenidos en G y, en todos los casos, por el lema 5.1, se comprueba que $G_i(\varphi) = E_i$ (aquí es donde se usa que los E_i son disjuntos 2 a 2, pues gracias a esto tenemos que $E_k \setminus \bigcup_{i=1}^{k-1} E_i = E_k$ en todos los casos). Esto prueba el lema.

Q.E.D.

□

A partir de los dos lemas anteriores, se justifican todos los pasos que daremos a continuación, teniendo además en cuenta la monotonía de la función de probabilidad y la siguiente observación: sea A un subconjunto acotado inferiormente de $\mathbb{R}$ (por lo tanto tiene ínfimo en $\mathbb{R}$), y sea $A' \subseteq A$ tal que $\forall a \in A, \exists a' \in A' : a' \leq a$; entonces, $inf\ A = inf\ A'$. Los pasos en cuestión son los siguientes:

$$Pr^*_{(a,s)}(\bigcup_{i=1}^{\infty} E_i) = inf\{Pr_{(a,s)}(G) \mid G \supseteq \bigcup_{i=1}^{\infty} E_i, G \in \mathfrak{A}_{(a,s)}\} =$$

$$inf\{Pr_{(a,s)}(G) \mid G \in \mathfrak{A}_{(a,s)}, G(\varphi) = \bigcup_{i=1}^{\infty} E_i\} =$$

$$inf\{Pr_{(a,s)}(\bigcup_{i=1}^{\infty} G_i) \mid G_i \in \mathfrak{A}_{(a,s)}, G_i \cap G_j = \emptyset \; si \; i \neq j, G_i(\varphi) = E_i, i \in \{1,2,3,...\}\} =$$

$$(*)\; inf\{\sum_{i=1}^{\infty} Pr_{(a,s)}(G_i) \mid G_i \in \mathfrak{A}_{(a,s)}, G_i \cap G_j = \emptyset \; si \; i \neq j, G_i(\varphi) = E_i, i \in \{1,2,3,...\}\} =$$

$$(**)\; inf\{\sum_{i=1}^{\infty} Pr_{(a,s)}(G_i) \mid G_i \in \mathfrak{A}_{(a,s)}, G_i(\varphi) = E_i, i \in \{1,2,3,...\}\} =$$

$$\sum_{i=1}^{\infty}\{Pr_{(a,s)}(G_i) \mid G_i \in \mathfrak{A}_{(a,s)}, G_i(\varphi) = E_i\} = \sum_{i=1}^{\infty} Pr^*_{(a,s)}(E_i)$$

El paso (*) utiliza la aditividad numerable de la $P_{(a,s)}$

El paso (**) está justificado porque, haciendo una construcción similar a la del lema 5.3, podemos encontrar siempre conjuntos disjuntos que, por lo demás, satisfagan las mismas propiedades.

Con esto tenemos el resultado, y terminamos también de probar el teorema.

Q.E.D.

□

Ahora que ya hemos visto que efectivamente nuestra propuesta "funciona", vamos a tratar de enumerar sus principales puntos fuertes en comparación con la propuesta original de Kooi. Para empezar, la principal ventaja que podemos observar es que no requiere de ningún tipo de condición sobre los modelos de Kripke probabilísticos, lo cual significa una generalización enorme de la misma. Además, nuestra propuesta *sí* preserva el carácter epistémico de los modelos, dado que, al menos en lo que a la parte modal se refiere, la definición es idéntica a la presentada en 3.2; esto la hace mucho más adecuada para lidiar, en principio, con el tipo de situaciones que se suelen tratar en el campo de la lógica epistémica dinámica

Otra ventaja importante que observamos en nuestra propuesta es su tratamiento de las "actualizaciones con probabilidad nula". En la propuesta de Kooi, la consecuencia de actualizar con una fórmula φ de probabilidad 0 es que la parte probabilística del modelo permanece igual que antes; esto podría considerarse una solución sencilla a los problemas derivados de dividir por 0 que aparecerían de otra forma, pero que, filosóficamente, no tiene una justificación del todo sólida. Nuestra propuesta tampoco es filosóficamente impenetrable, pero creemos que se le puede dar una interpretación algo más satisfactoria: ante el anuncio público de una fórmula cuya probabilidad es 0, la

asignación de probabilidad actualizada es la "vacía"[5]. En cierto modo, esto es "análogo" a transformar un modelo de Kripke probabilístico en uno puramente modal, dado que la parte probabilística deja de aportarnos información útil; este comportamiento puede considerarse coherente, dado que, si bien la actualización con una fórmula de probabilidad 0 "no tiene sentido" desde un punto de vista probabilístico, sí puede tenerlo desde un punto de vista modal.

No obstante, también reconocemos que esta forma de lidiar con las actualizaciones de probabilidad nula puede tener efectos secundarios no deseados, aunque en nuestro trabajo no hemos podido detectar ninguno; el único problema que sí hemos podido predecir tiene que ver con las perspectivas de proporcionar una axiomática completa para este sistema generalizando los axiomas "típicos" ya conocidos; de esto hablaremos en el capítulo 7. En este trabajo tratamos la parte axiomática de manera muy superficial y especulativa, de modo que no podemos pronunciarnos definitivamente sobre esto; no obstante, al menos en calidad de conjetura, proporcionamos algunas intuiciones sobre posibles modificaciones a dichos axiomas que nos permitan reutilizarlos en este nuevo sistema. Sobre esta cuestión, por lo tanto, solo puedo sugerir mayores esfuerzos de investigación en el futuro.

Un último apunte que queremos dejar por escrito sobre nuestra propuesta tiene que ver con los motivos para el uso selectivo de la medida exterior en algunas partes de la misma. El motivo principal es técnico: el lector que haya seguido las demostraciones anteriores podrá comprobar que "nada funcionaría" si se estuviese utilizando la medida interior en la definición de la medida de probabilidad actualizada. No obstante, es cierto que la decisión de actualizar la parte probabilística del modelo con la "asignación vacía" cuando $Pr^*_{(a,s)}(\varphi) = 0$ es algo más arbitraria, y aquí sí podríamos haber elegido utilizar la medida interior. Se pueden proporcionar argumentos técnicos a favor y en contra de ambas opciones; en nuestro caso, la razón por la que nos hemos acabado decantando por utilizar la medida exterior es que esto nos permite hacer razonamientos como en el ejemplo 5.4.2 que veremos más adelante.

5.4. Algunos ejemplos ilustrativos

5.4.1. Ejemplo: Alí Babá y los n ladrones

Antes de estudiar más propiedades de este nuevo formalismo, trataremos de ponerlo en práctica con un ejemplo clásico en el área de la criptografía; en particular, se trata de una ilustración de la idea de *prueba de conocimiento cero* presentada por Jean-Jacques Quisquater [14] en la forma de un "cuento para niños", donde los sucesivos éxitos en los atracos de unos ladrones a Alí Babá constituyen una prueba de cero conocimiento de que o bien su mala suerte es de magnitudes cósmicas, o bien existe un

[5]Más aún: según una interpretación filosófica radical, ante la actualización con una fórmula de probabilidad 0, podría considerarse, por analogía con *ex falso quodlibet*, que el nuevo espacio muestral sería "el conjunto de todas las cosas posibles e imposibles", un conjunto tan inconmensurablemente enorme que no puede expresarse matemáticamente.

pasadizo secreto en la cueva[6].

Ejemplo 5.1. **Alí Babá y los n ladrones.** Alí Babá se encontraba un buen día paseando por el bazar de Bagdad cuando, de improvisto, notó un tirón en el brazo: un ladrón le había robado su bolso y se había dado a la fuga. Inmediatamente, Alí Babá empezó a correr tras él; lo persiguió y lo persiguió hasta que, finalmente, lo vio esconderse en una cueva. Alí Babá trató de seguirlo allí, pero, poco después de entrar, se dio cuenta de que la cueva se bifurcaba, y no había alcanzado a ver cuál de los dos caminos había tomado el ladrón, de modo que decidió seguir uno de ellos al azar. Al llegar al final del camino y ver que el ladrón no estaba allí, simplemente pensó: "Diantres, debo haber tenido mala suerte. Habrá tomado el otro camino."

Al día siguiente, de nuevo mientras paseaba tranquilamente por el bazar, otro ladrón le robó su nuevo bolso a Alí Babá, y, de nuevo, se escondió en la misma cueva, repitiéndose los acontecimientos del día anterior; cuando, de nuevo, Alí Babá se encontró con que el ladrón no estaba al final del camino que había elegido, pensó una vez más: "¡Diantres! Debo haber tenido mala suerte otra vez. A ver si mañana, y durante los próximos $(n-2)$ días, no me pasa esto de nuevo."

Pasaron n días, y, con ellos, fueron n los ladrones que robaron su bolso a Alí Babá, refugiándose después en la cueva y desapareciendo siempre misteriosamente. Cada vez que ocurría esto, Alí Babá exclamaba: "¡Diantres! ¡Otra vez debo haber tenido mala suerte! ¡Y qué mal están las calles últimamente!". Pero, con cada día que pasaba, aumentaban también sus sospechas de que en aquella cueva ocurría algo extraño. De modo que, cuando llegó el día n, tras su rutinario atraco, decidió explorar detenidamente la cueva, y se percató de que esta tenía una forma un tanto peculiar 5.1: era *casi* como un donut, salvo por una delgada pared que separaba sus dos extremos. Observando esta pared con algo más de detenimiento, se percató de algo más: a ambos lados de la pared había un pequeño teclado, un tanto anacrónico en el Siglo XII, que permitía introducir códigos de cuadro dígitos en una pequeña pantalla, y, junto a este extraño aparato, una nota que decía: "¡NO OLVIDAR: EL CÓDIGO ES 1234!". Las sospechas ya comenzaban a acumularse en la cabeza de Alí Babá, de modo que, astutamente, decidió introducir estos mismos cuatro dígitos en el teclado; de repente, en la pared se abrió un pequeño pasadizo. Alí Babá comprendió finalmente por qué aquellos ladrones siempre lograban escapársele: aunque tomase el mismo camino que ellos, los ladrones simplemente podían usar el pasadizo para salir por el otro.

Al día siguiente, cuando el $(n+1)$-ésimo ladrón entró en la cueva y trató de usar el mismo truco, Alí Babá simplemente se quedó esperándolo fuera.[7]

Tratemos de modelar esta situación en términos de nuestro lenguaje, $\mathcal{LP}_{\mathcal{K}[]}$. Más concretamente, nuestro principal objetivo será representar el hecho de que, con las sucesivas iteraciones del atraco diario, Alí Babá considerará cada vez más improbable la posibilidad de que la cueva no tenga ningún pasadizo oculto (tras la n-ésima iteración,

[6] El ejemplo explora de forma más concreta muchas de las propiedades que ha de tener una prueba para poder considerarse "de cero conocimiento". Nosotros nos limitaremos, en un principio, a modelar lo fundamental de la situación.

[7] Siempre me he preguntado por qué no permaneció esperando fuera desde el principio, pero esa es otra historia.

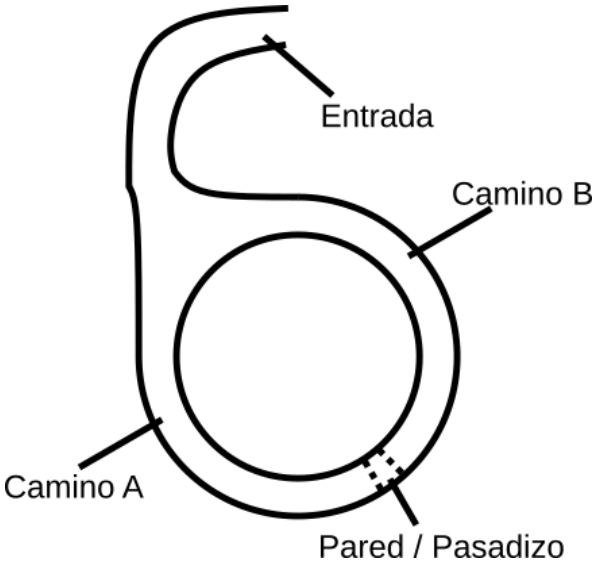

Figura 5.1: Una representación, no demasiado artística, de la cueva de Alí Babá (o más bien, de la cueva de sus atracadores).

Alí Babá considera que la improbabilidad de la situación es lo suficientemente significativa como para examinarla con más detenimiento). En particular, haremos uso de las siguientes notaciones:

- El átomo p representará la afirmación "Hay un pasadizo secreto en la cueva".
- Los átomos $I_1, ..., I_n, ...$ representarán, respectivamente, las afirmaciones "Se han producido al menos n atracos".
- El átomo F representa la afirmación "Se da al menos un intento de escape fallido"; su negación, por tanto, representa el hecho de que el ladrón logra escapar exitosamente en todos los intentos.

Proponemos el modelo de la figura 5.2 para representar la situación epistémica. Para aligerar la notación, hemos evitado escribir los casos en los que se tiene $\neg I_i$; debe sobreentenderse que, si no aparece explícitamente I_i, entonces se tiene su negación. Formalmente el modelo se define como sigue (a, Alí Babá, es el único agente):

- $S := S^0 \cup S^+ \cup S^-$, con
  - $S^0 := \{s_i^0, i \in \{1, 2, ...\}\}$
  - $S^+ := \{s_i, i \in \{1, 2, ...\}\}$
  - $S^- := \{\bar{s_i}, i \in \{1, 2, ...\}\}$
- $R_a := S \times S$
- $V_F = S^-$
- $V_{I_i} = \{s_j^0, s_j, \bar{s_j} \mid j \geq i\}$
- $V_p = S^0$
- $S_{(a,s_i^0)} = \{s_i^0\}$

- $S_{(a,s_i)} = S_{(a,\bar{s_i})} = \{s_i, \bar{s}_i\}$
- $\mathfrak{A}_{(a,s_i)} = \mathfrak{A}_{(a,\bar{s_i})} = \{\{s_i, \bar{s}_i\}, \{s_i\}, \{\bar{s}_i\}, \emptyset\}$
- $P_{(a,s_i)}(\{s_i\}) = P_{(a,\bar{s_i})}(\{s_i\}) = \frac{1}{2^k}$

Dicho modelo puede no parecer, en principio, la forma más intuitiva de representar la situación. No obstante, podemos ofrecer una interpretación sencilla para justificarlo: al principio, cuando todavía solo ha sufrido un atraco, Alí Babá considera la posibilidad de sufrir más atracos similares en el futuro. Además, en cada caso, subdivide la situación en tres sub-casos adicionales: en el primero, la cueva tiene un pasadizo secreto (o, por ser más fieles a nuestra versión del relato, "algo raro", sea un pasadizo u otro "truco" similar), y el ladrón siempre logrará escapar. En los sub-casos 2 y 3, no existe ningún "truco": el caso 2 representa, para cada número de "iteraciones", la situación en la que el ladrón logra escapar en cada una de ellas; por otra parte, el caso 3 representa, también para cada número de iteraciones, el conjunto de todas las situaciones en las que el ladrón es capturado al menos en una de dichas iteraciones.

En este modelo, el hecho de que se produzca un nuevo atraco puede representarse con el anuncio público del átomo I_{i+1}, donde i es el número de atracos actuales (inicialmente, $i = 1$); con cada anuncio de esta forma, el modelo restringido puede imaginarse como el modelo resultante de eliminar los estados de la columna correspondiente. En cualquiera de los estados del modelo inicial, se tiene:

$$M, s \vDash K_a\left(p \vee \left(\neg p \wedge P_a(\neg F) \leq \frac{1}{2}\right)\right)$$

Con el segundo atraco, la fórmula anterior puede ser sustituida por una más restrictiva:

$$M, s \vDash [I_2]K_a\left(p \vee \left(\neg p \wedge P_a(\neg F) \leq \frac{1}{4}\right)\right)$$

En general, con cada atraco sucesivo obtenemos una probabilidad cada vez menor en la segunda parte de la disyuntiva:

$$M, s \vDash [I_2]...[I_n]K_a\left(p \vee \left(\neg p \wedge P_a(\neg F) \leq \frac{1}{2^n}\right)\right)$$

Quizá un lector familiarizado con los conceptos y técnicas más comunes que se suelen aplicar en el área de la estadística haya establecido ya ciertos paralelismos entre estos y los razonamientos que tratamos de formalizar a través de las fórmulas anteriores. En efecto, la idea básica no se aleja demasiado de lo que en este área suele conocerse con el nombre de "contraste de hipótesis", cuya filosofía puede resumirse en la siguiente disyunción: "O bien aceptamos un hecho p, o bien aceptamos su negación $\neg p$ junto con la afirmación de que ha ocurrido un suceso altamente improbable".

Finalmente, otro aspecto importante que nuestra representación de esta situación mediante un modelo de Kripke probabilístico logra capturar es que basta con que el

ladrón fracase una sola vez en su huída para desmentir por completo la posibilidad de que exista un pasadizo; en efecto, no es difícil comprobar que el modelo resultante de realizar el anuncio público $[F]$ consistiría únicamente en la fila inferior de la figura 5.2, quedando reducidos los espacios muestrales "asociados" a dichos estados a los conjuntos unitarios formados por los propios estados, y, por lo tanto, dotando a dichos estados de probabilidad 1 en todos los casos (además de satisfacerse, obviamente, la fórmula $M_{|F}, \bar{s_i} \vDash K_a F$).

Una nota adicional sobre la interpretación de estos modelos es que no debe permitirse el anuncio público de la fórmula $\neg F$, o debe considerarse que tal anuncio no tendría un sentido claro. Esto puede parecer una solución improvisada para prevenir los problemas de interpretación resultantes de tal anuncio, y de hecho posiblemente lo sea dado que prevendría la posibilidad de escapes fallidos en iteraciones posteriores a la "actual". Una forma un tanto enrevesada de interpretar esto, no obstante, es que el anuncio público de $\neg F$ no representaría que el ladrón ha escapado exitosamente en todas las iteraciones hasta el momento, sino más bien que "sea cual sea el número de iteraciones que se produzcan, el ladrón jamás será atrapado"; para poder anunciar una afirmación como esta, sería necesario disponer de un "agente externo con información privilegiada sobre el futuro". En cambio, según esta interpretación sí que podría anunciarse F, dado que la afirmación que representa es que "en *alguna* iteración, el ladrón es capturado"; con que esto se produzca una sola vez, ya es cierto durante el resto de los tiempos, y por lo tanto puede anunciarse. Quizá podrían formalizarse estos razonamientos introduciendo conceptos adicionales de lógica temporal (otra de las ramas clásicas de la lógica modal), pero esto, desde luego, escapa por completo al ámbito de nuestro trabajo actual.

□

5.4.2. Ejemplo: Incertidumbre Probabilística e Incertidumbre Esencial (II)

En este otro ejemplo queremos ilustrar cómo nuestra definición de actualización proporciona interpretaciones satisfactorias a situaciones que de otra forma podrían parecer un tanto ambiguas.

Ejemplo 5.2. Supongamos que estamos de nuevo en la situación del ejemplo 4.1, y que por algún motivo consideramos que la forma más conveniente de representar la situación es a través del modelo **M0**, representado en el esquema 5.3. Recordemos que dicho modelo no nos permite asignar una probabilidad al conjunto definido por la proposición atómica a, y que la medida interior de dicho conjunto es 0. Supongamos ahora que se anuncia públicamente que el valor del input introducido ha sido, por ejemplo, el 1 (es decir, se anuncia públicamente la proposición B_1). Entonces es fácil comprobar que el modelo resultante es el de la figura 5.4, y, en este modelo, si podemos asignar una probabilidad (de 1/2) al conjunto definido por a.

Obsérvese que esto es posible porque hemos especificado explícitamente en nuestra semántica que las "actualizaciones nulas" solo se produzcan cuando la fórmula φ por la que se restringe satisface que $Pr^*_{(a,s)}(\varphi) = 0$, y, en este caso, la proposición atómica B_1

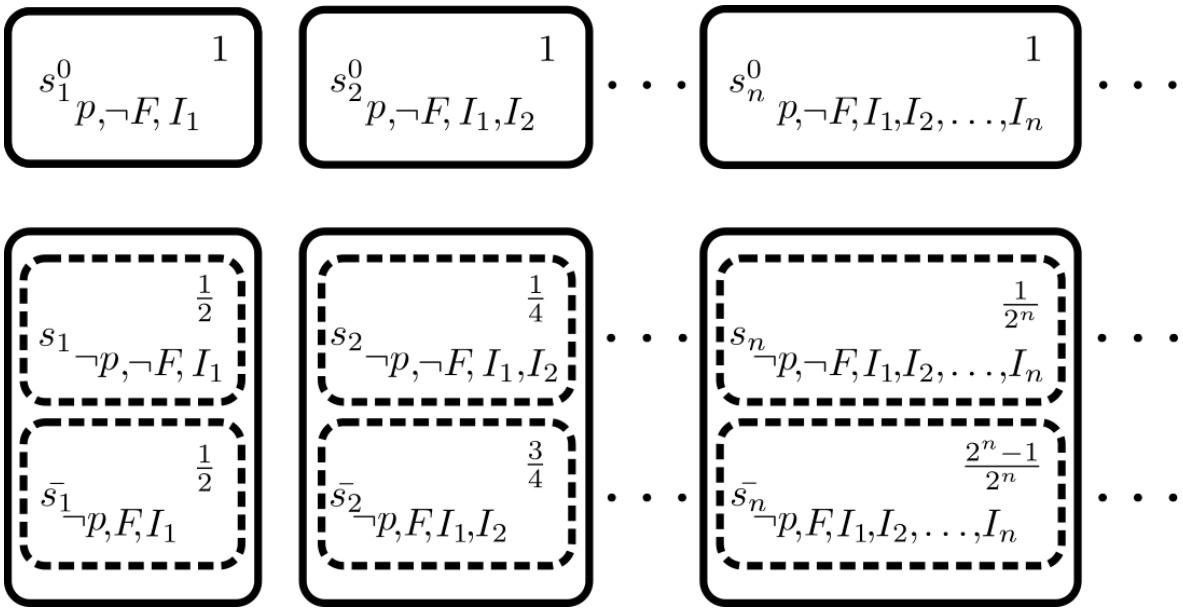

Figura 5.2: Esquema de un modelo de Kripke para representar la situación epistémica de Alí Babá. Las cajas de borde continuo representan los espacios muestrales asociados a cada asignación de probabilidades, mientras que las de borde discontinuo representan los conjuntos medibles no triviales (en este caso, los conjuntos unitarios). El modelo satisface **CONS**, y la relación de accesibilidad para Alí Babá es la total.

tiene medida exterior $\frac{1}{2}$ en todos los casos. En efecto, si nuestro criterio para hacer nulo un espacio probabilístico actualizado fuese la medida interior, habríamos obtenido un resultado mucho menos satisfactorio: se trataría de un modelo "de parte probabilística vacía" (pues la medida interior de la proposición atómica B_1 es 0), y cualquier fórmula tendría probabilidad 0 en el modelo resultante.

□

5.5. Algunas propiedades del lenguaje $\mathcal{LP}_{\mathcal{K}[]}$

A continuación presentamos algunos resultados básicos de nuestro lenguaje. Para empezar, nos gustaría comprobar si las propiedades presentadas en el capítulo anterior (**CONS**, **OBJ**, **SDP**, **UNIF**, **MEAS**) se conservan mediante el concepto de restricción que hemos propuesto.

Nota 5.2. En todas las demostraciones a continuación, representaremos con apóstrofes (') los elementos de los modelos restringidos, y sin apóstrofes los de los modelos originales.

□

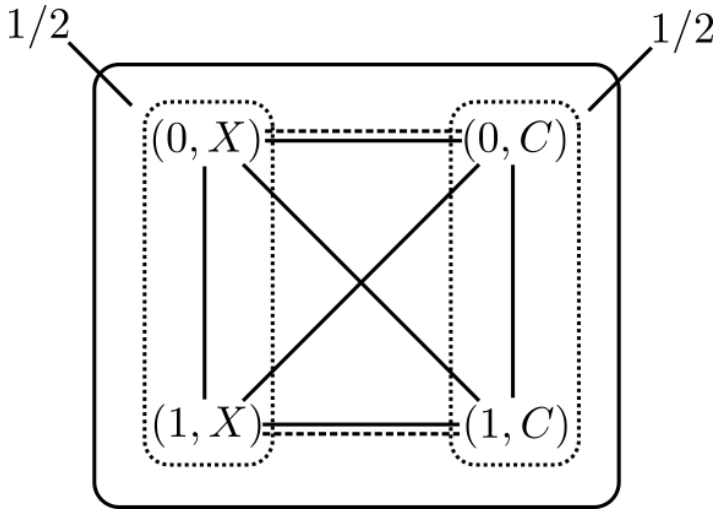

Figura 5.3: Representación del modelo **M0**. Las "cajas punteadas" representan los subconjuntos medibles de cada espacio muestral. En este caso, el único espacio muestral está representado con la caja de líneas continuas. Las rectas continuas representan la relación de accesibilidad del nodo a, y las rectas discontinuas representan la relación de accesibilidad del nodo b.

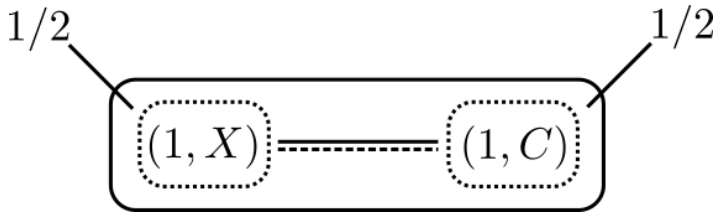

Figura 5.4: Representación del modelo $M0_{|B_1}$.

Proposición 5.2. **CONS se conserva mediante restricciones.** Sea M un modelo satisfaciendo **CONS** y $\varphi \in \mathcal{LP}_{\mathcal{K}[]}$ arbitraria. Entonces $M_{|\varphi}$ también satisface **CONS**.

Demostración. Sean $a \in A$ y $s \in S' = S(\varphi)$ arbitrarios. Sabemos que $S_{(a,s)} \subseteq R_a(s)$. Tenemos que probar que $S'_{(a,s)} \subseteq R'_a(s)$. Para empezar, en los casos en que $S'_{(a,s)} = \emptyset$, esto se tiene siempre. Consideremos pues los casos en que $S'_{(a,s)} = S_{(a,s)}(\varphi)$. No es difícil ver que:

$$R'_a(s) = R_a(s)(\varphi)$$

Sea $t \in S'_{(a,s)}$. Hay que ver que $t \in R'_a(s)$.

$$t \in S'_{(a,s)} \Leftrightarrow t \in S_{(a,s)} \ y \ M, t \vDash \varphi \Rightarrow t \in R_a(s) \ y \ M, t \vDash \varphi \Leftrightarrow$$
$$\Leftrightarrow t \in R_a(s)(\varphi) \Leftrightarrow t \in R'_a(s)$$

Q.E.D.

□

Proposición 5.3. **OBJ, SDP y UNIF se conservan mediante restricciones.** Sea M un modelo satisfaciendo **OBJ** (resp. **SDP**, **UNIF**). Sea $\varphi \in \mathcal{LP}_{\mathcal{K}[]}$. Entonces $M_{|\varphi}$ también satisface **OBJ** (resp. **SDP**, **UNIF**).

Demostración. Los tres casos son triviales. Veamos el caso de **UNIF**, que es marginalmente menos trivial.

Sea M un modelo satisfaciendo **UNIF**, es decir, dados $a \in A$ y $s, t \in S$ arbitrarios, $t \in S_{(a,s)} \Rightarrow \Pi_{(a,s)} = \Pi_{(a,t)}$. Sea $\varphi \in \mathcal{LP}_{\mathcal{K}[]}$; tenemos que probar que M_φ satisface **UNIF**.

Sean $a \in A$ y $s, t \in S(\varphi)$ arbitrarios. Supongamos primero que $S'_{(a,s)} = \emptyset$. Entonces $t \notin S'_{(a,s)}$, y por lo tanto se tiene la implicación. En otro caso, $t \in S'_{(a,s)} \Leftrightarrow t \in S_{(a,s)} \ y \ M, t \vDash \varphi \Rightarrow \Pi_{(a,s)} = \Pi_{(a,t)}$. A partir de esto es trivial comprobar que $\Pi'_{(a,s)} = \Pi'_{(a,t)}$.

Q.E.D.

□

Veamos el caso de **MEAS**. Debemos tener en cuenta que, en el contexto de $\mathcal{LP}_{\mathcal{K}[]}$, esta propiedad significa que los conjuntos $S_{(a,s)}(\varphi)$ son medibles *para cualquier fórmula de* $\mathcal{LP}_{\mathcal{K}[]}$, y no solo de $\mathcal{LP}_{\mathcal{K}}$ (como en el capítulo anterior). De hecho, para diferenciar ambas propiedades, diremos a partir de ahora que un modelo satisface **MEAS** si solo nos estamos limitando al caso de $\mathcal{LP}_{\mathcal{K}}$, y que satisface **MEAS-D** si también incluimos fórmulas con el operador de anuncio público.

Proposición 5.4. **MEAS-D se conserva mediante los anuncios públicos.** Sea M un modelo satisfaciendo **MEAS-D**, y sea $\varphi \in \mathcal{LP}_{\mathcal{K}[]}$. Entonces $M_{|\varphi}$ también satisface **MEAS-D**.

Demostración. Sea $\psi \in \mathcal{LP}_{\mathcal{K}[]}$. Queremos probar que para todo $a \in A$, $s \in S'$, $S'_{(a,s)}(\psi)$ es medible. Obviamos el caso trivial en el que $S'_{(a,s)} = \emptyset$. En el otro caso,

$$S'_{(a,s)}(\psi) = \{t \in S'_{(a,s)} \mid M_{|\varphi}, t \vDash \psi\} = \{t \in S_{(a,s)}(\varphi) \mid M_{|\varphi}, t \vDash \psi\} = \\ \{t \in S_{(a,s)} \mid M, t \vDash \varphi \ y \ M_{|\varphi} \vDash \psi\} = S_{(a,s)}(\langle\varphi\rangle\psi)$$

Y este último conjunto es medible porque M satisface **MEAS-D**.

Q.E.D.

□

En el capítulo anterior vimos una condición suficiente para que un modelo satisficiese **MEAS**. En este capítulo veremos una condición análoga para **MEAS-D**. No obstante, antes tenemos que probar varios resultados técnicos previos.

Proposición 5.5. **PMEAS se conserva mediante restricciones.** Sea M un modelo satisfaciendo **PMEAS** y $\varphi \in \mathcal{LP}_{\mathcal{K}[]}$ una fórmula arbitraria. Entonces M_φ satisface **PMEAS**.

Demostración. Sea $p \in At$. Sean $a \in A$, $s \in S' = S(\varphi)$ arbitrarios. Queremos probar que $S'_{(a,s)}(p) \in \mathfrak{A}'_{(a,s)}$.

Caso 1. $S'_{(a,s)} = \emptyset$. Entonces $S'_{(a,s)}(p) = \emptyset \in \mathfrak{A}'_{(a,s)}$.

Caso 2. $S'_{(a,s)} = S_{(a,s)}(\varphi)$. Por **PMEAS**, $S_{(a,s)}(p) \in \mathfrak{A}_{(a,s)}$. Por otra parte,

$$\mathfrak{A}'_{(a,s)} = \{E(\varphi) \mid E \in \mathfrak{A}_{(a,s)}\}$$

Queremos comprobar que $S'_{(a,s)}(p) \in \mathfrak{A}'_{(a,s)}$. En particular, afirmamos que

$$S'_{(a,s)}(p) = S_{(a,s)}(\varphi, p)$$

Escribimos las definiciones de ambos conjuntos:

$$S'_{(a,s)}(p) = \{t \in S'_{(a,s)} \mid M_{|\varphi}, t \vDash p\} = \{t \in S_{(a,s)} \mid M, t \vDash \varphi, \ M_{|\varphi}, t \vDash p\}$$

$$S_{(a,s)}(\varphi, p) = \{t \in S_{(a,s)} \mid M, t \vDash \varphi, \ M, t \vDash p\}$$

Es fácil ver que ambos conjuntos coinciden: en efecto, la cláusula $M_{|\varphi}, t \vDash p$ se puede reescribir como $t \in V(p)$ *y* $t \in S(\varphi)$, lo cual nos permite reescribir el primer conjunto como:

$$S'_{(a,s)}(p) = \{t \in S'_{(a,s)} \mid M_{|\varphi}, t \vDash p\} = \{t \in S_{(a,s)} \mid M, t \vDash \varphi,\ t \in V(p) \text{ y } t \in S(\varphi)\}$$

Pero $t \in S(\varphi)$ y $M, t \vDash \varphi$ están afirmando lo mismo, de modo que nos quedamos con el conjunto $S_{(a,s)}(\varphi, p)$.

Q.E.D.

□

Lema 5.4. Sea M un modelo de Kripke probabilístico, sea $\varphi \in \mathcal{LP}_{\mathcal{K}[]}$, y sea $M_{|\varphi}$ la restricción de M a φ. Sean $a \in A$ y $s \in S$ tales que tanto $\mathfrak{A}_{(a,s)}$ como $\mathfrak{A}'_{(a,s)}$ son no triviales (es decir, $Pr^*_{(a,s)}(\varphi) > 0$). Sea $E \in \mathfrak{A}'_{(a,s)}$. Entonces, existe $G \in \mathfrak{A}_{(a,s)}$ tal que $G(\varphi) = E$ y $Pr_{(a,s)}(G) = Pr^*_{(a,s)}(E)$.

Demostración. Por el lema 5.2, sabemos que

$$P := Pr^*_{(a,s)}(E) = inf\{Pr_{(a,s)}(G) \mid G \in \mathfrak{A}_{(a,s)} \text{ y } G(\varphi) = E\}$$

En particular, dado $\epsilon > 0$, existen G_i tales que

$$G_i(\varphi) = E \qquad y \qquad Pr_{(a,s)}(G_i) \geq P + \frac{\epsilon}{2^i}$$

Sea $G := \bigcap_{i=0}^{\infty} G_i$. Entonces $G(\varphi) = E$ y, además,

$$Pr_{(a,s)}(G) \geq P + \frac{\epsilon}{2^k} \quad \forall k \in \mathbb{N}$$

Es decir, $Pr_{(a,s)}(G) = P$.

Q.E.D.

□

Proposición 5.6. **Definición de SIGNIF + SIGNIF se conserva mediante las restricciones.** Decimos que un modelo M satisface **SIGNIF** (de "no tiene conjuntos in**SIGNIF**icantes") si, para todo $a \in A$ y $s \in S$, para todo $E \in \mathfrak{A}_{(a,s)}$ se tiene que $Pr_{(a,s)}(E) = 0 \Rightarrow E = \emptyset$.

Sea M un modelo satisfaciendo **SIGNIF** y sea $\varphi \in \mathcal{LP}_{\mathcal{K}[]}$ una fórmula arbitraria. Entonces $M_{|\varphi}$ también satisface **SIGNIF**.

Demostración. Sean $a \in A$, $s \in S(\varphi)$. Si $S'_{(a,s)} = \emptyset$, estamos en el caso trivial. Supongamos pues que $S'_{(a,s)} \neq \emptyset$.

Sea $E \in \mathfrak{A}'_{(a,s)}$, y supongamos que $E \neq \emptyset$; hay que ver entonces que $Pr'_{(a,s)}(E) > 0$. Por el lema anterior, sabemos que existe $G \in \mathfrak{A}_{(a,s)}$ tal que $G(\varphi) = E$ y cuya medida coincide con la medida exterior de E. Dado que $E \neq \emptyset$, G tampoco es vacío. Ahora bien, por **SIGNIF**, esto implica que la medida de G es positiva. Por lo tanto,

$$Pr'_{(a,s)}(E) = \frac{Pr_{(a,s)}(G)}{Pr^*_{(a,s)}(S_{(a,s)}(\varphi))} > 0$$

Q.E.D.

□

Teorema 5.2. ***Condición suficiente para MEAS-D en*** $\mathcal{LP}_{\mathcal{K}[]}$**.** *Sea M un modelo satisfaciendo* ***CONS***, ***OBJ***, ***UNIF***, ***PMEAS*** *y* ***SIGNIF***, *y sea $\varphi \in \mathcal{LP}_{\mathcal{K}[]}$. Entonces M satisface* ***MEAS-D****.*

Demostración. El caso base, así como los casos para las conectivas proposicionales, el operador modal y el operador probabilístico, se prueban como en la proposición 4.2. Veamos el caso para el operador de anuncio público (por inducción estructural sobre las fórmulas de $\mathcal{LP}_{\mathcal{K}[]}$).

Sean φ y ψ tales que, para cualquier modelo M satisfaciendo **CONS**, **OBJ**, **UNIF**, **PMEAS** y **SIGNIF** y tales que, para cualquier $a \in A$ y $s \in S$, los conjuntos $S_{(a,s)(\varphi)}$ y $S_{(a,s)}(\psi)$ son medibles. Tenemos que probar que $S_{(a,s)}([\psi]\varphi)$ es medible.

Caso 1. El caso trivial: $S_{(a,s)} = \emptyset$. En tal caso, $S_{(a,s)}([\psi]\varphi) = \emptyset$, que obviamente es medible.

Caso 2. $S_{(a,s)} \neq \emptyset$. Entonces,

$$\begin{aligned} S_{(a,s)}([\psi]\varphi) = \{t \in S_{(a,s)} \mid M, t \vDash \psi \Rightarrow M_{|\psi}, t \vDash \varphi\} = \\ \{t \in S_{(a,s)} \mid M, t \nvDash \psi \ o \ (M, t \vDash \psi \ y \ M_{|\psi}, t \vDash \varphi)\} = \\ \{t \in S_{(a,s)} \mid M, t \nvDash \psi\} \cup \{t \in S_{(a,s)} \mid M, t \vDash \psi \ y \ M_{|\psi}, t \vDash \varphi\} \end{aligned}$$

El primer conjunto de la unión en la última línea es $S_{(a,s)}(\neg\varphi)$, y es fácil comprobar que es medible usando la hipótesis de inducción y las propiedades del σ-álgebra.

El segundo conjunto de la unión podemos reescribirlo como:

$$\{t \in S_{(a,s)}(\psi) \mid M_{|\psi}, s \vDash \varphi\} =: (2)$$

Nuestro objetivo será probar que $(2) = S'_{(a,s)}(\varphi)$, dado que con esto obtendríamos el resultado: en efecto, por los lemas anteriores, el modelo $M_{|\psi}$ satisface **CONS**, **OBJ**,

UNIF, **PMEAS** y **SIGNIF**, y por hipótesis de inducción esto implica que $S'_{(a,s)}(\varphi)$ es medible. Probemos pues este objetivo.

Caso 2.1. $Pr_{(a,s)}(\psi) = 0$. Por **SIGNIF**, esto implica que $S_{(a,s)}(\psi) = \emptyset$, y, por otra parte, por definición, $S'_{(a,s)}(\varphi) = \emptyset$. Por lo tanto, $(2) = S_{(a,s)}(\psi)(\varphi) = S'_{(a,s)}(\varphi)$, y tenemos el resultado.

Caso 2.2. $Pr_{(a,s)}(\psi) > 0$. Entonces $S'_{(a,s)} = S_{(a,s)}(\psi)$, y, de la misma forma, tenemos el resultado.

Q.E.D.

□

Corolario 5.1. En particular, si M es un modelo satisfaciendo **CONS**, **OBJ**, **UNIF**, **PMEAS** y **SIGNIF**, entonces tanto este como su restricción a cualquier fórmula satisfacen **MEAS-D**.

□

Un último resultado que sería interesante tratar de probar sería algo análogo al teorema 3.1, que nos permite reescribir cualquier fórmula de $\mathcal{L}_{K[]}$ como una fórmula de $\mathcal{L}_K$. En este caso, presentamos un resultado que por sí solo es mucho más débil, pero que creemos que puede abrir el camino a otros resultados algo más interesantes.

Proposición 5.7. **Reescritura para el operador probabilístico en MEAS-D.** La siguiente equivalencia es cierta en modelos satisfaciendo **MEAS-D**:

$$[\psi]\sum_{i=1}^{n} Q_i Pr_a(\varphi_i) \geq Q \leftrightarrow \left(Pr_a(\varphi) > 0 \wedge \left(\varphi \rightarrow \sum_{i=1}^{n} Q_i Pr_a(\varphi \wedge [\varphi]\varphi_i) \geq Q\ Pr_a(\varphi)\right)\right) \vee (Pr_a(\varphi) = 0 \wedge (\varphi \rightarrow Pr_a(\bot) \geq Q))$$

Demostración. La prueba de la validez de la última fórmula es tediosa, pero fácil de ver a partir de la semántica de $\mathcal{LP}_{\mathcal{K}[]}$, teniendo en cuenta que se evalúa en modelos satisfaciendo **MEAS-D**.

Q.E.D.

□

Observación 5.2. La subfórmula $Pr_a(\bot) \geq Q$ básicamente expresa que $0 \geq Q$.

□

No hemos tratado de encontrar un contraejemplo a la "reescritura" del operador probabilístico en el caso general (modelo no satisfaciendo **MEAS-D**), pero, en principio, conjeturamos que debe existir tal contraejemplo, dado que la condición $Pr^*_{(a,s)}(\varphi) = Pr_{(a,s)}(\varphi) = Pr_{*(a,s)}(\varphi)$ parecería un requisito natural para que dicha

reescritura sea correcta en general[8]. No obstante, por supuesto, no podemos afirmar o negar nada sobre esta cuestión hasta que no se hayan encontrado resultados concluyentes a favor o en contra. Tampoco descartamos la posibilidad de encontrar una reescritura más compleja en el caso general, aunque nos decantamos más por la posibilidad de que, en el caso general, no sea posible hallar tal sistema de reescrituras.

Por otra parte, la proposición anterior adquiriría un peso mucho mayor si se lograse probar que, en adición a las reescrituras del teorema 3.1, proporciona una traducción de $\mathcal{LP}_{\mathcal{K}}$ a $\mathcal{LP}_{\mathcal{K}[]}$, al menos restringiéndonos a modelos que satisfacen **MEAS-D**. Una prueba así requeriría de la definición de algunas construcciones teóricas adicionales que no hemos podido cubrir en este trabajo, pero que tampoco son excesivamente complejas; en particular, nosotros creemos que es posible, y que no es difícil (disponiendo de las herramientas adecuadas), probar este hecho. En el caso de tenerse dicho resultado, se tendría también el siguiente corolario:

Conjetura 5.1. **Bisimulación en $\mathcal{LP}_{\mathcal{K}[]}$.** Sean M_1, M_2 dos modelos de Kripke probabilísticos satisfaciendo **MEAS-D**, y sean s_1, s_2 dos estados en los respectivos modelos. Entonces $M_1, s_1 \underline{\leftrightarrow} M_2, s_2 \Longrightarrow M_1, s_1 \equiv_{\mathcal{LP}_{\mathcal{K}[]}} M_2, s_2$.

□

[8] Aunque también barajamos la conjetura de que sea suficiente imponer una condición más débil sobre los modelos, como **SIGNIF**. Dicha propiedad asegura, en particular, que $Pr_{*(a,s)}(\varphi)$ es nula si y solo si lo es $Pr^{*}_{(a,s)}(\varphi)$.

6. Conocimiento Común

6.1. Idea y motivación

Hasta ahora, los formalismos presentados solo nos permiten hacer afirmaciones sobre nociones que, en última instancia, son reducibles al conocimiento individual de cada uno de los agentes involucrados. El operador de conocimiento común C_B que introducimos en este capítulo nos permite expresar, adicionalmente, propiedades epistémicas del sistema en conjunto, o de un subconjunto del mismo. Halpern y Fagin [3] presentan en su artículo una "versión probabilística" de este operador, que no cubriremos aquí pero cuya inclusión en investigaciones futuras también sería de gran interés.

La idea intuitiva del operador de conocimiento común puede expresarse en términos de los operadores grupales presentados en la anotación 2.7. Ya sabemos que la fórmula $E_B\varphi$ se leería "Todos los agentes en B saben φ". Consideremos ahora las iteraciones sucesivas de este operador: por ejemplo, la fórmula $E_B^2\varphi$ se leería "Todos los agentes en B saben que todos los agentes en B saben φ", y la fórmula $E_B^k\varphi,\ k > 0$ se leería "Todos los agentes en B saben que todos los agentes en B saben que (...) todos los agentes en B saben φ" (k veces). En este sentido, el operador de conocimiento común da el "salto al infinito", pues una forma natural de interpretar la fórmula $C_B\varphi$ es la siguiente:

$$\forall k \geq 0,\ E_B^k\varphi \tag{6.1}$$

La introducción de este operador de conocimiento común añade toda una dimensión de complejidad teórica a nuestros formalismos que, por desgracia, en este trabajo solo nos dará tiempo a sugerir de forma superficial – consideramos, por otra parte, que los principales aportes teóricos de este trabajo ya se hicieron en el capítulo anterior. En particular, una de las consecuencias teóricas más interesantes de la introducción de este operador es que el lenguaje $\mathcal{L}_{KC[]}$ ($\mathcal{L}_K$ + Operador anuncio público + Operador conocimiento común) es más expresivo que $\mathcal{L}_{KC}$ ($\mathcal{L}_K$ + Operador conocimiento común); una prueba de este hecho puede encontrarse en van Ditmarsch [1] (teorema 8.48). Esto es interesante porque, en cambio, el lenguaje $\mathcal{L}_K$ es, como vimos (de manera incompleta) en el capítulo 3, igual de expresivo que el lenguaje $\mathcal{L}_{K[]}$ ($\mathcal{L}_K$ + Operador anuncio público). Lo que esto indica es que la aportación que hacen los operadores de anuncio público y de conocimiento común en conjunto es "mayor" que la que hacen por separado. En investigaciones futuras sería interesante estudiar qué es lo que ocurre si se añade también el operador de anuncio público en toda esta mezcla.

Por último, la mayoría de los ejemplos realmente sugerentes que podríamos haber incluido en las secciones anteriores requerían del concepto adicional de conocimiento común, dado que, a menudo, es importante mostrar que un sistema satisface una propiedad que no es reducible a ninguno de los agentes individuales, sino al *conocimiento común* que comparten; de modo que será en esta sección donde presentaremos buena parte de estos ejemplos.

6.2. Sintaxis, semántica y propiedades básicas

Como hemos comentado en la introducción, el plano "teórico" quedará en un plano más bien secundario de este capítulo. Además, el lector ya estará cansado de ver una y otra vez el mismo tipo de construcciones y definiciones, de modo que nos tomamos la licencia de ser algo más escuetos en esta ocasión.

Definición 6.1. ***Sintaxis de los lenguajes*** $\mathcal{L}_{KC}$, $\mathcal{L}_{KC[]}$, $\mathcal{LP}_{KC}$, $\mathcal{LP}_{KC[]}$, $\mathcal{LP}_{KC[]_{Kooi}}$. *Para cualquiera de los lenguajes que hemos considerado a lo largo de este trabajo, el correspondiente lenguaje con conocimiento común es el generado por la BNF resultante de añadir el constructor* $C_B\varphi$ *a cualquiera de las BNFs originales, con* $B \subseteq A$.

□

Antes de presentar la semántica para poder interpretar este nuevo operador, es necesario introducir algunas construcciones teóricas previas.

Definición 6.2. ***Cierre transitivo reflexivo.*** *Sea* S *un conjunto numerable,* A *un conjunto finito no vacío, y, para cada* $a \in A$, *sea* R_a *una relación en el conjunto* S. *Sea* $B \subseteq A$. *Definimos:*

- $R_{E_B} = R_B := \bigcup_{b\in B} R_b$
- $R_{C_B} := R^*_{E_B}$

  Donde:

- *El cierre transitivo de una relación* R *se define como la menor relación* R^+ *tal que:*

  I) $R \subseteq R^+$

  II) *Para todos* $x, y, z \in S$, *se tiene que* $(Rxy \text{ y } Ryz) \Rightarrow Rxz$.

- *Si además imponemos que para todo* $x \in S$ *se tenga* R^+xx, *entonces decimos que es el cierre reflexivo transitivo de* R, *y lo denotamos* R^*.

  En particular, si R *ya es reflexiva, entonces* $R^+ = R^*$.

□

Observación 6.1. Se comprueba fácilmente que una forma de expresar la condición para la validez de la fórmula $E_B\varphi$ es la siguiente:

$$M, s \vDash E_B\varphi \iff \forall t \in S, R_{E_B} st \Rightarrow M, t \vDash \varphi$$

□

La siguiente proposición nos proporciona una forma sencilla de manejar e interpretar el cierre transitivo de una relación.

Proposición 6.1. Sea R una relación sobre un conjunto S. Consideremos las siguientes relaciones auxiliares:

- $R^0xy \Leftrightarrow x = y$
- $R^1xy \Leftrightarrow Rxy$
- $R^kxy \Leftrightarrow$ O bien $R^{k-1}xy$ o bien existe x_1 tal que Rxx_1 y $R^{k-1}x_1y$.

Entonces se tiene:

- R^+xy si y solo si existe $k \geq 1$ tal que R^kxy.
- R^*xy si y solo si existe $k \geq 0$ tal que R^kxy.

□

Definición 6.3. ***Semántica del operador de conocimiento común.*** *En cualquiera de los lenguajes estudiados más el operador C_B, la semántica para interpretar este operador es la siguiente:*

$$M, s \vDash C_B\varphi \Longleftrightarrow \forall t \in S, R_{C_B}st \Rightarrow M, t \vDash \varphi$$

Recordemos que utilizamos la notación $R_{C_B} := R^*_{E_B}$.

□

Observación 6.2. Obsérvese que la proposición 6.1 nos permite reescribir fácilmente la semántica del operador C_B para obtener la expresión de la ecuación 6.1.

□

A continuación presentamos varias propiedades básicas sobre la interacción entre el operador de conocimiento común y el operador de anuncio público. Comenzamos observando que no se satisface la reescritura "natural" que podríamos esperarnos por analogía con el teorema 3.1, a saber, $[\varphi]K_a\psi \leftrightarrow (\varphi \rightarrow K_a[\varphi]\psi)$, y de hecho no es difícil encontrar un contraejemplo. No obstante, sí se tiene la siguiente propiedad para "garantizar" el conocimiento público de una fórmula tras un anuncio público:

Proposición 6.2. Sean $\varphi, \psi, \chi \in \mathcal{LP}_{\mathcal{KC}[]}$. Entonces, si las fórmulas $\chi \rightarrow [\varphi]\psi$ y $(\chi \wedge \varphi) \rightarrow E_B\chi$ son válidas, también lo es $\chi \rightarrow [\varphi]C_B\psi$.

Demostración. Sea M un modelo de Kripke probabilístico, y sea s un estado en dicho modelo. Supongamos que $M, s \vDash \chi$; tenemos que probar entonces que, bajo las condiciones del enunciado, $M, s \vDash [\varphi]C_B\psi$. Si $M, s \nvDash \varphi$, esto siempre es cierto. Supongamos pues que $M, s \vDash \varphi$; tenemos que probar que $M_{|\varphi}, s \vDash C_B\psi$, es decir, dado $s \in S'$, R'^*_Bst implica que $M_{|\varphi}, t \vDash \psi$. Por la proposición 6.1, $R'^*_Bst \Leftrightarrow R'^k_Bst$ para algún k. Procedemos pues por inducción sobre k.

Caso $k = 0$**.** Entonces $s = t$, y $M_{|\varphi}, s \vDash \psi$ se tiene porque $M, s \vDash \chi$, y por la validez de $\chi \rightarrow [\varphi]\psi$.

Hipótesis de inducción. Para $i \leq k$, si $s \in S'$ es un estado en el que se satisface $M, s \vDash \chi$ y $M, s \vDash \varphi$, entonces $R'^i st \Rightarrow M_{|\varphi}, t \vDash \psi$.

Caso $k+1$**.** Sea t tal que $(R'_B)^{k+1} st$. Si $(R'_B)^k st$, entonces tenemos el resultado por hipótesis de inducción. En otro caso, existe $x \in S'$ tal que $R'_B sx$ y $(R'_B)^k xt$. Dado que $M, s \vDash \varphi$ y $M, s \vDash \chi$, y por la validez de la fórmula $(\varphi \wedge \chi) \rightarrow E_B \chi$, tenemos $M, x \vDash \chi$ (pues $R'_B sx \Rightarrow R_B sx$). Por otra parte, dado que $x \in S' = S(\varphi)$, tenemos que $M, x \vDash \varphi$. Aplicando la hipótesis de inducción en M, x, tenemos el resultado para t.

Q.E.D.

□

Corolario 6.1. $[\varphi]\psi$ es válida si y solo si lo es $[\varphi] C_B \psi$.

□

Este último resultado podría parecer contradictorio con la noción de que no se puede proporcionar una traducción de $\mathcal{L}_{KC[]}$ a $\mathcal{L}_{KC}$, pues algunos lectores podrían considerar que esto ya constituye la equivalencia que nos faltaría para tal sistema de reescrituras. No obstante, tener un conjunto de equivalencias de este tipo no es suficiente para que constituyan una traducción adecuada (es decir, que las fórmulas traducidas tomen el mismo valor de verdad para los mismos modelos en los mismos estados); por ejemplo, es posible construir modelos en los que la fórmula $\chi \rightarrow [\varphi]\psi$ es válida pero $\chi \rightarrow [\varphi] C_B \psi$ no lo es.

6.3. Algunos ejemplos

En esta sección estudiaremos algunos ejemplos clásicos que ponen en relieve la potencia conceptual de este operador de conocimiento común. En algunos casos, ofrecemos también "variaciones probabilísticas" de los mismos.

6.3.1. Números Consecutivos

Nuestro primer ejemplo es un problema sencillo de van Ditmarsch [1] que ilustra cómo, incluso en las situaciones más triviales, alcanzar conocimiento común sobre un hecho resulta no ser posible; en particular, esta situación distingue claramente la diferencia entre *conocimiento compartido por un grupo* (operador E_B) y *conocimiento común*.

Ejemplo 6.1. **Números consecutivos.** Dos agentes, Asmodeo (a) y Belfegor (b), se encuentran cara a cara. Cada uno de ellos tiene dibujado un número en su frente, de tal forma que a sabe el número de b y b sabe el número de a, pero ninguno de ellos conoce su propio número (pero sí saben que tienen algún número dibujado). Además, también conocen comunmente que se tienen las siguientes condiciones sobre los números:

- Son naturales mayores que 0.

- Son consecutivos

Una pregunta que podemos hacernos inicialmente es la siguiente: ¿bajo qué circunstancias alguno de los agentes (o ambos) puede deducir el número que tiene en su frente?

Empezemos describiendo la situación como un modelo de Kripke. El conjunto de estados sería $S := \{(n, m) \in \mathbb{N}^2 \mid n = m + 1 \ o \ n = m - 1\}$; esta representación de los estados es autodescriptiva en cuanto a las proposiciones atómicas que se satisfacen en cada uno de ellos, que representaremos como a_n ("el agente a tiene el número n") y b_m ("el agente b tiene el número m"). Las relaciones de accesibilidad pueden caracterizarse de la siguiente forma:

- $R_a(m, n)(m', n) \iff m = m' \ o \ (m = n{+}1 \ y \ m' = n{-}1) \ o \ (m = n{-}1 \ y \ m' = n{+}1)$
- $R_b(m, n)(m, n') \iff n = n' \ o \ (n = m{+}1 \ y \ n' = m{-}1) \ o \ (n = m{-}1 \ y \ n' = m{+}1)$

La figura 6.1 es una representación visual de este modelo. En particular, esta figura nos permite contestar rápidamente a la pregunta que nos hicimos antes: los únicos estados donde algún agente puede conocer su propio número son el $(1, 2)$, donde se tiene $M, (1, 2) \vDash K_b b_2$, y el $(2, 1)$, donde se tiene $M, (2, 1) \vDash K_a a_2$.

En cualquier otro estado, ninguno de los agentes conoce su propio número. Supongamos por ejemplo que nos encontramos en el estado $(3, 2)$; en tal caso, se tiene b_2, pero b considera posible b_4. En particular, se tiene $M, (3, 2) \vDash \neg b_4 \wedge \neg E_{a,b} \neg b_4$. Sí que se tiene $M, (3, 2) \vDash E_{a,b} \neg a_5$: en efecto, en todos los estados accesibles desde $(3, 2)$ tanto para a como para b se satisface la fórmula $\neg a_5$. No obstante, no se tiene $E_{a,b} E_{a,b} \neg a_5$, puesto que tenemos la "cadena" de estados accesibles $(3, 2) \sim_b (3, 4) \sim_a (5, 4)$, por lo que $\neg K_b K_a \neg a_5$; por lo tanto, tenemos $M, (3, 2) \vDash E_{a,b} \neg a_5 \wedge \neg E_{a,b} E_{a,b} \neg a_5$. Razonando de manera similar para números cada vez mayores, tenemos

$$M, (3, 2) \vDash E_{a,b} E_{a,b} \neg b_6 \wedge \neg E_{a,b} E_{a,b} E_{a,b} \neg b_6$$

$$M, (3, 2) \vDash E_{a,b} E_{a,b} E_{a,b} \neg a_7 \wedge \neg E_{a,b} E_{a,b} E_{a,b} E_{a,b} \neg a_7$$

y así sucesivamente. En definitiva, para cualquier número natural n, por inmenso que sea, no podemos afirmar que sea conocimiento común el hecho $\neg a_n$ si n es impar o $\neg b_n$ si n es par (supongamos que es par). Más aún: dado que un razonamiento absolutamente análogo podría hacerse en cualquier otro estado de la "rama superior" del modelo M – es decir, el conjunto $R^*_{a,b}((3, 2))$ – el hecho de que esto no sea conocimiento común *es* conocimiento común,

$$M, (3, 2) \vDash \neg C_{a,b} \neg b_n \wedge C_{a,b} \neg C_{a,b} \neg b_n$$

Esta situación puede resultar un tanto sorprendente, porque, en un principio, podría esperarse que, en el estado $(3, 2)$, ambos agentes pudiesen suponer que ninguno de ellos considera posible que ninguno de ellos considera posible (...) algo tan alejado de la realidad como que el número del agente b es el 666. Es precisamente por esto que

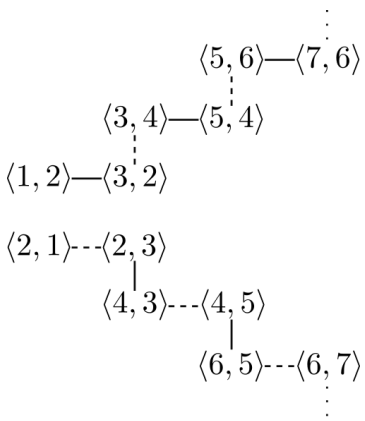

Figura 6.1: Una representación del modelo M en el ejemplo de los números consecutivos. Las líneas continuas representan la relación de accesibilidad del agente a, y las líneas discontinuas representan la relación de accesibilidad del agente b.

este ejemplo es especialmente bueno para mostrar lo restrictivas que son las condiciones para que se tenga conocimiento común de algún hecho.

□

6.3.2. Niños con Barro

La situación de los niños con barro, o de los niños embarrados, es uno de los ejemplos clásicos estudiados en la lógica epistémica. Se trata de un ejemplo sugerente, dado que plantea una situación en la que, tras n repeticiones de una misma frase, un conjunto de agentes logra obtener conocimiento común sobre un hecho que en las $n-1$ repeticiones anteriores desconocían.

Ejemplo 6.2. **Niños con barro.** Una serie de niños, convenientemente llamados $a_1, ..., a_n$, están jugando en un charco de barro fuera de su casa, ensuciándose en el proceso; en particular, algunos de estos niños se han manchado la frente con barro. En cierto momento, escuchan que su padre los llama, de modo que se dirigen de vuelta a su casa; una vez que han llegado, su padre les pide que se coloquen en círculo, de tal forma que cada uno de los niños puede ver si la frente de cualquiera de los otros niños está o no manchada (pero no sabe si está manchada o no su propia frente). Entonces el padre anuncia públicamente la siguiente afirmación:

> *"Al menos uno de vosotros tiene la frente manchada de barro."*

Acto seguido, les hace la siguiente petición:

> *"Que den un paso hacia delante aquellos que sepan si su frente está o no manchada."*

Si nadie da un paso hacia delante, el padre seguirá repitiendo la petición (hasta que alguien dé un paso hacia delante). Se puede probar entonces que, si m de los n niños tienen la frente manchada, entonces los m niños con la frente manchada darán un paso hacia delante tras m repeticiones de la petición.

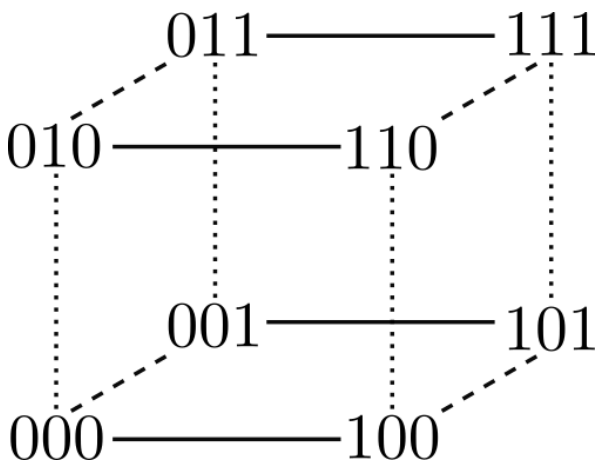

Figura 6.2: Situación inicial del ejemplo de los niños con barro. Las rectas continuas representan la relación de accesibilidad de a_1, las líneas punteadas representan la relación de accesibilidad de a_2, y las líneas discontinuas representan la relación de accesibilidad de a_3.

Supongamos que estamos en el caso particular en el que $n = 3$ y representémoslo con un modelo epistémico M. Consideremos las proposiciones atómicas m_1, m_2, m_3, que expresarían respectivamente el hecho de que el niño a_i correspondiente tiene la frente manchada. Podemos representar los estados posibles con las ternas $A_1A_2A_3$, donde cada A_i puede tomar los valores 0 o 1; de esta forma, si un estado tiene un 1 en la posición i, esto expresaría que en dicho estado se tiene la proposición atómica m_i (por ejemplo: en el estado 101 se tendrían m_1 y m_3, es decir, los niños a_1 y a_3 tendrían la frente manchada). Dos estados son mutuamente accesibles para un niño a_i si y solo si las cifras en las posiciones distintas de i de ambos estados coinciden (por ejemplo: $R_{a_2}101, 111$), pues cada niño sabe si las frentes de todos los demás están manchadas o no, y solo desconoce el estado de su propia frente. La figura 6.2 es una representación gráfica de este modelo.

En primer lugar, el padre anuncia públicamente que "al menos uno de los niños tiene la frente manchada con barro", es decir, la fórmula $m_1 \vee m_2 \vee m_3$. El modelo actualizado consistirá por tanto en el modelo resultante de eliminar el estado 000 del modelo inicial (figura 6.3). Supongamos ahora que nos encontramos en cualquiera de los estados en los que solo uno de los niños tiene la frente manchada; por ejemplo, supongamos que a_3 es el único niño con la frente manchada (estado 001). Entonces se observa rápidamente que $K_{a_3}m_3$, dado que el único estado accesible para a_3 desde 001 en el modelo restringido es el propio 001; una justificación informal de este hecho es que, al observar que sus dos hermanos tienen la frente limpia y teniendo en cuenta la afirmación de su padre, a_3 concluye que el único que puede tener la frente manchada es él mismo. Por lo tanto, ante la primera de las peticiones del padre, a_3 dará un paso hacia delante.

Ahora supongamos que estamos en algún estado en el que dos de los niños tienen la frente manchada (por ejemplo, 011). En tales circunstancias, la misma figura 6.3 nos permite observar rápidamente que ninguno de los tres niños puede estar seguro de la situación de su frente y, por lo tanto, ninguno de ellos dará un paso hacia delante.

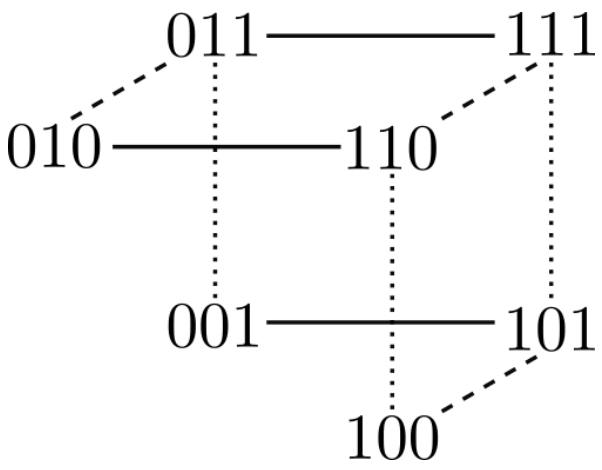

Figura 6.3: Situación tras anunciar $m_1 \vee m_2 \vee m_3$.

Ahora bien, ¿cómo interpretar esto?

Definimos una abreviatura para la siguiente fórmula:

$$alguienconoce := (K_{a_1} m_1 \vee K_{a_1} \neg m_1) \vee (K_{a_2} m_2 \vee K_{a_2} \neg m_2) \vee (K_{a_3} m_3 \vee K_{a_3} \neg m_3)$$

Una forma natural es que el hecho de que ningún niño haya dado un paso hacia adelante ante la petición de su padre puede interpretarse como el anuncio público de la fórmula $\neg alguienconoce$. Curiosamente, el modelo resultante de eliminar de $M_{|m_1 \vee m_2 \vee m_3}$ los estados donde se tiene $alguienconoce$ corresponde al modelo resultante de eliminar de este modelo los estados en los que solo un niño tiene manchada (figura 6.4); y, en dicho modelo, ahora los niños que tienen la frente manchada *sí* lo saben, de modo que tras el siguiente anunciodarían un paso hacia delante.

Por el mismo razonamiento, si nos encontramos en el estado en el que los tres niños tienen la frente manchada 111, las dos primeras peticiones del padre serán "ignoradas", lo que se traduce en dos anuncios públicos de la fórmula $\neg alguienconoce$; puede comprobarse que el modelo que nos queda al final consta únicamente del estado 111, por lo que, con la siguiente petición del padre, los tres niños darían un paso hacia delante.

Además, en cada caso, si definimos B como el conjunto de los niños que tienen la frente manchada en un estado s, es también fácil comprobar que tras $m-1$ iteraciones de la petición del padre, siendo m el número de niños con la frente manchada, cualquiera de los niños en B sabe de antemano que el resto de sus hermanos con la frente manchada también va a dar un paso hacia delante ante la siguiente petición de su padre (en particular, $E_B alguienconoce$), y que cualquiera de los niños en B sabe que cualquiera de los niños en B sabe de antemano que el resto de sus hermanos con la frente manchada también va a dar un paso hacia delante ante la siguiente petición de su padre (en particular, $E_B^2 alguienconoce$), y así sucesivamente; por lo tanto, $M, s \vDash [m_1 \vee m_2 \vee m_3][\neg alguienconoce]^{m-1} C_B alguienconoce$.

□

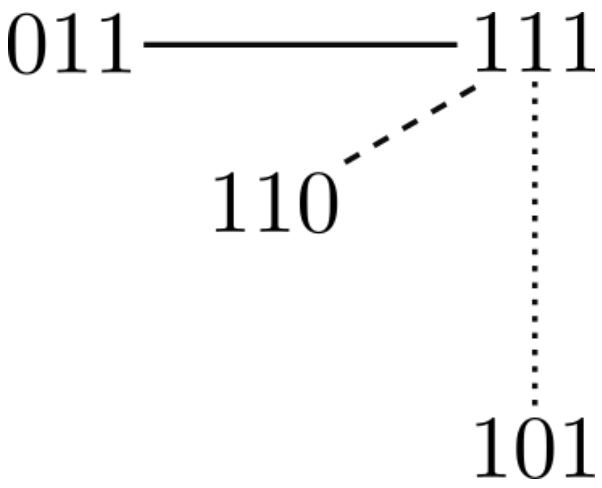

Figura 6.4: Situación tras anunciar ¬*alguienconoce*.

En este trabajo proponemos una variación probabilística de este ejemplo, donde podría ser interesante estudiar la evolución del "sistema" con las sucesivas actualizaciones dependiendo de una serie de parámetros. La situación en este caso es un tanto más abstracta y quizá su descripción no resulte tan pintoresca como en el ejemplo que la precede; no obstante, el paralelismo es fácil de captar.

Ejemplo 6.3. **"Niños con barro" probabilísticos.** Consideremos la siguiente situación: a_1, a_2 y a_3 son tres agentes, cada uno en posesión de dos números n_i y m_i $(i = 1, 2, 3)$ en las siguientes circunstancias:

- El número n_i es conocido para todos los agentes salvo para a_i.
- El número m_i solo es conocido para a_i.
- $n_i, m_i \in \{0, 1\} \forall i = 1, 2, 3$

Supongamos que, en cada caso, cada agente asigna una distribución de probabilidades equiprobable al conjunto de todos los estados que considera posible en el estado actual. Una forma general de formular la familia de cuestiones que podrían interesarnos es: ¿bajo qué circunstancias, o tras cuántos / cuáles anuncios públicos, se tiene conocimiento común sobre algún hecho por parte de todos, o de algún subconjunto de los agentes? Por ejemplo: ¿bajo qué circunstancias pueden conocer comúnmente todos los agentes el siguiente hecho: "La probabilidad de que alguno de los tres agentes posea el par $(1, 1)$ es mayor que P"?

No vamos a explorar en profundidad este ejemplo, pero sí daremos una imagen general de la situación inicial. Para empezar: ¿cómo definiríamos el modelo de Kripke para representar esta situación? Una posibilidad es la siguiente:

- Los átomos son $\{n_i(0), n_i(1), m_i(0), m_i(1) \mid i = 1, 2, 3\}$, donde $n_i(t)$ representa que el valor del número público del agente a_i es t, y $m_i(t)$ representa que el valor del número privado del agente a_i es t.
- Los estados son ternas de pares como $(10)(01)(00)$, donde cada par representa la situación del agente a_i correspondiente: el primer número del par representa su "número público", mientras que el segundo representa su "número privado".

- Dos estados son mutuamente accesibles para el agente a_i si coinciden los números privados del agente a_i y los números públicos de los demás agentes.
- Los espacios muestrales para cada agente a en cada estado s coinciden exactamente con los conjuntos $R_a(s)$. Los conjuntos medibles no triviales son los correspondientes a todos los estados individuales, y todas las uniones posibles de los mismos. Las probabilidades están equidistribuidas entre todos los estados. En particular, el modelo satisface **CONS**, **SDP** y **MEAS**. Obsérvese que no satisface **OBJ**.

Utilizaremos las siguientes abreviaturas:

- $(x,y)_i := n_i(x) \wedge n_i(y)$
- $alguno_{(x,y)} := (x,y)_1 \vee (x,y)_2 \vee (x,y)_3$
- $todos_{(x,y)} := (x,y)_1 \wedge (x,y)_2 \wedge (x,y)_3$

Supongamos que nos encontramos en el estado $s = (01)(10)(10)$. ¿Qué probabilidades asigna cada agente al evento $alguno_{(1,1)}$? Razonémoslo brevemente para el agente a_i:

$$P_{(a_1,s)}(alguno_{(1,1)}) = 1 - P_{(a_1,s)}(\neg alguno_{(1,1)}) = (por\ independencia) =$$
$$1 - P_{(a_1),s}(\neg(1,1)_1)P_{(a_1),s}(\neg(1,1)_2)P_{(a_1),s}(\neg(1,1)_3) = 1 - \left(\frac{1}{2}\right)^3 = \frac{7}{8}$$

Razonando análogamente para a_2 y a_3, obtenemos probabilidades de $\frac{1}{2}$ en ambos casos.

Por otra parte, los tres agentes conocen comunmente que la probabilidad que cualquiera de ellos asigna a este hecho es menor o igual que $\frac{7}{8}$, por el simple hecho de que esta es la probabilidad más alta que algún agente puede asignar a este hecho en cualquier estado en $R^*_A(s)$; en efecto, $(01)(10)(10) \sim_{a_1} (11)(11)(11)$, e intuitivamente puede observarse que en este estado todos los agentes asignarán la probabilidad más alta posible a este evento (que en todos los casos es $\frac{7}{8}$). Es decir:

$$M, s \vDash C_A P_A(alguno_{(1,1)}) \geq \frac{7}{8}$$

donde $P_A(\varphi) \geq Q := \bigwedge_{a \in A} P_a(\varphi) \geq Q$.

Similarmente, los tres agentes conocen comunmente que la probabilidad que cualquiera de ellos asigna a este hecho es mayor o igual que 0, pues $(01)(10)(10) \sim_{a_2} (00)(00)(10) \sim_{a_3} (00)(00)(00)$ y, de nuevo, intuitivamente se observa que en este estado todos los agentes asignarán la probabilidad más baja posible a este evento (que en todos los casos es 0). Una pregunta que podríamos hacernos es: ¿qué anuncios públicos (no triviales) podríamos hacer para mejorar estos umbrales?

Dejamos así este ejemplo, cuyo estudio creemos que puede dar lugar a desarrollos teóricos y prácticos interesantes en el área de la lógica epistémica dinámica, pero que

desgraciadamente no tenemos tiempo de cubrir en este trabajo. De hecho, proponemos brevemente algunas generalizaciones adicionales del mismo:

- Aumentar el número de agentes involucrados, o estudiar para un número de agentes arbitrario
- Hacer que la distribución de probabilidades no sea equiprobable (por ejemplo, permitiendo que el número 1 tenga una probabilidad mayor que el número 0 para algunos agentes, y viceversa para otros).
- Permitir más de dos números posibles además de 0 y 1, o incluso distintos rangos numéricos para distintos agentes.

□

6.3.3. Generales Bizantinos

Para finalizar esta sección, presentamos un ejemplo sencillo que pone de manifiesto el hecho de que la lógica epistémica dinámica no se limita exclusivamente a los lenguajes que se exponen en este trabajo, y que hay "todo un mundo de posibilidades" en cuanto a las propuestas que pueden hacerse para representar situaciones más allá de las que hemos planteado aquí.

El problema de los Generales Bizantinos es un problema clásico en el análisis de sistemas distribuidos, y es una ilustración alegórica de las dificultades que pueden presentarse a la hora de tratar de alcanzar un consenso en este tipo de sistemas. A continuación describimos una de sus variantes básicas.

Ejemplo 6.4. **Generales Bizantinos.** Consideremos la siguiente situación: dos generales a (Ascano) y b (Belisario), del mismo bando, están asediando una ciudad. La ciudad está situada en un valle entre dos montes, y cada uno de los generales (con sus respectivos ejércitos) se encuentra en uno de estos montes. Ambos generales saben que si atacan la ciudad juntos derrotarán fácilmente al enemigo pero, en cambio, si atacan por separado será el enemigo quien los derrotará a ellos. El problema radica, por tanto, en tratar de coordinarse para atacar al mismo tiempo[1].

El general a envía un mensajero al campamento de b con el siguiente mensaje: $m \equiv$ "Propongo que ataquemos mañana a mediodía". Es posible, no obstante, que el mensajero sea atrapado en el trayecto de a a b. Supongamos que el mensajero logra llegar al campamento de b, con lo que se tendría $K_b m$ e incluso $K_b K_a m$: ¿sería sensato para b tomar la decisión de atacar mañana a mediodía? En absoluto, pues b sabe que a no sabe si el mensaje ha llegado ($K_b \neg K_a K_b m$), con lo que b tomaría la decisión prudente de no atacar. Por lo tanto, b decide enviar al mensajero de vuelta a a con el mensaje "He recibido tu mensaje y estoy de acuerdo", que codificamos como $K_b m$. Ahora bien, b tiene el mismo problema que tenía a al principio, dado que tampoco es seguro esta vez que el mensajero logre transmitir el mensaje exitosamente; supongamos una vez más, que el mensajero logra llegar al campamento de a. ¿Sería ahora sensato

[1] No pueden hacer señales de humo: imaginemos, por ejemplo, que está lloviendo.

para a tomar la decisión de atacar mañana a medio día? De nuevo, no, pues a sabe que b no sabe si su mensaje de confirmación $K_b m$ ha llegado ($K_a \neg K_b K_a K_b m$).

¿Cuántos mensajes tienen que enviarse mutuamente ambos generales para poder atacar la ciudad con la certeza de que el otro general también lo hará? Un razonamiento inductivo informal nos permite ver rápidamente que tal "consenso" nunca podrá alcanzarse. Sea $n+1$ el número de mensajes que se ha enviado (de esta forma, el primer mensaje tendrá asociado un valor $n = 0$); denotemos $n = 2k$ si es par y $n = 2k + 1$ si es impar. Para n par, los mensajes son de la forma $(K_a K_b)^k m$ (con $(K_a K_b)^0 m = m$), y para n impar son de la forma $K_b(K_a K_b)^k m$. Supongamos ahora, por ejemplo, que b acaba de recibir el mensaje n (por lo tanto n es par; el caso impar es análogo). Entonces se tiene $K_b(K_a K_b)^k m$, pero no se tiene $(K_a K_b)^{k+1} m$, por lo que $K_b \neg (K_a K_b)^{k+1} m$, y por lo tanto ni b ni a pueden ponerse de acuerdo.

En particular, la condición de consenso que buscamos es capturada por la fórmula con conocimiento común $C_{\{a,b\}}\varphi$, que para todo $k \geq 0$ implica $(K_a K_b)^k m$ y $K_b(K_a K_b)m$.

□

El lector habrá observado que en esta versión "clásica" del problema de los generales bizantinos, hemos hecho alusión a razonamientos que no están formalizados en ninguno de los lenguajes que hemos estudiado en este trabajo. En efecto, los anuncios públicos no son suficientes para representar la situación que tenemos entre manos: al fin y al cabo, los mensajes que se están enviando no son anunciados públicamente, sino enviados de forma privada y con la posibilidad de ser interceptados. En general, este tipo de situaciones forman parte de una teoría más general de *acciones epistémicas*, utilizando la terminología de van Ditmarsch [1].

En relación con esto, y para poner en relieve el hecho de que los lenguajes que hemos expuesto en este trabajo no son de ninguna manera los "definitivos" o los "más adecuados" independientemente de las circunstancias, sugerimos a continuación, también de una manera un tanto informal, otro concepto de *actualización* de un modelo probabilístico que diverge por completo de la propuesta de Kooi y, por lo tanto, también de nuestra propia generalización de la misma. En particular, hacemos esta propuesta en el contexto de los generales bizantinos para mostrar cómo, bajo ciertas condiciones y aceptando ciertas suposiciones, puede considerarse que el problema tiene una solución *aproximada* o *probabilística*. Más concretamente aún, la idea fundamental tras este "enfoque alternativo" para la actualización de un modelo probabilístico tiene sus bases teóricas en la predicción *a posteriori* del parámetro p de una variable tipo Bernouilli según la metodología bayesiana (para más detalles consultar Gelman [15], capítulo 2).

Ejemplo 6.5. **Generales Bizantinos tras un curso de estadística bayesiana.** Consideremos de nuevo la situación del ejemplo anterior, pero añadiendo ahora una serie de suposiciones adicionales. Primero, supongamos que, de no ser atrapado, el mensajero tarda a lo sumo un tiempo T en llegar de un campamento al otro. Los generales sabrán, por tanto, que el mensajero ha sido atrapado si pasa un periodo mayor que $2T$ sin haberse obtenido un mensaje de confirmación del otro general. Consideremos ahora el siguiente "protocolo", que ambos generales habrían acordado previamente:

- Valores iniciales:

- $n = 0$
- $p_0 < \frac{1}{2}$ arbitrario, "probabilidad umbral"
- $p = \frac{1}{n+2} = \frac{1}{2}$

- Protocolo para el general a:

  I) a envía un mensaje m con las instrucciones de ataque a b

  II) Si en un periodo inferior a $2T$ llega una respuesta de la forma $K_b(K_aK_b)^{n/2}m$:

    a) Hacer $n := n + 2$

    b) Hacer $p_a := \frac{1}{n+2}$

    c) Hacer $p_b := \frac{1}{n+1}$

    d) Si $p_a < p_0$:

      1) Si $p_b < p_0$, FINALIZAR PROTOCOLO

      2) En otro caso, enviar $(K_aK_b)^{(n/2)+1}$ y FINALIZAR PROTOCOLO

    e) En otro caso, enviar $(K_aK_b)^{(n/2)+1}$ y volver al paso II

  III) Si en un periodo mayor a $2T$ todavía no ha llegado una respuesta de b, ABORTAR PROTOCOLO

- Protocolo para el general b:

  I) b espera indefinidamente un mensaje m con las instrucciones de ataque de a. Si llega el mensaje:

    a) Hacer $n := n + 1$

    b) Hacer $p_a := \frac{1}{n+2}$

    c) Si $p_b < p_0$, enviar el mensaje K_bm; FINALIZAR PROTOCOLO

    d) En otro caso, enviar el mensaje K_bm y continuar

  II) Si en un periodo inferior a $2T$ llega una respuesta de la forma $K_b(K_aK_b)^{n/2}m$:

    a) Hacer $n := n + 1$

    b) Hacer $p_b := \frac{1}{n+2}$

    c) Hacer $p_a := \frac{1}{n+1}$

    d) Si $p_b < p_0$:

      1) Si $p_a < p_0$, FINALIZAR PROTOCOLO

      2) En otro caso, enviar $K_b(K_aK_b)^{(n-1)/2}$ y FINALIZAR PROTOCOLO

    e) En otro caso, enviar $K_b(K_aK_b)^{(n-1)/2}$ y volver al paso II

  III) Si en un periodo mayor a $2T$ todavía no ha llegado una respuesta de b, ABORTAR PROTOCOLO

□

A continuación ofreceremos una explicación detallada de lo que ocurre en este protocolo, aunque no se trata de algo excesivamente complejo. No obstante, como siempre, consideramos que la comprensión del mismo se hace mucho más sencilla si se tiene claro *lo que se pretende conseguir* con él. Informalmente, este protocolo, si se realiza "hasta el final", garantiza (obviamente) que el general que ha recibido el último mensaje (por ejemplo, b) sepa que lo ha recibido, pero también que el general que lo ha enviado (por ejemplo, a) sepa que la probabilidad de haberlo recibido de b es superior a $1-p_0$; más aún, b también sabe que a sabe que la probabilidad de que b haya recibido el mensaje es superior a $1-p_0$, y a también sabe que la probabilidad de que b conozca el hecho anterior también es superior a $1-p_0$; y así sucesivamente. En la descripción del protocolo anterior, p_a y p_b representan en cada caso la probabilidad que cada general asigna al evento "$I \equiv$ El mensajero es interceptado" en el momento en el que uno de ellos recibe un nuevo mensaje. El hecho de que *ambas* probabilidades tengan que ser menores que el "umbral" para detener el envío de mensajes es crucial para que se cumplan las propiedades que expondremos a continuación; por otra parte, si solo una de estas probabilidades (a saber, la del propio general y no la de su compañero) es la que está por debajo de este umbral, dicho general envía un último mensaje pero no espera respuesta.

Supongamos, por ejemplo, que el general a es el primero en asignar una probabilidad inferior al umbral al evento I. Entonces a envía un último mensaje, sabiendo que la probabilidad de llegue a su destino es mayor que $1-p_0$. Supongamos también que, efectivamente, el mensaje llega a su destino (en este caso el general b); entonces, b también sabe que a sabe que la probabilidad de que su mensaje haya llegado a su destino es mayor que $1-p_0$. Por otra parte, a no *sabe* que el mensaje ha llegado a su destino, pero sí sabe que *si* el mensaje ha llegado a su destino, entonces b sabe que a sabe que la probabilidad de que el mensaje haya llegado a su destino es mayor que $1-p_0$; es decir, a *sabe* que la probabilidad de que b sepa que a sepa que la probabilidad de que el mensaje haya llegado a su destino es mayor que $1-p_0$ es mayor que $1-p_0$. El general b, por su parte, también conoce el hecho anterior, y podemos seguir iterando de esta forma.

Formalmente, esto se traduciría en la validez de las siguientes fórmulas:

$$\begin{gathered}
K_a(P_a(\tilde{K}_{ab}^{(n+1)}m) > 1-p_0) \\
K_bK_a(P_a(\tilde{K}_{ab}^{(n+1)}m) > 1-p_0) \\
K_aP_a((K_bK_a(P_a(\tilde{K}_{ab}^{(n+1)}m) > 1-p_0)) > 1-p_0) \\
K_bK_a(P_a((K_bK_a(P_a(\tilde{K}_{ab}^{(n+1)}m) > 1-p_0)) > 1-p_0)) \\
\vdots \\
(K_bK_aP_a^{>1-p_0})^n m \quad \forall n \in \mathbb{N} \\
K_a(P_a((K_bK_aP_a^{>1-p_0})^n m) > 1-p_0) \quad \forall n \in \mathbb{N}
\end{gathered} \tag{6.2}$$

El razonamiento informal expuesto en el ejemplo 6.4 excluye desde un primer momento la posibilidad de obtener conocimiento común sobre algún hecho que tenga que ser transmitido a través de un mensaje; no obstante, las iteraciones sucesivas de

los operadores ($K_b K_a P_a^{>1-p_0}$) apuntan a posibles "generalizaciones probabilísticas" del mismo, en la línea de la propuesta de Halpern y Fagin [3].

Si el protocolo se realiza *hasta el penúltimo paso*, es decir, uno de los generales (por continuar con el ejemplo anterior, a) llega a FINALIZAR PROTOCOLO pero el otro general ABORTA PROTOCOLO (porque no le llega el último mensaje de confirmación), entonces no son ciertas aquellas de las fórmulas cuyo operador "más externo" es K_b; por lo tanto, el general a tendría un problema, ya que acabaría atacando la ciudad él solo. No obstante, lo que garantiza la parte probabilística del protocolo es precisamente que la probabilidad de que esto ocurra sea considerada "ínfima".

Por otra parte, si alguno de los generales ABORTA PROTOCOLO antes de que tanto p_a como p_b sean menores que p_0, el otro también ABORTA PROTOCOLO, y simplemente decidirían no atacar (y esperar otra oportunidad mejor, o algo así).

De hecho, técnicamente ni siquiera es necesario que el protocolo hubiese sido acordado de antemano: en su primer mensaje, junto con las instrucciones para atacar la ciudad, el general a podría haber incluido también las directrices del protocolo (más aún: si el general b recibe el mensaje y considera que la "probabilidad umbral" es demasiado grande, puede proponer un umbral menor en su primera respuesta). Evidentemente no estamos teniendo en cuenta otras consideraciones, como la de la confidencialidad o la autenticidad de los mensajes, o las posibles estrategias que podría seguir un potencial "interceptor antagónico" que conociese los contenidos de los intercambios; no obstante, creemos que la sencillez de este ejemplo es parte de lo que lo hace sugerente y extrapolable a casos mucho más complejos.

No descartamos la posibilidad de obtener otros protocolos similares que resuelvan este problema de forma todavía más satisfactoria; la fórmula general para la *predicción a posteriori* del parámetro p tras la realización de n experimentos bayesianos independientes sucesivos es $\frac{y+1}{n+2}$, donde y es el número de éxitos y n es el número de realizaciones, de modo que los generales podrían tener en cuenta también el número de mensajes interceptados; simplemente habría que detallar algo más la cuestión de cómo se llegaría a un consenso, en el caso de que el mensajero fuese interceptado, sobre si lo ha sido "en el camino de ida" o "en el camino de vuelta". También podría estudiarse en este sentido la posibilidad de establecer un umbral superior para la "probabilidad de intercepción" a partir de la cual ambos generales se resignarían a simplemente no atacar.

7. Sistemas axiomáticos

7.1. Idea y motivación

La idea de interpretar la verdad de fórmulas en relación a modelos constituye un enfoque relativamente moderno en el desarrollo de la lógica como campo de estudio. Mucho más común, al menos en términos históricos (podemos remontarnos, de nuevo, hasta al menos Aristóteles), es encontrarse con una concepción de los sistemas, teorías y razonamientos lógicos como un conjunto de proposiciones que se obtienen aplicando unas determinadas reglas de deducción a una serie de premisas o *axiomas* – tal concepción es, además, mucho más cercana a la que existe en el "imaginario colectivo"[1].

Que dicha idea sea más antigua no significa que de alguna forma esté pasada de moda: por el contrario, podría decirse que es precisamente a partir del Siglo XX cuando vemos una enorme intensificación cuantitativa, de la mano de un mayor refinamiento conceptual en las herramientas y técnicas utilizadas, en los esfuerzos para tratar de desarrollar y estudiar tales sistemas de *deducción* o *cálculo lógico*, con figuras tan colosales como David Hilbert sentando las bases y la dirección de esta nueva deriva. Se pueden citar muchos motivos para el interés en este tipo de cuestiones, pero la mayoría puede trasladarse al contexto de nuestro trabajo como sigue: si bien el enfoque *semántico* en el que nos hemos centrado aquí nos permite interpretar sin problemas si una determinada fórmula es o no verdadera en un determinado estado de un modelo, no nos proporciona un procedimiento *conveniente* (en términos computacionales, por ejemplo) para saber *qué fórmulas* son verdaderas en dicho modelo, o más aún, en una determinada familia de modelos.

En este capítulo presentamos brevemente un conjunto de sistemas axiomáticos, ya conocidos, para los lenguajes que hemos estudiado, sin proporcionar ejemplos y solo explicaciones muy breves de los mismos; en este sentido, este capítulo podría entenderse más bien como un anexo. Sí que queremos observar, como "hecho curioso", que muchos de los axiomas y reglas de deducción que aparecen aquí tienen su análogo en capítulos anteriores como "fórmulas válidas en una familia de modelos" – por ejemplo, algunas de las propiedades de omnisciencia lógica en $\mathcal{L}_K$ (proposición 2.2). Para $\mathcal{LP}_{\mathcal{K}[]}$ y $\mathcal{LP}_{\mathcal{KC}[]}$, proporcionamos algunas propuestas a modo de conjetura, en aras de confirmarlas o refinar sus aspectos problemáticos a través de investigaciones posteriores.

7.2. Sistemas axiomáticos

Informalmente hablando, un sistema axiomático puede entenderse como una "forma sintáctica" de especificar una lógica (a diferencia de la "forma semántica" que hemos estudiado a lo largo de este trabajo). Consta de un conjunto de "fórmulas básicas",

[1] Si se me permite decir algo en tono personal, puedo atestiguar que, de hecho, esta es la imagen que yo mismo tenía de la lógica antes de empezar a escribir este trabajo.

los axiomas[2], y un conjunto de reglas de inferencia, que permitiría deducir el resto de fórmulas (válidas) a partir de los axiomas dados. Más concretamente:

Definición 7.1. ***Derivación de una fórmula en un sistema axiomático.*** *Sea X una axiomática con axiomas $Ax_1, ..., Ax_n$ y reglas $Ru_1, ..., Ru_k$, donde cada regla Ru_j es de la forma "de $\varphi_1, ..., \varphi_{ar(j)}$, deducir φ"; decimos que $ar(j)$ es la "aridad" de la regla. Una derivación de φ en X es una sucesión finita $\varphi_1, ..., \varphi_m$ de fórmulas tales que:*

- $\varphi_m = \varphi$
- *Para cada φ_i en la sucesión se tiene:*
  - *O bien φ_i es una instancia de uno de los axiomas en X*
  - *O bien es el resultado de aplicar alguna de las reglas Ru_j a una cantidad $ar(j)$ de fórmulas en la sucesión cuyo índice sea menor que i.*

Si φ tiene una derivación en X, escribimos $\vdash_X \varphi$, o simplemente $\vdash \varphi$ si X está claro por el contexto. En tal caso decimos que φ es un teorema de X.

□

Las dimensiones semántica y axiomática de un lenguaje lógico se relacionan a través de los siguientes conceptos:

Definición 7.2. ***Solidez y completitud*** *Sea $\mathcal{L}$ un lenguaje lógico. En lo que sigue, X será un sistema axiomático sobre $\mathcal{L}$, y $\mathcal{X}$ será una clase semántica tal que el lenguaje $\mathcal{L}$ será interpretable sobre ella.*

- *X es completo con respecto a $\mathcal{X}$ para $\mathcal{L}$ si para toda fórmula $\varphi \in \mathcal{L}$, se tiene $\mathcal{X} \vDash \varphi \Rightarrow \vdash_X \varphi$.*
- *X es sólido[3] con respecto a $\mathcal{X}$ para $\mathcal{L}$ si para toda fórmula $\varphi \in \mathcal{L}$, se tiene $\vdash_X \varphi \Rightarrow \mathcal{X} \vDash \varphi$.*

□

Como apunte final sobre estas anotaciones teóricas previas, en el contexto de la lógica modal existen conceptos más restrictivos de solidez y completitud, que básicamente restringen el uso de los operadores modales de necesitación (como nuestro operador epistémico K_a) a la hora de hacer derivaciones lógicas *a partir de premisas* (otro concepto que tampoco hemos estudiado aquí). Como hemos comentado ya, el objetivo de esta sección es proporcionar *lo más básico* para que un lector no familiarizado con estos conceptos pueda sacar alguna conclusión aproximadamente útil de los sistemas axiomáticos que proporcionamos a continuación. No obstante, si el lector desea profundizar sobre este asunto, lo referimos a las partes relevantes de van Ditmarsch

[2]Los axiomas suelen proporcionarse como *especificaciones*, y no como fórmulas concretas. Por ejemplo: si un axioma es $K_a(\varphi \to \psi) \to (K_a\varphi \to K_a\psi)$, esto indica que *cualquier* fórmula de esta forma (con φ y ψ fórmulas arbitrarias) es válida en el sistema axiomático. Por otra parte, las implementaciones concretas de los axiomas serán llamadas *instancias* de los mismos.

[3]A menudo esta propiedad también se conoce con el nombre de *adecuación*.

[1] (para una introducción más completa que la nuestra recomendamos el capítulo 2, sección 2.2.3; para quien *realmente* quiera profundizar, recomendamos los capítulos 7 y 8).

7.2.1. $\mathcal{L}_K$

Definición 7.3. ***Sistema axiomático*** K***.*** *Definimos K como el sistema axiomático constituido por:*

- *Todas las instancias de tautologías proposicionales*
- ***Distribución de la implicación a través de*** K_a

$$K_a(\varphi \to \psi) \to (K_a\varphi \to K_a\psi)$$

- ***Modus ponens.*** *De φ y $\varphi \to \psi$, inferir ψ.*
- ***Necesitación de*** K_a***.*** *De φ inferir $K_a\varphi$.*

□

Definición 7.4. ***Sistema axiomático*** $S5$***.*** *Definimos $S5$ como el sistema axiomático resultante de añadir a K los siguientes axiomas:*

- ***Verdad***

$$K_a\varphi \to \varphi$$

- ***Introspección positiva***

$$K_a\varphi \to K_aK_a\varphi$$

- ***Introspección negativa***

$$\neg K_a\varphi \to K_a\neg K_a\varphi$$

□

Teorema 7.1. *El sistema axiomático K es sólido y completo respecto a la clase semántica $\mathcal{K}$ (es decir, la de todos los modelos de Kripke) para $\mathcal{L}_K$. El sistema axiomático $S5$ es sólido y completo respecto a la clase semántica $\mathcal{S}5$ (modelos epistémicos) para $\mathcal{L}_K$.*

□

7.2.2. $\mathcal{L}_{KC}$

Definición 7.5. ***Sistema axiomático*** $S5C$**.** *Definimos* $S5C$ *como el sistema axiomático resultante de añadir a* $S5$ *los siguientes axiomas y reglas:*

- ***Distribución de la implicación a través de*** C_B

$$C_B(\varphi \to \psi) \to (C_B\varphi \to C_B\psi)$$

- ***Mix***

$$C_B\varphi \to (\varphi \wedge E_BC_B\varphi)$$

- ***Inducción del conocimiento común***

$$C_B(\varphi \to E_B\varphi) \to (\varphi \to C_B\varphi)$$

- ***Necesitación de*** C_B**.** *De* φ*, inferir* $C_B\varphi$*.*

□

Teorema 7.2. *El sistema axiomático* $S5C$ *es sólido y completo respecto a la clase semántica* $\mathcal{S}5$ *para* $\mathcal{L}_{KC}$*.*

□

7.2.3. $\mathcal{L}_{K[]}$

Definición 7.6. ***Sistema axiomático*** PA**.** *Definimos* PA *como el sistema axiomático resultante de añadir a* $S5$ *los siguientes axiomas:*

- ***Permanencia atómica***

$$[\varphi]p \leftrightarrow \varphi \to p$$

- ***Anuncios y conjunción***

$$[\varphi]\,(\psi \wedge \chi) \leftrightarrow ([\varphi]\psi \wedge [\varphi]\chi)$$

- ***Anuncios y negación***

$$[\varphi]\,\neg\psi \leftrightarrow (\varphi \to \neg[\varphi]\psi)$$

- ***Anuncios y conocimiento***

$$[\varphi]\,K_a\psi \leftrightarrow (\varphi \to K_a[\varphi]\psi)$$

- ***Composición de anuncios***

$$[\varphi]\,[\psi]\chi \leftrightarrow [\varphi \wedge [\varphi]\psi]\chi$$

□

Observación 7.1. Obsérvese que los nuevos axiomas de PA se corresponden con las reescrituras proporcionadas en el teorema 3.1.

□

Teorema 7.3. *El sistema axiomático PA es sólido y completo respecto a la clase semántica $\mathcal{S}5$ para $\mathcal{L}_{K[]}$.*

□

7.2.4. $\mathcal{L}_{KC[]}$

Definición 7.7. ***Sistema axiomático*** *PAC. Definimos PAC como el sistema axiomático resultante de añadir las siguientes reglas a PA:*

- ***Necesitación de*** *C_B. De φ inferir $C_B\varphi$.*
- ***Necesitación de*** *$[\psi]$. De φ inferir $[\psi]\varphi$.*
- ***Anuncios y conocimiento común.*** *De $\chi \to [\varphi]\psi$ y $\chi \wedge \varphi \to E_B\chi$, inferir $\chi \to [\varphi]C_B\psi$.*

□

Observación 7.2. Obsérvese que el axioma **Anuncios y conocimiento común** se corresponde con el teorema 6.2.

□

Teorema 7.4. *El sistema axiomático PAC es sólido y completo respecto a la clase semántica $\mathcal{S}5$ para $\mathcal{L}_{KC[]}$.*

□

7.2.5. $\mathcal{LP}_{\mathcal{K}}$

Definición 7.8. ***Sistema axiomático*** *$S5P(MEAS)$. Definimos $S5P(MEAS)$ como el sistema axiomático resultante de añadir a $S5$ los siguientes axiomas y reglas:*

- ***Axiomas para razonar sobre desigualdades lineales*** *(consultar Halpern-Fagin [3])*
- ***No negatividad***
$$P_a(\varphi) \geq 0$$
- ***Evento seguro***
$$P_a(\top) = 1$$
- ***Aditividad***
$$P_a(\varphi \wedge \psi) + P_a(\varphi \wedge \neg\psi) = P_a(\varphi)$$

- ***Distributividad.*** *Si* $\varphi \leftrightarrow \psi$ *es una tautología proposicional, entonces*

$$P_a(\varphi) = P_a(\psi)$$

□

Definición 7.9. ***Sistema axiomático*** $S5P$***.*** *Definimos* $S5P$ *como el sistema axiomático resultante de cambiar en* $S5P(MEAS)$ *el axioma* ***Aditividad*** *por los dos siguientes:*

- ***Principio de inclusión-exclusión***

$$P_a(\varphi_1 \wedge ... \wedge \varphi_k) \geq \sum_{I \subseteq \{1,...,k\}, I \neq \emptyset} (-1)^{\#(I)+1} P_a\left(\bigwedge_{i \in I} \varphi_i\right)$$

- ***Evento imposible***

$$P_a(\bot) = 0$$

□

Teorema 7.5. *El sistema axiomático* $S5P$ *es sólido y completo para la clase semántica* $\mathcal{S}5\mathcal{P}$ *(modelos de Kripke probabilísticos cuya "parte modal" está en* $\mathcal{S}5$*) para* $\mathcal{LP}_{\mathcal{K}}$*. El sistema axiomático* $S5P(MEAS)$ *es sólido y completo respecto a la clase semántica* $\mathcal{S}5\mathcal{P}(MEAS)$ *(ídem pero solo los modelos que satisfacen* ***MEAS****) para* $\mathcal{LP}_{\mathcal{K}}$*.* □

Teorema 7.6. ***Propiedades adicionales en modelos probabilísticos.*** *Sea* **Props** *un subconjunto de* $\{\mathbf{CONS}, \mathbf{OBJ}, \mathbf{UNIF}, \mathbf{SDP}\}$*, y sea* **Axiomas** *el correspondiente subconjunto de* $\{Ax_1, Ax_2, Ax_3, Ax_4\}$*, donde los* Ax_i *son los axiomas que definiremos a continuación. Entonces* $S5P \cup$ **Axiomas** *(resp.* $S5P(MEAS) \cup$ **Axiomas***) constituye una axiomatización sólida y completa respecto a la clase de los modelos epistémicos probabilísticos satisfaciendo las propiedades en* ***Props*** *(resp. satisfaciendo las propiedades en* ***Props*** *y* ***MEAS****).*

- Ax_1

$$K_a\varphi \rightarrow (P_a(K_a\varphi) = 1)$$

- Ax_2

$$(Q_1P_a(\varphi_1) + ... + Q_kP_a(\varphi_k) \geq Q) \rightarrow (Q_1P_b(\varphi_1) + ... + Q_kP_b(\varphi_k) \geq Q)$$

- Ax_3 *Si* φ *es una fórmula de la forma* $Q_1P_a(\varphi_1)+...+Q_kP_a(\varphi_k) \geq Q$ *o la negación de una fórmula de tal forma, entonces*

$$\varphi \rightarrow (P_a(\varphi) = 1)$$

- Ax_4 *Si* φ *es una fórmula de la forma* $Q_1P_a(\varphi_1)+...+Q_kP_a(\varphi_k) \geq Q$ *o la negación de una fórmula de tal forma, entonces*

$$\varphi \rightarrow K_a\varphi$$

□

7.2.6. $\mathcal{LP}_{KC}$

Definición 7.10. ***Sistemas axiomáticos*** $S5CP$ ***y*** $S5CP(MEAS)$**.** *Definimos* $S5CP$ *como el sistema axiomático resultante de añadir a* $S5C$ *los axiomas "nuevos" de* $S5P$*, y* $S5CP(MEAS)$ *como el sistema axiomático resultante de añadir a* $S5C$ *los axiomas "nuevos" de* $S5P(MEAS)$*.*

□

Teorema 7.7. *Los sistemas axiomáticos* $S5CP$ *y* $S5CP(MEAS)$ *son sólidos y completos, respectivamente, respecto a las clases semánticas* $\mathcal{S}5$ *y* $\mathcal{S}5(MEAS)$ *para el lenguaje* $\mathcal{LP}_{KC}$*.*

□

7.2.7. $\mathcal{LP}_{K[]}$ y $\mathcal{LP}_{KC[]}$

En esta última sección, basándonos en intuiciones informales e impresiones sobre las axiomáticas que hemos presentado anteriormente, así como en el estudio más riguroso que hemos hecho a través del resto de nuestro trabajo de la semántica de los lenguajes $\mathcal{LP}_{K[]}$ y $\mathcal{LP}_{KC[]}$, nos atrevemos a hacer algunas sugerencias sobre *posibles* sistemas axiomáticos para estos lenguajes (que sean sólidos y completos). También nos inspiramos parcialmente en la axiomática presentada por Kooi [4] para su propuesta de lógica probabilística con anuncios públicos.

La idea básica de nuestra propuesta es directa: el lector se habrá percatado de que cada nuevo componente (anuncios públicos, conocimiento común, probabilidades) que añadimos al lenguaje $\mathcal{L}_K$ básico es, por lo general, modular en cuanto a los axiomas y reglas de deducción que "implica"; es decir, el sistema axiomático para $\mathcal{L}_{KC}$ es simplemente el sistema axiomático "estándar" de $\mathcal{L}_K$ más los axiomas nuevos requeridos para el operador C_B; el sistema axiomático para $\mathcal{LP}_{\mathcal{K}}$ es simplemente el sistema axiomático "estándar" de $\mathcal{L}_K$ más los axiomas nuevos requeridos para el operador probabilístico; etcétera. No obstante, hay algunas excepciones en esto: el sistema axiomático para $\mathcal{L}_{KC[]}$, como ya hemos observado en varias ocasiones, es "mayor" que la unión de los sistemas axiomáticos para $\mathcal{L}_{KC}$ y $\mathcal{L}_{K[]}$ por separado, porque de alguna manera existe una "interacción" entre el operador de anuncio público y el de conocimiento común. En este sentido, en nuestras propuestas apostaremos que tal tipo de "interacción" no existe entre el operador probabilístico y los demás operadores, y que los sistemas axiomáticos correspondientes se obtienen simplemente "combinando" los que ya tenemos, modificando algunos detalles.

Por otra parte, el sistema axiomático que proponemos será para modelos satisfaciendo **MEAS-D**, dado que de lo contrario no podemos asegurar que se tenga la equivalencia que presentamos en la proposición 5.7, y que vuelve a aparecer aquí como uno de los axiomas de nuestro sistema. De antemano, podemos afirmar también que uno de los axiomas típicos para la lógica con probabilidades tampoco se satisface en nuestro sistema: en efecto, en un estado donde la asignación de probabilidades sea la "nula" para algún agente, no se tendrá el axioma de **Evento seguro** (pues en este caso

$P_a(\top) = 0$). No obstante, creemos tener un arreglo sencillo para sustituir este axioma por otro casi idéntico.

Propuesta 7.1. **Sistemas axiomáticos** $PAP(MEASD)$ **y** $PACP(MEASD)$**.** Definimos $PAP(MEASD)$ como el sistema axiomático resultante de añadir al sistema axiomático PA los siguientes axiomas y reglas de los sistemas axiomáticos $S5P(MEAS)$ y $S5P$:

- **Axiomas para razonar sobre desigualdades lineales**
- **No negatividad**
- **Aditividad**
- **Distributividad**
- **Evento imposible**[4]

  Y los siguientes axiomas nuevos:

- **Evento seguro en condiciones plausibles**

$$Pr_a(\varphi) > 0 \to Pr_a(\top) = 1$$

- **Anuncios públicos y probabilidad I**

$$(Pr_a(\psi) > 0) \to \left(\left([\psi]\sum_{i=1}^{n} Q_i Pr_a(\varphi_i) \geq Q\right) \leftrightarrow \left(\sum_{i=1}^{n} Q_i Pr_a(\varphi \wedge [\varphi]\varphi_i) \geq Q\ Pr_a(\varphi)\right)\right)$$

- **Anuncios públicos y probabilidad II**

$$(Pr_a(\psi) = 0) \to \left(\left([\psi]\sum_{i=1}^{n} Q_i Pr_a(\varphi_i) \geq Q\right) \leftrightarrow (Pr_a(\bot) \geq Q)\right)$$

El sistema axiomático $PACP(MEASD)$ se definiría análogamente pero partiendo del sistema axiomático PAC en lugar de PA.

□

Nuestra conjetura es que estos sistemas axiomáticos podrían ser sólidos y completos respecto a la clase semántica $\mathcal{S}5(MEASD)$ para los lenguajes $\mathcal{LP}_{\mathcal{K}[]}$ y $\mathcal{LP}_{\mathcal{KC}[]}$, respectivamente, aunque no lo afirmamos con demasiada convicción; no obstante, sí creemos que, si este sistema axiomático falla en algo, un sistema axiomático adecuado no debería alejarse demasiado de nuestra propuesta. Obsérvese nuestro "parche" para el axioma **Evento seguro**: dado que las situaciones en las que $P_a(\top) = 0$ son aquellas en las que *cualquier* proposición tiene probabilidad 0, basta con que haya *alguna* proposición con probabilidad no nula para que se tenga $P_a(\top) = 1$. En particular para $PACP(MEASD)$, hay que estudiar cómo interacciona esta propiedad con el axioma **Necesitación de** $[\psi]$ de PAC, y es probable que sea necesario modificar también este axioma.

[4] Es posible que sea redundante.

8. Conclusiones y posibles líneas de investigación

Originalmente no teníamos en mente un objetivo especialmente concreto y bien definido para nuestro trabajo, y, en un principio, el objetivo provisional era simplemente el de realizar una investigación sobre (algunos de) los diversos sistemas formales derivados de $\mathcal{L}_K$ que se han propuesto en el campo de la lógica epistémica. En este sentido, la dirección que ha ido tomando el trabajo se ha desarrollado de forma paulatina y orgánica, y podría decirse incluso que ha llegado a "tomar vida propia", empujándonos precisamente a dotarlo de la forma que "debía tener" desde un principio sin que nosotros lo supieramos.

Consideramos que esto es bueno, y, sin ánimos de vanagloriarnos, que es una cualidad indicativa de que algo es "buena ciencia". En efecto, ¿cuál es el objetivo de la ciencia? ¿Simplemente elegir una "parcela" al azar de la realidad y tratar de buscar en ella patrones aleatorios, o inventar sobre ella reglas arbitrarias? Si no puede decirse que "la realidad" (tanto la "exterior" como la "interior", si es que realmente es legítimo hacer tal diferenciación) sea algo infinito, inabarcable e insondable, puede decirse, al menos, que a efectos prácticos lo es para la mente humana (al menos en su forma actual), y aunque tratemos de negarnos o resistirnos activamente a ello, nuestra "intuición" o "instinto de lo que es importante" juega un papel fundamental en determinar qué es lo que decidimos investigar y qué es lo que no, así como la forma de la que desarrollamos esta investigación. Es en este sentido que nos parece que si la sensación con la que el investigador concluye su estudio de unos determinados datos, formalismos, o fenómenos de cualquier otra clase es la de que las conclusiones a las que se ha llegado se corresponden con un desarrollo "natural" de su entendimiento de las cosas, esto, si bien no es un signo definitivo de su valor como ciencia, sí es, como mínimo, una buena señal[1].

Regresando a los mucho más concretos términos del campo de estudio que concierne a este trabajo, queremos empezar observando que hemos podido cumplir la mayor parte de los objetivos que nos hemos llegado a proponer en algún punto. Nuestro aporte principal, como ya hemos observado en varias ocasiones a lo largo del trabajo, ha sido nuestra propuesta para unificar la "variante estándar" de los anuncios públicos en la lógica epistémica dinámica, tal y como aparece en van Ditmarsch [1], con una variante también bastante estándar de la lógica epistémica probabilística, tal y como aparece en Halpern y Fagin [3]. Kooi [4] ya había presentado algunos desarrollos en esta dirección, pero su propuesta, además de poner importantes restricciones sobre el tipo de modelos probabilísticos que se podían considerar, se ajusta más bien a un enfoque *doxástico* que a uno *epistémico* (dejamos de lado la discusión sobre si el enfoque más apropiado

[1] Somos conscientes de que lo más probable es que esta sensación sea simplemente una percepción psicológica, pero esto no la hace menos importante: al fin y al cabo, no es sino nuestro juicio humano *subjetivo* (o, como mucho, el agregado de diversos juicios subjetivos – y todavía más, ¿qué juicio es el encargado de realizar dicha *agregación*?) el que, en última instancia, tiene la potestad de determinar *qué ciencia es valiosa y qué ciencia no lo es.*

para estudiar razonamientos probabilísticos es el doxástico o el epistémico).

En este sentido, nuestra propuesta generaliza por una parte la clase de modelos probabilísticos sobre la que es aplicable, y, por otra, se adecúa más fielmente a los principios de la lógica epistémica. No obstante, esto no significa que nuestra propuesta *tenga* que adecuarse a los principios de la lógica epistémica; más concretamente, lo que queremos decir es que, en principio, no vemos dificultades para introducir variaciones en el concepto de actualización que hemos definido de forma que, por ejemplo, se asemeje más a la propuesta original de Kooi, sin llegar a perder su generalidad en cuanto a la clase de modelos en la que "funciona".

Una pregunta de carácter técnico que algunos lectores podrían haberse hecho sobre nuestra propuesta es la siguiente: *¿Qué necesidad hay de exigir aditividad numerable a la función de probabilidad actualizada, $P'_{(a,s)}$, si en el lenguaje solo tendremos que interpretar fórmulas con cantidades finitas de términos?* La respuesta más inmediata es que, si bien el lenguaje no permite escribir fórmulas como $\sum_{i=1}^{\infty} P_a(\varphi_i) \geq Q$, sí puede ocurrir que una fórmula φ se corresponda con un conjunto $S(\varphi)$ que, a su vez, pueda reescribirse como una unión infinita de conjuntos disjuntos $S(\psi_1), ..., S(\psi_k), ...$, lo cual podría introducir ciertas inconsistencias de no exigirse aditividad numerable. No obstante, una respuesta más visionaria podría ser que se ha exigido esto precisamente con las perspectivas de poder ofrecer una generalización del lenguaje que admita *también* sumatorios infinitos de probabilidades.

Uno de nuestros mayores dilemas a la hora de plantear nuestra propuesta para la actualización de modelos probabilísticos ha provenido del tratamiento de las "actualizaciones nulas". Esto tiene que ver, principalmente, con la siguiente cita de Kooi [4]:

> *"There are some approaches in probability theory for updating with sentences that have probability zero. The most common one is to leave it undefined. (...) Another approach is to assign probability zero to everything. (...) This would seem to go against the laws of modal logic; after learning a sentence with probability zero, even the truth would be assigned probability zero. (...) So both choices would make it difficult to provide a complete proof system."*

En nuestro trabajo no nos hemos preocupado excesivamente sobre la cuestión de los sistemas axiomáticos. No obstante, esto no significa que no los hayamos tenido en cuenta de manera subliminal durante la mayor parte de la escritura del mismo; en efecto, este es el motivo por el que hemos tratado de introducir nuestro "parche" en nuestra propuesta especulativa 7.1 para la axiomática de $\mathcal{LP}_{\mathcal{K}[]}$. Consideramos enteramente posible que este parche no sea suficiente para arreglar los problemas derivados de tener situaciones con $P_a(\top) = 0$; por este motivo, hemos considerado, ya a posteriori, algunas modificaciones a nuestra propuesta original. Una de las posibilidades alternativas que hemos considerado es la siguiente:

$$
\begin{aligned}
S' &:= S(\varphi)\\
R'_a &:= R_a \cap (S' \times S')\\
V' &:= V \cap S'\\
S'_{(a,s)} &:= \begin{cases} S_{(a,s)}(\varphi) & si\ Pr^*_{(a,s)}(\varphi) > 0\\ S' & e.o.c. \end{cases}\\
\mathfrak{A}'_{(a,s)} &:= \begin{cases} \{E(\varphi) \mid E \in \mathfrak{A}_{(a,s)}\} & si\ Pr^*_{(a,s)}(\varphi) > 0\\ \{\emptyset, S'\} & e.o.c \end{cases}\\
Pr'_{(a,s)}(E) &:= \begin{cases} 1 & si\ E = S'_{(a,s)}\\ 0 & si\ E = \emptyset\\ \frac{Pr^*_{(a,s)}(E)}{Pr^*_{(a,s)}(\varphi)} & e.o.c. \end{cases}
\end{aligned}
$$

Es posible que esta propuesta sea más consistente de cara a tratar de proporcionar un sistema axiomático, y, por otra parte, también nos parece que su interpretación filosófica es bastante satisfactoria ("ante el anuncio de un hecho de probabilidad 0, cualquier hecho tiene probabilidad 0, salvo aquellos hechos que caractericen el conjunto definido por todos los estados concebibles (incluso imposibles) del mundo"). Por otra parte, también creemos que la mayoría de los resultados teóricos que hemos proporcionado para nuesta propuesta de actualización son fácilmente traducibles a esta otra; en cualquier caso, dejamos estos aspectos en manos de investigaciones futuras.

Otra de las posibles líneas de investigación tiene que ver con tratar de encontrar conjuntos de condiciones suficientes menos restrictivas, o simplemente diferentes, de las ofrecidas en el teorema 5.2 para garantizar que se satisfaga **MEAS-D** en un modelo de Kripke probabilístico (de hecho, ya sería interesante de por sí tratar de encontrar otras caracterizaciones de la propiedad **MEAS** al márgen de aspectos dinámicos). Una de las propuestas alternativas que hemos llegado a intuir, pero que no hemos llegado a comprobar, es que el conjunto $\{\mathbf{MEAS}, \mathbf{SIGNIF}\}$ constituye también condiciones suficientes para lapropiedad **MEAS-D**; animamos a otros investigadores a que traten de comprobar si esto es cierto y, en el caso de que no lo sea, de encontrar los "arreglos" necesarios.

Muchas otras líneas de investigación ya se han insinuado a lo largo del propio trabajo: nuestro ejemplo final en el capítulo 6, sin ir más lejos, propone informalmente un concepto de *actualización* de una probabilidad en un modelo que en algunas situaciones resultaría más natural, basándonos en los principios de la estimación de parámetros en la inferencia bayesiana. En el mismo ejemplo, mencionamos también la existencia de propuestas probabilísticas para generalizar el operador de conocimiento común (como la que aparece en Halpern y Fagin [3]); sería sin lugar a dudas interesante estudiar los efectos de la introducción de algunas de estas propuestas en nuestro lenguaje $\mathcal{LP}_{KC}$.

Finalmente, y aunque durante el resto del trabajo no hemos hecho nunca énfasis en esta cuestión (salvo en una nota a pie de página), consideramos que una de las limitaciones más importantes de las propuestas que existen en el área de la lógica epistémica dinámica, o al menos de las que nosotros hemos estudiado y presentado en

este trabajo, es que se excluye desde un principio la posibilidad de que algún agente *mienta*. Es cierto que esta cuestión es difícil de tratar, dado que, de nuevo, "a partir de una mentira puede deducirse cualquier cosa" (aunque en la literatura ya existen propuestas en este ámbito, véase [16]). No obstante, consideramos que el enfoque probabilístico proporciona precisamente una salida en este sentido: en efecto, como se ha visto en los correspondientes capítulos, es muy distinto (al menos en los términos de nuestro lenguaje) que algo tenga *probabilidad nula* a que sea *falso* o *imposible*; por este motivo, consideramos que una forma prometedora de tratar esta cuestión podría ser precisamente la de expresar el *grado de confianza* en que un agente sea *veraz* en términos probabilísticos, y la actualización de estos grados de confianza podría depender de una "función de utilidad de mentir frente a decir la verdad"; en cualquier caso, desarrollar una propuesta de este tipo requeriría por sí mismo de una monografía aparte.

Bibliografía

[1] Hans van Ditmarsch, Wiebe van der Hoek, y Barteld Kooi. *Dynamic Epistemic Logic*. Springer, 2008.

[2] Johan van Benthem. *Modal Logic for Open Minds*. Center for the study of language and information, 2010.

[3] Ronald Fagin y Joseph Y. Halpern. Reasoning about knowledge and probability. *Journal of the Association for Computing Machinery*, 41(2):340–367, 1994.

[4] Barteld P. Kooi. Probabilistic dynamic epistemic logic. *Journal of Logic, Language and Information*, 12:381–408, 2003.

[5] George G. Roussas. *An Introduction to Probability and Statistical Inference*. Academic Press, 2003.

[6] Mordechai Ben-Ari. *Mathematical Logic for Computer Science*. Springer, 1993.

[7] Stanford Encyclopedia of Philosophy. Modal logic. Disponible en https://plato.stanford.edu/entries/logic-modal/, (2023/04/29).

[8] Stanford Encyclopedia of Philosophy. Intensional logic. Disponible en https://plato.stanford.edu/entries/logic-intensional/, (2023/04/28).

[9] Stanford Encyclopedia of Philosophy. Medieval theories of modality. Disponible en https://plato.stanford.edu/entries/modality-medieval/, (2023/06/15).

[10] Immanuel Kant. *Crítica de la Razón Pura*. Taurus, 2005. Una traducción al español de Pedro Ribas. Con fines comparativos, incluye varias de las ediciones originales y de algunas notas al pie de página.

[11] Edmund L. Gettier. Is justified true belief knowledge? *Analysis*, 23:121–123, 1963. Texto disponible en la página web http://www.ditext.com/gettier/gettier.html (2023/06/01).

[12] Joseph Y. Halpern. Lexicographic probability, conditional probability, and nonstandard probability, 2009. Disponible en https://arxiv.org/abs/cs/0306106v2 (2023/06/17).

[13] Unifying zero-knowledge proofs of knowledge, 2009. Disponible en https://crypto.ethz.ch/publications/files/Maurer09.pdf.

[14] Jean-Jacques Quisquater, Louis Guillou, y Tom Berson. How to explain zero-knowledge protocols to your children. In *Advances in Cryptology – CRYPTO '89: Proceedings. Lecture Notes in Computer Science*, volume 435, pages 628–631, 1990.

[15] Andrew Gelman, John B. Carlin, Hal S. Stern, y Donald B. Rubin. *Bayesian Data Analysis*. Columbia University, 1995. Disponible en http://www.stat.columbia.edu/~gelman/book/.

[16] Hans van Ditmarsch. Dynamics of lying, 2011. Disponible en https://personal.us.es/hvd/fpubs/f3KRA_lyingNancy.pdf (07-11-2023).