UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN FILOSOFÍA DE LA CIENCIA
Facultad de Ciencias
Instituto de Investigaciones Filosóficas
Facultad de Filosofía y Letras

EL CONOCIMIENTO EN EL SUDOKU:
UN EJEMPLO DE EPISTEMOLOGÍA FORMAL

T E S I S

QUE PARA OBTENER EL GRADO DE:
MAESTRA EN FILOSOFÍA DE LA CIENCIA

P R E S E N T A
MELISA GUTIÉRREZ VIVANCO

TUTOR:
Dr. FRANCISCO HERNÁNDEZ QUIROZ
Facultad de Ciencias, UNAM.

UNAM POSGRADO
Filosofía de la Ciencia

MÉXICO, D. F., NOVIEMBRE DE 2012

Agradecimientos

Muchas gracias a mi familia que siempre ha sido tan amorosa y ha estado a mi lado cuidándome y apoyándome. Gracias Kat, abuelita y tío Mario.

Agradezco profundamente a la UNAM; es un orgullo pertenecer a una universidad que además de proveer enseñanza de primer nivel, fomenta valores como la cultura, el arte, la diversidad, la pluralidad y la equidad. Gracias al Instituto de Investigaciones Filosóficas y en particular al Programa de Estudiantes Asociados. Gracias por proporcionarme condiciones (una oficina, una biblioteca, un ambiente, etc.) propicias para desarrollarme dentro de un ambiente filosófico. Agradezco también a CONACYT (sin dejar de ser crítica ante las medidas gubernamentales de reducción en materia de ciencia, tecnología, educación e investigación) por los dos años de beca que me otorgaron. El apoyo que recibí durante mi maestría fue fundamental para llevar a cabo mis estudios.

Gracias a mi tutor Francisco Hernández Quiroz. Gracias por apoyarme desde el principio. Nunca olvidaré que sin conocerme, estuviste dispuesto a ayudarme y a enfrentar un proceso de admisión partiendo desde cero. Gracias por las asesorías, por creer en mí tanto como lo has hecho, por las lecciones de lógica y de gramática, y por fomentar una relación tan cordial como la que mantuvimos durante estos dos años.

Agradezco a mis sinodales. Gracias a Ivonne Pallares por el apoyo que me dio, por la prontitud de sus respuestas y por la calidez que se percibía hasta por correo electrónico. Gracias a Max Fernández, no sólo por haber leído la tesis y ayudarme con sus valiosos comentarios, sino también por haberme dado la clase que quizá sea la que más me ha gustado en la vida (esa clase determinó lo que hoy es mi futuro académico inmediato). Agradezco a Axel Barceló por sus comentarios. En particular por aquella idea que me dio y que me sacó de un estancamiento de semanas. Quiero agradecer muy especialmente a Ángeles Eraña: mi maestra y lectora. Gracias Ángeles por todo, gracias por tu clase, gracias por el apoyo para entrar al programa de estudiantes asociados, gracias por tus comentarios de la tesis. Pero sobre todo, gracias por tu cariño y por las palabras de aliento. Te admiro y te quiero profundamente.

Gracias a mis queridos amigos, Mónica, Miguel, Tatiana, Diego Cali y Aurora por todos esos pasteles (hicieron más fácil el trayecto). Especialmente a

Alfonso Anaya por ser la única persona en el mundo que entiende que nunca ¡nunca! es suficiente queso Philadelphia. ¡Los quiero!

Gracias a todos mis profesores de la maestría. Especialmente a Maricruz Galván quien me enseñó que la filosofía se trata de tomarse un momento para pensar antes de hablar y no de proceder impulsivamente.

Agradezco también a mis profesores de Ciencias por haberme apoyado incluso cuando les dije que quería venir a Filosofía. Gracias por ser siempre incluyentes y fomentar el trato respetuoso, cordial e incluso amistoso con los alumnos. Gracias también a mis alumnos y a mis compañeros por dejarme seguir formando parte de una comunidad tan maravillosa.

Gracias a la chulísima War, a Pedro Miguel, a Marina Taibo, a Penagos, a Palomita, a la hermosa Afrodita y a todos mis amigos de la campaña. Esta vez no lo conseguimos (U.u) pero estoy segura de que de alguna manera, lograremos hacer de éste un mejor país.

Índice general

Introducción

La eterna tensión

Una razón por la cual durante siglos la humanidad ha mostrado preocupación por estudiar el conocimiento, es que éste juega un papel muy importante en el debate, las decisiones, y las acciones del día a día de la vida, tanto de la cotidiana, como de la científica. El estudio sistemático y detallado del conocimiento, sus criterios de adquisición, y sus límites y modos de justificación, es con frecuencia lo que se conoce como epistemología (Hendricks 2006). Pero en la epistemología como en muchos otros ámbitos de la filosofía se ha generado desde hace tiempo una especie de tensión entre los enfoques de tipo formal y aquellos más concentrados en las cuestiones de contenido. Sin embargo, esta tensión ha sido particularmente fructífera en la historia de la filosofía. Por ejemplo, después de que el positivismo lógico aportara grandes avances en materia de teorías de la verdad y la justificación (entre otras), la búsqueda por derrotar sus intentos de encontrar un fundamento último para el conocimiento (Ayer 1959) dio lugar a un nuevo paradigma en epistemología en el que surgieron una gran variedad de propuestas alternativas a las dadas, así como nuevos conceptos tales como *la carga teórica de la observación* (Hanson 1958), *la inconmensurabilidad* de teorías (Kuhn 1962), etc. Pero éste no es el único ejemplo en el que esta tensión ha resultado una especie de catalizador que contribuye a la elaboración teorías filosóficas. Después de que en 1948 Hempel publicara su modelo Nomológico-Deductivo para la explicación científica, múltiples críticas le fueron presentadas (e.g. Ruben 1990).

Una de las críticas hechas a Hempel por parte de algunos filósofos de la ciencia tiene que ver con la presuposición de leyes de la naturaleza, pero muchas otras tienen que ver con la mencionada tensión entre forma y contenido; por ejemplo, para Salmon, de ninguna manera, el modelo de explicación científica puede ser dado en términos de un argumento lógico, pues esto genera problemas concernientes a la relevancia de las hipótesis respecto a la conclusión (Salmon 1984). Para resolver el problema originado por el modelo formal de Hempel, los filósofos presentaron distintas alternativas. Algunas de estas alternativas se desarrollaron dentro de marcos teóricos principalmente formales (por ejemplo, utilizando lógicas de la relevancia), y otras, incluso pretendían mostrar la inconveniencia de una metodología de tipo formal (eg. Riben 1990).

El objetivo del presente trabajo es ofrecer un panorama del estudio formal de ciertos problemas filosóficos, en particular en el contexto del conocimiento. Mostraremos la eficacia y conveniencia del uso de sistemas formales en epistemología, buscando así dar una respuesta parcial favorable a la pregunta sobre qué podemos decir acerca de los tratamientos de epistemología a partir del desarrollo en materia de sistemas formales que se ha dado en las últimas décadas. Para concluir, a manera de ejemplo, presentaremos la primera parte de un proyecto cuyo propósito es modelar el fenómeno del conocimiento en el sudoku. Esta parte del proyecto es pertinente por que distingue el aspecto computacional del aspecto epistémico en el juego.

Con el fin de lograr los objetivos mencionados, en el capítulo 1 presentamos dos cuestiones filosóficas para motivar el estudio por la vía formal de problemas epistemológicos. Estas cuestiones son: i) la necesidad de responder al escéptico -vía el problema del error-, y ii) el problema de la justificación -vía una propuesta fiabilista-. Los capítulos 2 y 3 tienen como objetivo principal mostrar el funcionamiento de dos de las teorías más paradigmáticas correspondientes al área de epistemología formal. En el capìtulo 2 presentamos los principios de la *lógica epistémica*, destacando cómo nociones típicas de conocimiento pueden expresarse efectivamente en un lenguaje formal. La lógica epistémica puede ubicarse en el contexto de la búsqueda por una definición satisfactoria de conocimiento. En particular como un recurso para defender una concepción fiabilista de conocimiento. De esta manera, el capítulo encuentra una motivación en el inciso ii) mencionado anteriormente. El capítulo 3 está dedicado al proceso conocido como *revisión de creencias*. Esta teoría encuentra motivación en el inciso i): una posible respuesta al *problema del error* planteado por el escéptico cartesiano es que si bien, es clara la existencia de la posibilidad del error ante las afirmaciones que pretendemos catalogar como conocimiento, la epistemología formal, en particular el proceso de revisión de creencias, brinda una alternativa para probar que los mundos en los cuales cometemos errores (como aquél en el que un genio maligno me engaña) no son relevantes para nuestro conocimiento. Otra respuesta posible a este problema es que una función de la revisión de creencias es detectar los “errores”dentro de un sistema de creencias, y una vez que éstos sean localizados, ejecutar procesos que establezcan estrategias para resolver los conflictos derivados de dichos errores para así lograr conocimiento. Los métodos de revisión de creencias nos permiten modelar los cambios en un conjunto de proposiciones (creencias) ante la aparición de nueva información. Estos procesos logran capturar cómo es que ante la aparición de nueva información, se modifican los estados y las actitudes epistémicas. En el presente trabajo nos enfocaremos en el modelo paradigmático AGM y en la versión modal. El capítulo 4 es el más importante de todos. En este capítulo vamos a construir nuestro propio modelo formal como parte de la solución de un problema epistemológico. El problema es el de describir el conocimiento en el sudoku. Nuestro propósito es construir un lenguaje que nos permita distinguir entre las proposiciones que describen estados epiestémicos y aquellas que distinguen estados de cosas dentro del juego. Este ejemplo puede dar luz a diferentes contextos del conocimiento, pues el sudoku es un juego cuyo éxito depende, entre otras cosas, de la interacción entre nuestro conocimiento (interno al juego) y nuestras creencias acerca de él. Es muy importante aclarar que el modelo no pretende servir como mecanismo para llegar a la solución, sino describir interacciones entre conocimiento y creencias, así como buscar una justificación (en el sentido fiabilista) para poder llamar “conocimiento” a las proposiciones verdaderas de las que podemos tener certeza en un juego de sudoku. Antes de entrar en materia, daremos una breve introducción sobre el juego del sudoku con el objetivo de que a lo largo del trabajo, al lector que no esté familiarizado con el juego se le faciliten las referencias al mismo.

El sudoku

El sudoku es un juego de tipo rompecabezas que se ha popularizado durante los últimos años a partir de su internacionalización en 2005. El objetivo del sudoku es rellenar, respetando ciertas restricciones, una cuadrícula de 9×9 casillas, dividida en subcuadrículas de 3×3 con las cifras del 1 al 9 partiendo de algunos números ya dispuestos en algunas de las casillas. Las restricciones consisten en que no pueden repetirse cifras en el mismo renglón, columna o subcuadrícula. A los números que se encuentran ya dispuestos al inicio del juego, les llamamos “pistas iniciales”, y dado que estas pistas corresponden a las casillas en las que se encuentran colocadas, nuestra creencia acerca de que cada uno de los números mostrados corresponden a la casilla en la que se encuentran es, junto con las reglas, el conocimiento inicial del juego. La idea es que a partir de este conocimiento, podemos ir construyendo razonamientos que nos permiten conocer la correspondencia entre cada uno de los números y las casillas restantes en el juego. Pero con sólo poseer el conocimiento inicial, es muy difícil que tengamos certeza sobre el resto de las correspondencias casilla-cifra; el grado de dificultad dependerá del número de pistas de las que dispongamos. Así pues, tenemos que ir formando conjeturas acerca de cuál es el número que le corresponde a cada una de las casillas. Si estas conjeturas se forman de la manera correcta, lo ideal es que sean creencias que se van reforzando hasta convertirse en conocimiento (una vez que tenemos certeza sobre la correspondencia casilla-cifra). Por supuesto que éste no siempre será el caso, pues es posible que nos equivoquemos y una o más conjeturas sean incorrectas. Lo importante es que mientras transcurre el juego, vamos adquiriendo conocimiento. Ya sea porque las conjeturas han sido correctas y las creencias se refuerzan cada vez más hasta convertirse en conocimiento, o porque una vez que nos equivocamos y revisamos las creencias que nos condujeron hasta el error sabemos que existe una o varias asignaciones que no son posibles dentro del juego.

La estrategia para modelar la interacción entre las creencias (incluidas algunas que son falsas) y el conocimiento que se va adquiriendo en un juego de sudoku será definir una lógica modal similar a la lógica epistémica; vamos a valernos de diferentes recursos como lógica multimodal, lógicas multivaluadas y por supuesto, lógica epistémica, para construir un sistema donde sean válidos los axiomas que tradicionalmete han sido aceptados en los modelos de conocimiento (justamente, por que al parecer, son los que mejor coinciden con nuestras intuiciones acerca de lo que el conocimiento debería ser).

Capítulo 1

Epistemología formal

Como parte de la motivación filosófica del presente trabajo, vamos a tomar como punto de partida dos vertientes desde las que tradicionalmente se han estudiado ciertos problemas en epistemología. La primera de ellas está ubicada en el contexto del clásico problema de dar respuesta al escéptico, concentrándonos específicamente en el problema del error. La segunda vertiente que motivará nuestro planteamiento formal tiene que ver con el problema de encontrar una definición satisfactoria de conocimiento. Nuestro enfoque se centrará en una respuesta fiabilista acerca de la noción de justificación dentro del análisis tripartito de conocimiento.

1.1. El problema del error

La epistemología puede ser vista como el estudio de la posibilidad del conocimiento, y este estudio no sería posible sin estudiar también nuestra propensión a equivocarnos. Durante siglos, uno de los grandes retos de la epistemología ha sido el de derrotar al escepticismo. El escepticismo puede verse como la postura que parte de la pregunta sobre cómo es posible el conocimiento dada la posibilidad de error. Por ejemplo, el escéptico cartesiano comienza su argumento con la posibilidad del error. Pareciera entonces que el error es el primer obstáculo a vencer en contra de la posibilidad del conocimiento. Desde esta perspectiva, la epistemología muestra un interés primordial en dar réplica al escepticismo y a sus retos, en particular al problema del error.

Una de las vías para enfrentar el reto del escéptico es desarrollar una teoría que garantice que donde quiera que el escéptico cite la posibilidad del error como un argumento en contra de la afirmaciones dadas como conocimiento, se puede mostrar que, aún si aceptamos que las afirmaciones tienen posibilidades de error, fallan en cuanto a tener posibilidades relevantes de error. Algunas posibilidades de error simplemente no son genuinas: son muy remotas, muy especulativas o, simplemente van demasiado lejos. Estas posibilidades, en consecuencia, podrían ser descartadas y no ser consideradas en adelante en el proceso de adquisición de conocimiento. La tarea de desechar estas posibilidades ha sido abordada desde distintos enfoques. A veces la tarea es tratada más en terminos tradicionales de epistemología, y a veces en términos más formales. Tres trabajos que podríamos considerar como una especie de "puente.entre los enfoques usuales (o tradicionales) y los formales son: el *fiabilismo epistémico* de Goldman, *la epistemología contrafactual de Nozick* y la *epistemología contextual* de Lewis. Si bien Goldman, Nozick y Lewis son filósofos tradicionales, las teorías antes mencionadas proponen ciertos *procesos*, lo que de alguna manera coincide con la elaboración de teorías y/o herramientas de tipo formal.

El fiabilismo de Goldman (1979, 1986) reconoce las capacidades cognitivas limitadas del agente. La idea de esta propuesta es intentar resolver el problema de Gettier a partir del análisis del concepto de conocimiento en otros términos (en términos causales). De acuerdo al fiabilismo epistémico, no es necesario que los métodos de adquisición del conocimiento sean infalibles. Esto implica que aún cuando los seres humanos somos propensos a equivocarnos, existen procesos cognitivos de elaboración de creencias en los cuales podemos confiar, pues éstos tienden a producir creencias verdaderas. Otra propuesta de tipo falibilista es la de Nozick.

El objetivo del análisis contrafáctico de conocimiento presentado por Nozick (1981) es evitar la Gettierización, dando así buenas razones para defender que el conocimiento es posible. Se podría decir que de cierta manera, la definición dada por Nozick, *evita* (en el mismo sentido que lo hace la propuesta de Goldman) tener que dar una justificación del conocimiento bajo los estándares puestos por el escéptico. Nozick cree que los condicionales contrafácticos ponen de manifiesto un aspecto importante de nuestra comprensión intuitiva del conocimiento, así pues, el agente sólo requiere tener éxito en todos los mundos posibles suficientemente cercanos al mundo actual. Un último ejemplo que presentaremos de conocimiento falibilista es la noción contextual de Lewis.

El método de Lewis no parte del reto escéptico. La epistemología contextual supone el conocimiento de una gran cantidad de cosas en una variedad de diferentes contextos. Los interlocutores de una conversación determinan qué mundos posibles son reales o relevantes y también por qué o cuándo. El conocimiento que se posee en un momento dado puede evaporarse si cambia el contexto, pero en cualquier caso, el punto es que tenemos conocimiento y la epistemología parte de ahí, no de la ignorancia o de la demostración de la mera posibilidad del conocimiento. Lo que necesitamos para obtener conocimiento son reglas dictadas por el actual contexto (conversacional) para descartar los mundos posibles y que entonces sea posible describir cómo evitamos el error y ganamos verdad en los mundos posibles restantes. En los tres ejemplos presentados pareciera que la noción de conocimiento determina también una noción de relevancia para los estados de cosas en los cuales evaluamos el conocimiento. Al ser falibilistas las tres propuestas, la noción de relevancia parece descartar aquellos mundos en los que somos engañados por un genio maligno, somos cerebros en cubetas, etc. Por ejemplo, en el caso de Goldman podemos decir que aquellos estados epistémicos en los cuales las creencias no son producto de procesos cognitivos fiables, resultan irrelevantes para lo que consideramos conocimiento. Todas aquellas creencias producto de alucinaciones o engaños entrarían dentro de esta clasificación.

Por el lado de los tratamientos formales, existe una serie de propuestas epistemológicas que comparten la estrategia de establecer una noción de relevancia para los estados desde los cuales evaluamos el conocimiento. Las afirmaciones del conocimiento pueden ser *depuradas* por restricciones algebraicas definidas por la relación de accesibilidad entre mundos posibles (como veremos en el capítulo 2), lo cual parece ser el fundamento de las técnicas mencionadas en los párrafos anteriores y en general, el fundamento para la lógica epistémica.

La lógica epistémica se origina con Von Wright (1951) y fue propuesta más notablemente por Hintikka (1962). Las propiedades algebraicas de la relación de accesibilidad entre mundos posibles pueden ser definidas de tal modo que el escéptico no tenga a dónde ir. La idea básica es que cuando un agente se encuentra frente a un problema de aprendizaje epistémico, el problema determina un conjunto de mundos posibles en cada uno de los cuales el agente ha tenido éxito en solucionar el problema y además ha adquirido conocimiento.

Una propuesta más que cabe en el enfoque aquí presentado de epistemología formal es la presentada por Hendricks (2001), a la que llamó *epistemología de operador modal*. La epistemología de operador modal es una mezcla de un modo epistémico y lógica alética con algunos conceptos surgidos de la lógica computacional. Fue desarrolllada con el objetivo de estudiar la validez de crear límites para el conocimiento convergente (Hendricks 2001). A fin de encontrar este límite, el agente tiene que converger a la hipótesis verdadera sólo en los mundos posibles consistentes con lo que se ha observado hasta ahora. Según Hendricks (2007), la epistemología puede ser presentada de dos diferentes maneras:

- La aproximación tradicional (*mainstream*), que busca condiciones de necesidad y suficiencia para la atribución de conocimiento utilizando ampliamente consideraciones de sentido común, así como ejemplos y contraejemplos muy particulares y,

- la aproximación formal, que incluye epistemología lógica, epistemología computacional y epistemología de operador modal; ésta utiliza procesos axiomáticos o se concentra en el aprendizaje o la adquisición de conocimiento, utilizando herramientas de lógica y teoría de la computación.

Lamentablemente, estas dos tradiciones no se han puesto mucha atención una a la otra, pero ambas utilizan en el sentido aquí planteado el mismo tipo de técnica (revisar las creencias y descartar el error), para combatir el escepticismo.

1.2. La justificación como proceso

El segundo debate en epistemología contemporánea que utilizaremos para motivar nuestro enfoque formal se centra en la parte de justificación del análisis tripartito de conocimiento. Desde hace tiempo, los filósofos han trabajado en la noción de conocimiento con el fin de evitar la "gettierización"(Gettier 1963) y otras dificultades epistémicas. Se supone que la condición de justificación asegura que las condiciones de verdad y creencia de la definición tripartita están "conectadas adecuadamente".

Las estrategias estándar para *perfeccionar* esta noción de conocimiento han sido agregar requisitos a la condición de justificación, o complementarla con una cuarta condición. A lo largo de los años han surgido diversas teorías que proponen agregar una cuarta condición; algunas de ellas resuelven el problema original de Gettier y otras no, pero en cualquier caso, lo que suele ocurrir es que surgen nuevos cuestionamientos derivados del problema original (Dancy 1993). Esta es entre otras, una razón por la cual una propuesta fiabilista puede resultar muy atractiva; pues en lugar de añadir condiciones (lo que muy probablemente añadiría probremas), pretende modificar la definición tripartita de conocimiento vinculando las nociones de creencia y de verdad con la noción de justificación. Esto es, entendemos una creencia como resultado de un proceso cognitivo y decimos que una creencia está justificada, cuando el proceso que la genera es un proceso fiable. La fiabilidad se da cuando el proceso tiende a generar creencias que resultan verdaderas. Aunque Goldman es el epistemólogo "fiabilista"por antonomasia, no es el único que ha desarrollado teorías en este sentido. Muchos epistemólogos contemporáneos también han intentado dar una noción de justificación en términos de fiabilidad con el fin de resolver los problemas de Gettier. Nozick (1981), por ejemplo, insiste en una estrategia más "radical.ªpelando a un procedimiento altamente fiable que a pesar de variaciones en sus condiciones relevantes, sea conducente a la verdad, utilizando para esto condicionales contrafácticos. Notemos que aún cuando haya distintas concepciones de lo que "fiabilidad"significa, todas las propuestas de este tipo coinciden en que el criterio recae sobre el proceso mediante el cual se produce la creencia. Este enfoque sugiere de manera natural un tratamiento formal del concepto de justificación. Como Hendricks (2006) dice:

> La condición de la justificación en epistemología, equivale a la metodología vista como el estudio sobre cómo la ciencia llega a sus enunciados, es decir, sobre cómo las creencias están justificadas por los cánones, normas o recomendaciones y el funcionamiento interno del método aplicado, o el agente en cuestión.

En filosofía de la ciencia, por ejemplo, la metodología es la que se encarga del estudio de la naturaleza de la justificación y su conexión con la racionalidad. La metodología puede ser (someramente) caracterizada como el estudio de los métodos (o procesos) mediante los cuales la ciencia accede a sus verdades postuladas. Los metodologistas, así como los filósofos de la mente formales, cuentan con una amplia gama de herramientas para analizar e, idealmente, poder garantizar que los métodos utilizados en la ciencia sean conducentes a la verdad. Estas técnicas van desde varias lógicas inductivas y no monotónicas, hasta el bayesianismo; desde la teoría de juegos y la teoría de revisión de creencias, hasta la teoría del aprendizaje formal, y así sucesivamente. En cierto sentido, el discurso de un filósofo "tradicional"sobre justificación es *equivalente* al de un filósofo "formal"sobre metodología. Así pues, una de las tareas de este trabajo es utilizar un método fomal para estudiar un fenómeno epistémico en el contexto de un juego, y de esta manera ofrecer una clara justificación para las creencias verdaderas a las que intuitivamente denominaríamos como *el conocimiento en el juego.*

Entre otras virtudes, un tratamiento formal podría proporcionar algún tipo de claridad a la noción de justificación. En este sentido, la metodología puede ser un recurso para aumentar la claridad del concepto. Para Bonjour (1976) una teoría epistemológica adecuada debe establecer una conexión entre cómo da cuenta de la verdad y cómo da cuenta de la justificación, es decir, debe mostrar que la justificación -como la plantea la teoría- es conducente a la verdad (p.75). Una labor central en filosofía de la ciencia es estudiar la conducción a la verdad del método. Si bien estas cuestiones se relacionan directamente con el status epistémico del método, también es cierto que llevan indirectamente a la naturaleza de la justificación racional. Si el uso del método conduce a la verdad, entonces, dada la relación entre el método y la justificación, la garantía que proporciona el método también es una garantía para la verdad (Sankey, 1999). Independientemente de que aceptamos o no la interpretación que da Hendricks del papel heurístico que juegan las teorías formales en una noción fiabilista de justificación, la herramienta formal nos proporciona una salida para resolver dos de los cuestionamientos que se le presentan a estas propuestas. Éstos son, el problema de cuándo sabemos que un proceso es fiable y el problema de la individuación de procesos. Revisar un proceso cognitivo con mecanismos como lógica epistémica o revisión de creencias puede determinar cuándo áste es fiable. Además, dado que una creencia puede ser el resultado de diversos procesos, un mecanismo formal puede ayudar a especificar exactamente de qué proceso es resultado una creencia dada.

Lo mencionado en el pàrrafo anterior es muy claro en el caso del sudoku: no diríamos que tenemos un buen método si éste no nos conduce a la solución del juego (que puede ser descrita por un conjunto de proposiciones verdaderas). Por otro lado, cuando el agente resuelve un juego de sudoku, en general no está siguiendo un procedimiento formal, pero un modelo de lógica epistémica que distinga las conjeturas del conocimiento, sí puede determinar qué tan fiable es el proceso mediante el cual el agente resuelve el juego. Es muy importante para los objetivos del presente trabajo marcar la distinción entre un método para solucionar el juego (que no es lo que haremos, ya que esto sería materia de un problema matemático y hay bastantes resultados al respecto en teoría de la computación), y uno que nos proporcione información acerca de las creencias en el juego; por ejemplo, que nos permita distinguir creencias de conocimiento. Hablando en términos de comocimiento y del modelo que presentaremos en el capítulo 4, el modelo debe mostrar que la interacción entre nuestras diferentes creencias conduce a proposiciones verdaderas que son las que constituyen nuestro conocimiento final. Lo importante es determinar entonces qué elementos deben componer el método para conseguir el objetivo. Por ejemplo, en el sudoku (como en muchos otros casos) el error juega un papel importante en el camino hacia el conocimiento. Una creencia incorrecta, (por ejemplo "El número 5 corresponde a la casilla ijk") en sí misma no aporta nada, pero una vez que somos conscientes de que la proposición es falsa, las posibilidades se van reduciendo de manera tal que dicho procedimiento nos llevará a darle estatus de conocimiento a una proposición (verdadera) del tipo "El número n corresponde a la casilla ijk"donde $n \neq 5$. Así pues, parece que las proposiciones en el sudoku quedarán justificadas por la fiabilidad del método con el que determinemos su verdad, es decir, por lo "bien hecho"(coherente, preciso, exhaustivo, etc.) del modelo. Por ejemplo, una de las condiciones que este modelo debe de cumplir, es la de dar cuenta de qué proposiciones, por el hecho de ser falsas, no se deben formular, lo que en el juego se traduce en jugadas que no se pueden realizar.

1.3. Epistemología formal

Otra diferencia metodológica característica de un tratamiento formal en epistemología es que el enfoque aborda los problemas "por partes". Una cierta cantidad del entendimiento conceptual es presupuesto, o ciertos parámetros conceptuales se fijan desde el comienzo. Los parámetros que se fijan pueden ser muy variados: el objetivo epistémico, la actitud, el método, la fuerza (o validez) del conocimiento o de la creencia, etc. Incluso, pueden ser combinaciones de estos elementos. Tener estos parámetros fijos, más la estructura adicional impuesta por el aparato formal elegido (lógica, teoría de aprendizaje, teoría de la probabilidad, teoría de juegos, revisión de creencias, etc.) da lugar a un modelo formal de investigación. En este modelo, nociones particulares como validez, fiabilidad, computabilidad, y racionalidad son naturalmente seleccionadas por el análisis formal. La idea es entonces estudiar lo que se sigue a partir del modelo con respecto a los conceptos que nos interesan. Aún cuando los resultados de este análisis conceptual parcial en un marco formal, pueden no aportar nada al concepto global de conocimiento, es posible que nos proporcione información acerca de la estructura de los componentes que integran dicho concepto. De esta manera, diversos trabajos que en principio parecen estar alejados del *verdadero* objeto estudio de la epistemología, aportan elementos y arrojan luz a cuestiones que resultan relevantes en el estudio de los fenómenos epistemológicos.

En las últimas dos décadas, la epistemología formal ha cobrado importancia en distintos departamentos e institutos de investigación en el mundo. La epistemología formal se ocupa de temas tanto de la teoría tradicional del conocimiento, como de la epistemología moderna de las creencias (considerando a las creencias como conjuntos de proposiciones). Para abordar estos problemas se hace uso de diversas herramientas formales de la lógica y las matemáticas. Por ejemplo, en el Departamento de Filosofía de la Universidad de Constanza hay actualmente un proyecto llamado "*Formal Epistemology*". Este proyecto se divide en tres partes: creencias y su revisión, grados de creencia y creencia, y revisión de creencias en lógica dinámica epistémica.

La revisión de creencias (que veremos a detalle en el capítulo 3) es el proceso mediante el cual revisamos nuestros conjuntos de creencias a partir de nueva información (que se presenta en forma de proposiciones). La revisión de creencias es una teoría epistémica, donde por *teoría epistémica* entendemos un mecanismo conceptual que sirve para estudiar los cambios de conocimiento y creencias de un agente racional. Presentar los componentes de una teoría epistémica podría ayudarnos a entender la relación que hay entre nuestras intuiciones y las cuestiones tratadas tradicionalmente, y un aparato formal que pretende abordar problemas de epistemología. Tenemos entonces que los componentes de una teoría epistémica son:

- Estados Epistémicos: Los estados epistémicos son usados para representar el estado actual, o posibles estados cognitivos de un agente racional en un determinado momento. Un estado epistémico se dice *en equilibrio* si satisface los criterios de racionalidad.

- Actitudes Epistémicas: Son la condición de las creencias con respecto al estado epistémico. Por ejemplo en un modelo proposicional las posibles actitudes serían *ser aceptada, rechazada, indeterminada.* En un modelo probabilístico podrían *ser probable, deseable*, entre otros.

- Entradas Epistémicas: Si asumimos que el conjunto de creencias o el estado epistémico es internamente estable, las actualizaciones requerirán de un estímulo externo: las entradas epistémicas. Dichas entradas provocarán un *cambio de creencias* transformando el estado epistémico original en un nuevo estado epistémico.

- Criterios de Racionalidad: Se encuentran situados en el metanivel de la teoría epistémica y son usados para determinar el comportamiento de los cambios de creencias. Por ejemplo *mínimo cambio de las creencias preexistentes, primacía de la nueva información, consistencia*, entre otros.

Adicionalmente algunos autores (eg. Dalal 1988) creen que existen principios básicos que deberían ser adoptados en toda teoría epistémica:

- Concordancia de la representación: La representación de un estado epistémico después de producido un cambio de creencias debe ser del mismo tipo que la representación del estado epistémico antes del cambio.

- Imparcialidad: Si existen varios estados epistémicos candidatos como posibles resultados de una función de cambio, entonces ninguno de ellos debe ser elegido arbitrariamente como resultado.

Si la nueva información es compatible con la totalidad de las viejas creencias, simplemente debemos añadirla al conjunto. El caso más interesante es cuando la nueva información es incompatible con las viejas creencias. En este caso, las reglas de actualización tales como la condicionalización estricta se colapsan. Entonces surge la cuestión de cómo combinar las viejas y las nuevas creencias. Un problema que presenta la revisión de creencias es que parte del supuesto de que nosotros revisamos nuestras creencias si y sólo si existe una contradicción lógica dentro del conjunto, cosa que no siempre es cierta.

El estudio de los grados de creencia tiene una parte cualitativa que se enfoca en la noción categórica de cierto tipo de creencias (las creencias de sí-o-no) y las leyes a las que éstas obedecen, así como una parte cuantitativa (que es la que será tratada más adelante) cuyo interés son los grados de las creencias y las leyes que éstos deben obedecer. El objetivo del proyecto *Degrees of Belief and Belief* del Departamento de Filosofía de la Universidad de Constanza es desarrollar teorías que puedan dar cuenta de estados epistémicos que comprenden tanto la creencia categórica como los grados de creencia y verificar la conexión entre estos dos aspectos (Foley 1992, Hawthrone & Bovens 1999). Las teorías de grados de creencia pueden ser relacionadas con la plausibilidad e informatividad de una teoría (Hubert 2008). Esto es, que una hipótesis es aceptable si es plausible y suficientemente informativa. Por ejemplo, las tautologías, mientras que son maximalmente plausibles, son también poco (o nada) informativas para *merecer* ser llamadas aceptables. Por otro lado, una hipótesis completa puede ser altamente implausible. Teorías tales como la de la probabilidad subjetiva analizan toda la gama que existe entre estos dos extremos.

Al igual que la revisión de creencias, la lógica modal epistémica también ha proporcionado un campo fértil en el que se pueden estudiar problemas epistemológicos (entre otros). Los estados epistémicos tales como el conocimiento y la creencia pueden ser modelados con lógica epistémica. Si agregamos un componente dinámico obtenemos lógica dinámica epistémica, la cual nos permite representar estados epistémicos y acciones de múltiples agentes (van Ditmarsch & van der Hoek & Kooi 2007). Recientemente, han surgido diversas propuestas para la revisión de creencias que incorporan lógica dinámica epistémica (e.g. van Benthem 2007), lo cual proporciona recursos para estudiar además de los estados epistémicos, las dinámicas entre ellos, capturando cuestiones como la actualización de la información de los agentes.

Dada su importancia en el contexto de la epistemología formal, y que será la herramienta que utilizaremos para modelar el conocimiento en el sudoku, en el capítulo 3, presentaremos un panorama detallado sobre lógica modal epistémica. Este tratamiento nos permitirá ver cómo esta herramienta nos permite capturar condiciones epistemológicas complejas como son el conocimiento compartido y el conocimiento común.

Antes de entrar en materia con las cuestiones formales, es pertinente detenernos un poco para discutir qué queremos decir cuando utilizamos expresiones como *modalidades* o *mundos posibles*.

1.4. La modalidad desde el enfoque usual

Las nociones contemporáneas de conocimiento se presentan con frecuencia en versiones modales. Esto es porque el conocimiento es definido con respecto a posibles estados de cosas distintos al estado de cosas actual. De esta manera, los enfoques tradicional y formal de la epistemología han convergido en definiciones de conocimiento con respecto a otros mundos posibles. Como lo señala Hintikka en (2003), "Con el fin de de hablar de lo que una persona a sabe y no sabe, tenemos que suponer una clase ("espacio") de posibilidades. Estas posibilidades serán llamadas escenarios. Los filósofos típicamente los llaman *mundos posibles*."(p.19).

Una vez que contamos con el recurso conceptual de los mundos posibles, es necesario establecer criterios que nos ayuden a determinar cuáles de ellos deben ser considerados relevantes para tener éxito epistémico. Por ejemplo, una noción clásica de infalibilismo requiere que el agente, con el fin de tener conocimiento acerca de una hipótesis, deba estar capacitado para eliminar todas las posibilidades de error asociadas a la hipótesis en cuestión. El problema es que al pensar en cosas tales como "El conjunto de *todos* los mundos posibles tales que...", el asunto podría volverse inmanejable. Estas clases podrían ser tan grandes que podrían incluso salirse de la categoría de los conjuntos. Pero además, es claro que no sólo se trata de algo que podría ser conceptualmente complicado, sino que seguramente es innecesario. Es claro que gran parte de esas clases está constituida por mundos tan extraños y extravagantes que quedan realmente *lejos* del mundo actual y que tener conocimiento de lo que en ellos ocurre podría no ser más que una pérdida de tiempo. Suponiendo que fuera posible, si estos mundos fueran considerados relevantes todo el tiempo, el escéptico tendría ahí un recurso constante para cuestionar la posibilidad de conocimiento. Parece entonces que la epistemología se ve orillada a aceptar una noción falibilista de conocimiento (pues el agente podría no tener la posibilidad de demostrar cuestiones como que la proposición p es verdadera en todos los mundos posibles, incluso en aquél en donde todo lo falso es verdadero). Sin embargo, algunos epistemólogos podrían seguir sosteniendo que una noción falibilista no puede calificar como conocimiento en el más amplio sentido de la palabra (pues el conocimiento implica la verdad de lo que se cree). A lo más calificaría como una descripción de cómo es aquello que en la práctica consideramos conocimiento. Pero si aceptamos una noción infabilista de conocimiento, éste no puede ser definido respecto a todos los mundos posibles. Si aceptamos una definición en términos de mundos posibles, ésta debe ser infalibilista y sigue pendiente la tarea de buscar criterios que puedan determinar cuáles son los mundos *irrelevantes*. Aquí es donde algunos epistemólogos (tradicionales y formales) ejercen su labor mediante la creación de procesos que sirven como "depuradores"del error que finalmente nos proveen de una gran variedad de teorías en epistemología. El principio general de estos procesos podría verse de esta forma:

> Si el escéptico cuestiona la posibilidad del conocimiento a partir de la posibilidad del error, el reto es mostrarle que las posibilidades de error fallan en ser genuinas en el sentido de su relevancia. Donde la relevancia está dada por el contexto de la teoría epistemológica.

El enfoque también tiene que ver con los términos en los que se dan los mundos *relevantes*. Por un lado, se pueden buscar condiciones que determinen el conjunto de los mundos relevantes como subconjunto de todos los mundos posibles. En general, los lógicos y epistemólogos formales definen una relación de accesibilidad dentro del producto cartesiano del universo de mundos posibles. Lo deseable es que no haya una distinción importante entre *relevancia* y *accesibilidad*. Un epistemólogo informal por ejemplo, daría la relevancia en términos de condiciones de equivalencia perceptual, o proximidades contrafácticas (Lewis 1973,Stalnaker 1968), o contextos conversacionales que circuscriban a los mundos posibles, mientras que un epistemólogo formal o un lógico, mediante las variaciones de las propiedades de la relación de accesibilidad (reflexividad, serialidad, transitividad, etc.) obtiene diferentes sistemas modales epistémicos válidos cuyo objetivo es determinar justamente los mundos relevantes relativos a nuestros intereses.

Hemos argumentado a favor de una noción fiabilista del conocimiento, proponiendo también una solución al problema de cómo determinar que un proceso cognitivo es fiable. Los sistemas formales sirven para evaluar en qué medida los razonamientos producen creencias verdaderas. Por ejemplo, un sistema lógico puede determinar si los procesos inferenciales son correctos. Así mismo (aunque aquí no desarrollamos este punto), este tipo de procedimientos parecen disponer de elementos capaces de individuar procesos cognitivos. Entonces, hay razones para pensar que trabajar en el desarrollo de estrategias formales podría robustecer de manera importante las posturas fiabilistas en epistemología.

La herramienta formal nos permite dar una respuesta al cuestionamiento del escéptico acerca de cómo podemos hacer afirmaciones de conocimiento dado que siempre existe la posibilidad del error; herramientas formales como teoría de juegos (en particular árboles de decisión), revisión de creencias, etc., proponen ejecutar acciones concretas que abren paso al conocimiento, una vez que se detectan las creencias incorrectas (proposiciones falsas).

La construcción de un modelo formal (utilizando una lógica multimodal epistémica trivaluada) para estudiar el conocimiento en el sudoku que presentaremos en el capítulo 4 satisface las motivaciones planteadas en este capítulo. Por un lado, el sistema modal permite determinar, dado un proceso cognitivo para resolver el sudoku, el grado en el que éste arroja verdades. Por otro lado, el modelo apunta a que en una segunda etapa podremos determinar la función que el error (las conjeturas falsas) tiene en el proceso de adquisición de conocimiento dentro del juego.

La evidencia presentada juega un papel a favor del uso de sistemas formales en la investigación de fenómenos epistémicos. Un trabajo posterior sería generalizar el modelo construido para el caso del sudoku y llevarlo a contextos más relevantes en el campo de la teoría de la computación y de la epistemología.

Capítulo 2

Lógica modal epistémica

La lógica modal es una de las herramientas formales que más se ha popularizado durante los últimos años. La razón es, entre otras cosas, que ha tenido grandes alcances en diversas disciplinas. Por ejemplo, uno de sus logros es que ha resultado sumamente útil en la modelación del conocimiento. A la lógica modal utilizada con este fin se le conoce como lógica epistémica. El principal objetivo de la lógica modal epistémica es crear un lenguaje con el cual se puedan expresar las creencias, el conocimiento, y sus interacciones en un contexto dado. En este capítulo presentaremos la sintaxis y la semántica, así como algunos ejemplos y observaciones acerca de estos sistemas.

Como se menciona en la introducción, el conocimiento juega un papel determinante en las decisiones que se toman, tanto en la vida cotidiana, como en el quehacer científico. Pero en ocasiones, las decisiones no sólo dependen del conocimiento que tenemos de los hechos, sino también del conocimiento que tenemos acerca del conocimiento (o creencias) que tienen otros agentes. Los juegos ejemplifican muy bien esta situación. Por ejemplo, en el pókar, la idea que existe detrás del *bluffing*, es generar creencias en el resto de los jugadores acerca de nuestro propio conocimiento de la situación del juego. Un caso similar ocurre en economía cuando las operaciones bursátiles son determinadas a partir no sólo de lo que se sabe, sino de lo que se cree (o se sabe) que otros agentes saben. La lógica epistémica nos permite razonar acerca del conocimiento; sus alcances llegan no sólo a dar cuenta del conocimiento que tenemos acerca de los hechos, sino también del conocimiento que poseemos acerca del conocimiento (o creencias) que poseen otros agentes. Un lenguaje que nos permita expresar hechos acerca del mundo, y que también nos permita expresar el conocimiento que cada agente tiene acerca de estos hechos resulta muy útil para estudiar casos como los mencionados.

Hacer la distinción entre los hechos del mundo y el conocimiento que cada agente tiene de ellos implica que, dado un hecho, no necesariamente cada uno de los agentes tiene conocimiento acerca de él. En estos casos, cuando el agente no sepa si una proposición es verdadera o falsa en alguna situación particular, considerará posible un mundo en el cual dicha proposición es verdadera, pero también considerará posible otro mundo en el cual la proposición es falsa. Si su incertidumbre no es sólo respecto a un hecho sino a varios, considerará posibles una mayor cantidad de mundos. La semántica de mundos posibles pretende describir este tipo de situaciones.

2.1. Mundos posibles

La mayoría de los intentos de modelar el conocimiento se han basado en el modelo de los mundos posibles (Hintikka, 1962). En lógica proposicional clásica, para determinar cuando una fórmula es falsa o verdadera, es necesario verificar el valor de verdad de cada una de las proposiciones atómicas que la componen. Fórmulas tales como "Está soleado en la Ciudad de México y está lloviendo en Londres" reciben un valor de verdad de acuerdo al valor de verdad de las proposiciones "Está soleado en la Ciudad de México" y "Está lloviendo en Londres".

Supongamos que está soleado en la Ciudad de México y un agente A se encuentra ahí. El agente A sabe entonces que está soleado en la Ciudad de México, pero no sabe si está lloviendo o no en Londres. El agente A considera posible la situación en la cual está soleado en la Ciudad de México y está lloviendo en Londres, pero también considera posible la situación en la cual está soleado en la Ciudad de México y está nevando en Londres. Lo único que A sabe de hecho, es que está soleado en la Ciudad de México (que es lo único que él puede asegurar que es verdadero en las dos situaciones que considera posibles). Podemos decir que, intuitivamente, entre menos situaciones considere posibles el agente A, menor será su incertidumbre acerca de cual es la situación real. Esta es la noción de mundos posibles que Hintikka presentó en 1962 con la finalidad de estudiar situaciones en las cuales un agente no está seguro acerca de cuál es la situación real en la que se encuentra. Esta presentación es muy útil para describir una situación en un tiempo t no sólo en términos de las proposiciones que son verdaderas en t, sino también, en términos de lo que el agente sabe acerca de estas proposiciones. Dado entonces un conjunto de proposiciones atómicas, podemos describir qué es lo que sabe un agente acerca de estas proposiciones mediante un conjunto de mundos posibles (en cada uno de los cuales son ciertas o no estas proposiciones), así como relaciones entre estos mundos (que nos dicen cuándo un mundo es compatible con el conocimiento de un agente).

A continuación presentaremos el lenguaje clásico de la lógica modal epistémica. Es pertinente señalar que la formulación original (Hintikka, 1962) está pensada para un solo agente. En el caso del modelo del conocimiento en el sudoku tampoco es necesario considerar una lógica multiagente, sin embargo, aquí la presentamos por que ésta juega un papel importante en los tratamientos actuales de epistemología formal.

2.2. Sintaxis

El lenguaje clásico de la lógica modal epistémica ($\mathcal{LE}$) se construye a partir de:

- Un conjunto de proposiciones atómicas (denotado como Φ) en cuyos términos es posible describir el mundo sobre el cual deseamos que los agentes razonen.
- Un conjunto de agentes (comúnmente denotado como $\mathcal{A}$).
- Un operador modal K_i para cada agente $i \in \mathcal{A}$, y los conectivos lógicos $\neg$ (negación) y $\vee$ (disyunción).

Este lenguaje se define formalmente de la siguiente manera:

Definición 2.2.1 *(Lenguaje de la lógica epistémica $\mathcal{LE}$). Las fórmulas que conforman el lenguaje de la lógica modal epistémica $\mathcal{LE}$ se construyen a partir de un conjunto numerable no vacío de fórmulas atómicas Φ y un conjunto finito de agentes A siguiendo las reglas que se muestran a continuación:*

$$
\begin{array}{lll}
\top \in \mathcal{LE} & & \\
\textit{Si } p \in \Phi & \textit{entonces} & p \in \mathcal{LE} \\
\textit{Si } \varphi \in \mathcal{LE} & \textit{entonces} & \neg\varphi \in \mathcal{LE} \\
\textit{Si } \varphi \in \mathcal{LE} \textit{ y } \psi \in \mathcal{LE} & \textit{entonces} & \varphi \vee \psi \in \mathcal{LE} \\
\textit{Si } i \in \mathcal{A} \textit{ y } \varphi \in \mathcal{LE} & \textit{entonces} & K_i\varphi \in \mathcal{LE}
\end{array}
$$

En notación BNF, las fórmulas $\varphi \in \mathcal{LE}$ se definen con las siguientes reglas:

$$\varphi ::= \top \mid p \mid \neg\varphi \mid (\varphi \vee \psi) \mid K_i\varphi$$

Donde $p \in \Phi$, $i \in \mathcal{A}$ y $\varphi, \psi \in \mathcal{LE}$.

La fórmula $\top$ representa la fórmula que siempre es verdadera, $\neg\varphi$ representa la negación de φ, $(\varphi \vee \psi)$ representa la disyunción de φ y ψ, y $K_i\phi$ expresa "El agente i sabe ϕ".

Como podemos observar, el lenguaje $\mathcal{LE}$ es un conjunto que se forma a partir de la proposición $\top$, las proposiciones atómicas $p \in \Phi$, y que es cerrado bajo la disyunción, la negación y los operadores modales K_i para cada agente $i \in \mathcal{A}$. Las conectivas lógicas $\wedge, \rightarrow, \leftrightarrow$ (conjunción, implicación y doble implicación respectivamente) y los operadores modales P (posibilidad) y E (todos saben)se definen como:

$$\begin{aligned}
\bot &\equiv_{def} \neg\top \\
(\varphi \wedge \psi) &\equiv_{def} \neg(\neg\varphi \vee \neg\psi) \\
(\varphi \rightarrow \psi) &\equiv_{def} (\neg\varphi \vee \psi) \\
(\varphi \leftrightarrow \psi) &\equiv_{def} ((\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)) \\
P_i\varphi &\equiv_{def} \neg K_i \neg\varphi \\
E_{\mathcal{B}}\varphi &\equiv_{def} \bigwedge_{i \in \mathcal{B}} K_i\varphi
\end{aligned}$$

Las conectivas $\wedge, \rightarrow$ y $\leftrightarrow$ tienen el significado habitual, las fórmulas del tipo $P_i\varphi$ intuitivamente significan "No es cierto que el agente i sepa que no se cumple φ", otra posible interpretación es "El agente i considera posible φ". Finalmente, las fórmulas de la forma $E_{\mathcal{B}}\varphi$ se leen como "Todos los agentes que pertenecen al conjunto $\mathcal{B}$ saben φ (donde $\mathcal{B}$ es un subconjunto del conjunto de agentes $\mathcal{A}$).

Como podemos notar, sintácticamente, lo único que le hemos agregado a la lógica proposicional clásica es el operador modal K_i para cada agente $i \in \mathcal{A}$. Como mencionamos anteriormente, cada fórmula de la forma $K_i\varphi$ se lee como .El agente i sabe (conoce) φ". Por ejemplo:

$K_i\varphi$ El agente i sabe φ.

$K_i\varphi \wedge K_j\psi$ El agente i sabe φ y el agente j sabe ψ.
$K_i(\varphi \rightarrow \psi)$ El agente i sabe que φ implica ψ.
$K_i\varphi \wedge \neg K_j\varphi$ El agente i sabe φ y no es cierto que el agente j sepa φ.

El lenguaje tiene la ventaja de que podemos expresar conocimiento de órdenes superiores. Es decir, no sólo podemos expresar lo que un agente sabe acerca de los hechos, sino también lo que él sabe acerca de lo que los otros agentes saben.

$K_i\varphi$ El agente i sabe φ.
$K_iK_j\varphi$ El agente i sabe que el agente j sabe que φ.
$K_iK_jp \wedge \neg K_jK_iK_jp$ El agente i sabe que el agente j sabe p, pero el agente j no sabe que el agente i sabe que el agente j sabe p.
$\neg K_i\psi \wedge P_jK_i\psi$ No es cierto que el agente i sepa ψ, pero el agente j lo considera posible.

Estrictamente hablando, no existe un operador análogo al operador modal de posibilidad, pero podemos definirlo. Por ejemplo, en nuestra definición tenemos que:

$$P_i\varphi \equiv_{def} \neg K_i\neg\varphi$$

Sin embargo, es importante tener presente que su interpretación no corresponde, al menos en principio, con una noción intuitiva de creencia. Para el modelo del sudoku, definiremos un operador dual de conocimiento que claramente, no coincide con lo que sería una creencia dentro del juego. Esto a primera vista, pareciera ser un punto débil de la lógica modal epistémica.

La sintaxis presentada nos sirve para escribir fórmulas que pertenezcan al lenguaje, asi como para dar la interpretación intuitiva de estas fórmulas. Para determinar cuándo una fórmula es verdadera o falsa es necesario presentar la semántica de nuestro lenguaje.

2.3. Semántica

Como ya hemos mencionado, la mayoría de los trabajos realizados en lógica epistémica parten del trabajo que Jakko Hintikka presentó en *Knowledge and Belief: An Introduction to the Logic of the Two Notions* (1962). Hintikka combinó sus conocimientos matemáticos con las ideas de lógica modal de von Wright en un manuscrito que fue el primero en dar una semántica clara de las nociones de *creencia* y *conocimiento*. Esto no sólo puso fin a una era de intentos de la filosofía por encontrar medios para debatir formalmente estas nociones sino que también marcó el punto de partida de una nueva era en la que, más que el desarrollo de nuevas cuestiones epistemológicas, surgieron nuevos enfoques mediante los cuales se abordaron los problemas tradicionales.

Von Wright (1951) fue el primero en dar cuenta formal (principalmente de manera sintáctica) de los razonamientos acerca del conocimiento en términos de lógica modal. A principios de la década de los 60, la semántica de mundos posibles tuvo un fuerte auge. Uno de los primeros trabajos que se hicieron desde esta vertiente fue el de Carnap (1946), el cual posteriormente fue enriquecido por la noción de *accesibilidad* entre mundos dada por Hintikka (1957) y perfeccionado por Kripke (1963). Estas propuestas semánticas han demostrado ser muy eficientes para interpretar nociones tan diversas como *ser epistémico*, *ser doxástico*, *ser temporal*, *ser deóntico*, *ser dinámico*, así como aquellas concernientes a la probabilidad. Esta fue la era en la que la lógica modal se convirtió en una importante herramienta de razonamiento en diversas disciplinas.

La idea de usar la semántica de mundos posibles para el conocimiento y la creencia consiste en pensar en la información que un agente tiene en términos de los mundos posibles que son consistentes con dicha información. Decimos que estos mundos posibles son accesibles para el agente. Así, el conocimiento y la creencia pueden ser definidos en términos de estos mundos; un agente sabe que algo es el caso, si y sólo si es el caso en todos los mundos que son accesibles al agente. Esta semántica para la lógica epistémica nos permite formalizar razonamientos acerca del conocimiento y la creencia. Por otra parte, la semántica proporciona una manera concisa para representar la información de un agente, y además, permite añadir fácilmente más agentes, de tal manera que también es capaz de representar la información que cada uno de los agentes tiene acerca de la información que tienen los otros.

Definición 2.3.1 *(Estructura de mundos posibles). Dado un conjunto de proposiciones atómicas* Φ *y un conjunto de agentes* $\mathcal{A}$*, una estructura de mundos posibles es una terna* $M = \langle W, R_i, V \rangle$ *donde:*

- W *es un conjunto tal que* $W \neq \emptyset$ *y cada elemento de* W *es llamado un mundo posible.*

- *Para cada agente $i \in \mathcal{A}$, tenemos una relación (llamada de accesibilidad), $R_i \subset W \times W$.*

- *$V : \Phi \to 2^W$ es una función que asigna a cada proposición atómica en Φ un conjunto de mundos posibles.*

En la literatura, a la estructura de mundos posibles también se le conoce como estructura epistémica, modelo de mundos posibles, modelo epistémico o modelo de Kripke. A la pareja $\langle W, R_i \rangle$ se le conoce como marco de mundos posibles, marco epistémico o marco de Kripke. Algunos autores llaman "Marco de Kripke" a lo que acabamos de presentar como "Estructura de mundos posibles"(e.g. Zalta 1995), es decir, al conjunto de mundos junto con la relación de accesibilidad y la función de evaluación.

Una de las ventajas de utilizar marcos de Kripke es que éstos pueden ser representados de manera gráfica. Cada elemento $w \in W$ es representado con un punto un nodo, y cada relación de accesibilidad con una flecha etiquetada de acuerdo al agente al que corresponde. Es decir, si $(w_1, w_2) \in R_i$ (si w_2 es accesible a través de R para el agente i desde w_1), trazamos una flecha de w_1 a w_2 etiquetada por i.

Ejemplo 2.3.2 *(Representación gráfica de un marco de Kripke). Sea $\mathcal{M} = \langle W, R_i \rangle$ un marco de Kripke tal que*

$W = \{w_1, w_2, w_3\}$

$R_1 = \{(w_1, w_3), (w_1, w_2), (w_2, w_1), (w_3, w_2)\}$ *y*

$R_2 = \{(w_1, w_2), (w_2, w_1), (w_3, w_3), (w_2, w_3)\}$.

Una representación gráfica de $\mathcal{M}$ puede ser la siguiente:

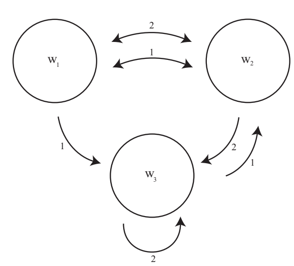

Figura 2.1: Marco de Kripke

El valor de verdad de las fórmulas de lógica epistémica está determinado por la estructura de mundos posibles (a diferencia de la lógica proposicional en donde el valor de vardad de cada fórmula está determinado por el valor de verdad de cada una de las proposiciones atómicas que la componen). El valor de verdad de las fórmulas del tipo $K_i\varepsilon$ se obtiene verificando el valor de verdad de ε en todos los mundos u tal que $(w, u) \in R_i$, que son precisamente aquellos mundos que el agente i considera posibles a partir del mundo real w. El valor de verdad de las fórmulas que no utilizan operadores modales se obtiene de la misma forma que se haría con lógica proposicional (de acuerdo al valor de las proposiciones atómicas) relativo a los mundos posibles en los que se evalúa. Esto quiere decir que para conocer el valor de verdad de las fórmulas en $\mathcal{LE}$ es necesario indicar cuál de los mundos posibles es el mundo real. A la pareja (M, w) donde $M = \langle W, R_i, V\rangle$ es una estructura de mundos posibles y $w \in W$ se le conoce como *estructura puntuada*. Esta estructura nos indica cuál es la situación real desde la que partimos. A w también se le conoce como *punto de evaluación*.

Definición 2.3.3 *(Semántica para $\mathcal{LE}$). Definimos la relación $\models$ entre una estructura puntuada (M, w) (donde $M = \langle W, R_i, V\rangle$) y una fórmula $\varphi \in \mathcal{LE}$ (con M y $\mathcal{LE}$ definidos en términos del conjunto de proposiciones atómicas Φ y del conjunto de agentes $\mathcal{A}$) de manera inductiva sobre la estructura sintáctica de φ.*

$(M, w) \models \top$		
$(M, w) \models p$	*si y sólo si*	$w \in V(p)$
$(M, w) \models \neg\varphi$	*si y sólo si*	*no es cierto que* $(M, w) \models \varphi$
$(M, w) \models (\varphi \vee \psi)$	*si y sólo si*	$(M, w) \models \varphi$ *o* $(M, w) \models \psi$
$(M, w) \models K_i\varphi$	*si y sólo si*	$(M, u) \models \varphi$ *para toda* $u \in W$ *tal que* $(w, u) \in R_i$.

donde $p \in \Phi$, $i \in \mathcal{A}$ y $\varphi, \psi \in \mathcal{LE}$.

Cuando $(M, w) \models \varphi$ se dice que φ es verdadera en la estructura puntuada (M, w). En el caso en que no sea cierto que $(M, w) \models \varphi$ se escribe $(M, w) \not\models \varphi$ y se dice que φ es falsa en (M, w).

Como ya hemos mencionado, $\top$ es la proposición que siempre es verdadera, en particular, será verdadera para cualquier estructura acentuada dada. La semántica para las conectivas lógicas $\neg$ y $\vee$ es la usual: $\neg\varphi$ es verdadera en (M, w) si y sólo si φ es falsa en (M, w) y $\varphi \vee \psi$ es verdadera en (M, w)

si y sólo si φ es verdadera en (M, w) o ψ es verdadera en (M, w). Para el operador modal K_i la semántica es la siguiente: Decimos que K_i es verdadera en (M, w) si y sólo si φ es verdadera en toda estructura acentuada (M, u) tal que $(w, u) \in R_i$. Es decir, el agente i sabe que φ es el caso en (M, w) si y sólo si φ es verdadera en todos aquellos mundos u que i considera posibles a partir de w.

El valor de verdad de la fórmula $\perp$ y las fórmulas compuestas por las conectivas $\wedge, \rightarrow, \leftrightarrow$ y los operadores modales P y E se derivan a partir de los elementos presentados en el párrafo anterior y la definición 2.3.3.

Definición 2.3.4 *(Satisfactibilidad y validez). Sea (M, w) una estructura puntuada tal que $M = \langle W, R_i, V\rangle$, y sea φ una fórmula en $\mathcal{LE}$.*

1. *Se dice que φ es satisfactible en M si $(M, w) \models \varphi$ para algún $w \in W$.*

2. *Se dice que φ es satisfactible si es satisfactible para alguna estructura M.*

3. *Se dice que φ es verdadera en M si $(M, w) \models \varphi$ para todo $w \in W$. En este caso usaremos la notación*

$$M \models \varphi$$

4. *Se dice que φ es válida si es verdadera para cualquier estructura M. En este caso usaremos la notación*

$$\models \varphi$$

Ejemplo 2.3.5 *(Modelo de una situación utilizando lógica epistémica). Retomando el ejemplo de 2.1 consideremos las siguientes proposiciones:*

p= "Está soleado en la Ciudad de México"

q= "Está lloviendo en Londres"

Y las siguientes situaciones que el agente i considera posibles:

Está soleado en la Ciudad de México y está lloviendo en Londres (denotaremos esta situación con w).

Está soleado en la Ciudad de México y no está lloviendo en Londres (denotaremos esta situación con u).

La incertidumbre de i respecto a si está lloviendo o no en Londres puede representarse con el modelo $\langle M, R_i, V\rangle$. Donde

- $W = \{w, u\}$
- $R_i = \{(w, w), (w, u), (u, w), (u, u)\}$
- $V : \Phi \to 2^W$ *tal que* $V(p) = \{w, u\}, V(q) = \{w\}$

Podemos representar M *gráficamente de la siguiente manera:*

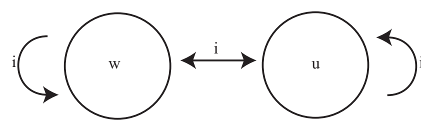

Figura 2.2: Modelo $\langle M, R_i, V\rangle$

Notemos que dado este modelo y considerando a w *el mundo posible que describe la situación real, tenemos que:*

- $(M, w) \models K_i p$ *El agente* i *sabe que está soleado en la Ciudad de México ya que en los dos mundos que considera posibles (*w *y* u*), la proposición* p *es verdadera.*
- $(M, w) \models \neg K_i q \wedge \neg K_i \neg q$ *El agente* i *no sabe si en Londres llueve o no, ya que en uno de los mundos que considera posibles (en* w*) se satisface* q*, y en el otro no.*

En el cuadro 2.1 se muestra la semántica para la lógica epistémica proposicional.

Ya antes habíamos mencionado que una de las razones a las que se debe el éxito que la lógica modal, y en particular de la epistémica, es que dan cuenta clara de algunas distinciones importantes. Por ejemplo, las nociones de *conocimiento común* y *conocimiento compartido* pueden determinarse de manera precisa utilizando lógica epistémica.

2.4. Conocimiento común

La necesidad de expresar nociones que involucran a múltiples agentes ha conducido a la lógica epistémica a presentarse frecuentemente como una lógica multimodal (con una modalidad por agente). En el caso del conocimiento

$$
\begin{array}{ll}
(M,w) \not\models \bot & \\
 & \\
(M,w) \models (\varphi \wedge \psi) & \text{ssi } (M,w) \models \neg(\neg\varphi \vee \neg\psi) \\
 & \text{ssi } (M,w) \not\models (\neg\varphi \vee \neg\psi) \\
 & \text{ssi } (M,w) \not\models \neg\varphi \text{ y } (M,w) \not\models \neg\psi \\
 & \text{ssi } (M,w) \models \varphi \text{ y } (M,w) \models \psi \\
 & \\
(M,w) \models (\varphi \to \psi) & \text{ssi } (M,w) \models (\neg\varphi \vee \psi) \\
 & \text{ssi } (M,w) \models \neg\varphi \text{o} (M,w) \models \psi \\
 & \text{ssi } (M,w) \not\models \varphi \text{ o } (M,w) \models \psi \\
 & \\
(M,w) \models (\varphi \leftrightarrow \psi) & \text{ssi } (M,w) \models ((\varphi \to \psi) \wedge (\psi \to \varphi)) \\
 & \text{ssi } (M,w) \models (\varphi \to \psi) \text{ y } (M,w) \models (\psi \to \varphi) \\
 & \text{ssi } ((M,w) \not\models \varphi \text{ o } (M,w) \models \psi) \text{ y} \\
 & \qquad ((M,w) \not\models \psi \text{ o } (M,w) \models \varphi) \\
 & \text{ssi } ((M,w) \not\models \varphi \text{ y } ((M,w) \not\models \psi \text{ o } (M,w) \models \varphi)) \text{ o} \\
 & \qquad ((M,w) \models \psi \text{ y } ((M,w) \not\models \psi) \text{ o } (M,w) \not\models \varphi) \\
 & \text{ssi } [(M,w) \not\models \varphi \text{ y } ((M,w) \not\models \psi \text{ o } (M,w) \not\models \varphi)) \text{ y} \\
 & \qquad ((M,w) \models \varphi] \text{ o } [(M,w) \models \psi) \text{ y } (M,w) \models \varphi)] \\
 & \text{ssi } ((M,w) \not\models \varphi \text{ y } ((M,w) \not\models \psi \text{ o } (M,w) \models \psi)) \text{ y} \\
 & \qquad ((M,w) \models \varphi) \\
 & \\
(M,w) \models P_i\varphi & \text{ssi } (M,u) \models \neg K_i \neg\varphi \\
 & \text{ssi } (M,u) \not\models K_i \neg\varphi \\
 & \text{ssi existe un } u \in W \text{ tal que } (w,u) \in R_i \text{ y } (M,w) \not\models \neg\varphi \\
 & \text{ssi existe un } u \in W \text{ tal que } (w,u) \in R_i \text{ y } (M,w) \models \varphi \\
 & \\
(M,w) \models E_{\mathcal{B}}\varphi & \text{ssi } (M,w) \models \displaystyle\bigwedge_{i \in \mathcal{B}} K_i\varphi \\
 & \text{ssi } (M,w) \models K_i\varphi \text{ para todo } i \in \mathcal{B}
\end{array}
$$

Cuadro 2.1: Satifacción en la lógica epistémica en el sudoku estas nociones no serán necesarias, pues el juego involucra sólo un agente. Es decir, dependiendo de las necesidades del fenómeno que se pretende modelar es el conjunto de agentes respecto al que vamos a considerar las modalidades. Sin embargo, es importante para los objetivos del presente trabajo, señalar aquellos aspectos epistémicamente relevantes que logran ser capturados por las herramientas formales que aquí presentamos.

Con el fin de motivar la noción de conocimiento común presentaremos el siguiente ejemplo (van Ditmarsch, van der Hoek & Kooi, 2007).

Ejemplo 2.4.1 *Dos agentes, a (Ana) y b (Beto) se encuentran uno frente a otro. Cada uno de ellos lleva un número en la frente y puede ver el número en la frente del otro. Los números son números consecutivos n y $n+1$ para una cierta $n \in \mathbb{N}$. Ambos agentes saben esto, y ambos agentes saben que lo saben, etc. Sin embargo, ellos no tienen ningún tipo de conocimiento a priori acerca de su propio número. Consideremos las proposiciones atómicas a_n y b_m para cada $n, m \in \mathbb{N}$, que expresan que el número en la frente de Ana es n y el número en la frente de Beto es m.*

Supongamos que de hecho a_3 y b_2 son verdaderas. Supongamos también que los agentes sólo ven cada uno el número del otro y que es conocimiento común (ambos saben, ambos saben que saben, etc.) que los números son consecutivos. Podemos realizar las siguientes observaciones.

1. $K_a b_2$

   Ana puede ver el número de b.

   Por la misma razón tenemos que $K_b a_3$.

2. $K_a(a_1 \vee a_3)$

   Ana sabe que su numero o es 1, o es 3.

   Así mismo tenemos que $K_b(b_1 \vee b_3)$.

3. $KaKb(b_0 \vee b_2 \vee b_4)$

   Esto se sigue de 2.

   Similarmente, tenemos que $KbKa(a_1 \vee a_3 \vee a_5)$.

4. $K_a K_b K_a(b_0 \vee b_2 \vee b_4)$

   Pues Ana sabe que lo que Beto sabe sobre lo que ella sabe, sólo puede ser a partir de lo que él sabe sobre su propio número y lo que él sabe que ella sabe acerca del número de Beto.

Por una razón análoga tenemos que $K_bK_aK_b(b_1 \vee b_3 \vee b_5)$.

Esta secuencia de verdades para el estado real, caracterizado por $(b_2 \wedge a_3)$, *tiene un gran número de consecuencias lógicas para el conocimiento iterado correspondiente a* a *y* b. *Esto se podría ver facilmente si enumeraramos todas las posibilidades epistémicas para los agentes.*

Supongamos que el objetivo del juego es que cada uno de los agentes adivine el número que tiene en la frente. Para expresar esto introduciremos la proposición $\mathbf{gana_a}$ *para "Ana sabe cuál es el número que tiene en la frente". Similarmente, introducimos* $\mathbf{gana_b}$ *para "Beto sabe cuál es el número que tiene en la frente".*

5. $\hat{K}_a a_1 \wedge \hat{K}_a a_3$

   Ana considera posible que su número sea 1, pero también considera posible que su número sea 3.

   De manera similar para Beto tenemos que $\hat{K}_b b_2 \wedge \hat{K}_b b_4$.

6. $K_a(\neg gana_a \wedge \neg gana_b)$

   Notemos que sólo si uno de los números es 0, alguien puede saber con certeza cuál es su propio número: el agente que ve el 0 sabe que él (o ella) tiene el 1. Ana considera dos situaciones posibles: una en la que es cierto que $a_1 \wedge b_2$ *y otra en la que es cierto que* $a_3 \wedge b_2$. *Como en ninguna de las situaciones alguien tiene el 0 en la frente, entonces Ana sabe que no hay ganador.*

   Similarmente, tenemos que $K_b(\neg gana_b \wedge \neg gana_a)$.

7. $E_{\{a,b\}}\neg a_5 \wedge \neg E_{\{a,b\}}E_{\{a,b\}}\neg a_5$

   Tanto Ana como Beto saben que Ana no tiene el 5 (por que Beto sabe que Ana tiene un 3 y Ana sabe que tiene un 1 o un 3). Sin embargo, no todos saben que todos saben esto. Como tenemos que $\hat{K}_b\hat{K}_a a_5$: *Beto piensa que es posible la situación en la que Ana tiene un 3 y Beto tiene un 4, en cuyo caso, Ana podría ver el 4 de Beto por lo que consideraría posible que ella tuviera ¡un 5! Lo cual no es cierto. Una vez más, una semantica clara puede ayudar a verificar este tipo de situaciones.*

8. $E_{\{a,b\}}(\neg gana_a \wedge \neg gana_b) \wedge \neg E_{\{a,b\}}E_{\{a,b\}}(\neg gana_a \wedge \neg gana_b)$

   Aunque todos saben que ninguno de los agentes puede ganar (se sigue de 6.), no es el caso que este hecho sea sabido por todos; Notemos que

> *Ana piensa que es posible que ella tenga el número 1 y que Beto tenga el número 2, en cuyo caso Beto consideraría posible que el número de Ana fuera 1 y el suyo fuera 0. Pero en este caso, Ana sabría cuál es su propio número:* $\hat{K}_a\hat{K}_b(a_1 \wedge K_a a_1)$.

Una versión de lógica epistémica más completa que la que presentamos en la sección anterior es la siguiente.

Definición 2.4.2 *(Lenguaje de lógica epistémica con conocimiento común* $\mathcal{LE}^C$*). El lenguaje de la lógica epistémica con conocimiento común* $\mathcal{LE}^C$ *se define siguiendo las reglas que definen al lenguaje de la lógica epistémica* $\mathcal{LE}$ *(reglas 1 a 5, ver definición 2.2.1) más las reglas que permites crear nuevas fórmulas utilizando los operadores* E *y* C*. Esto es:*

$$\begin{array}{lll}
\top \in \mathcal{LE}^C & & \\
\textit{Si } p \in \Phi & \textit{entonces} & p \in \mathcal{LE}^C \\
\textit{Si } \varphi \in \mathcal{LE}^C & \textit{entonces} & \neg\varphi \in \mathcal{LE}^C \\
\textit{Si } \varphi \in \mathcal{LE}^C \textit{ y } \psi \in \mathcal{LE}^C & \textit{entonces} & \varphi \vee \psi \in \mathcal{LE}^C \\
\textit{Si } i \in \mathcal{A} \textit{ y } \varphi \in \mathcal{LE}^C & \textit{entonces} & K_i\varphi \in \mathcal{LE}^C \\
\textit{Si } \mathcal{B} \subseteq \mathcal{A} \textit{ y } \varphi \in \mathcal{LE}^C & \textit{entonces} & E_{\mathcal{B}}\varphi \in \mathcal{LE}^C \\
\textit{Si } \mathcal{B} \subseteq \mathcal{A} \textit{ y } \varphi \in \mathcal{LE}^C & \textit{entonces} & C_{\mathcal{B}}\varphi \in \mathcal{LE}^C
\end{array}$$

En notación BNF, las fórmulas $\varphi \in \mathcal{LE}^C$ *se definen con las siguientes reglas:*

$$\varphi ::= \top \mid p \mid \neg\varphi \mid (\varphi \vee \psi) \mid K_i\varphi \mid E_{\mathcal{B}}\varphi \mid C_{\mathcal{B}}\varphi$$

Donde $p \in \Phi$, $i \in \mathcal{A}$, $\mathcal{B} \subseteq \mathcal{A}$ *y* $\varphi, \psi \in \mathcal{LE}^C$.

$K_a\varphi$ se leerá como "El agente a sabe que φ".

$E_{\mathcal{B}}\varphi$ se leerá como "Todos los agentes en $\mathcal{B}$ saben que φ".

$C_{\mathcal{B}}\varphi$ se leerá como "Es conocimiento común entre los agentes en $\mathcal{B}$ saben que φ".

Para definir la semántica del operador C, recordemos que intuitivamente, el hecho de que φ sea conocimiento común entre los agentes que pertenecen a $\mathcal{B}$ quiere decir que además de que todos los agentes del conjunto $\mathcal{B}$ saben φ, todos saben que todos saben φ.

Sean $i \in \mathcal{A}$, $\mathcal{B} \subseteq \mathcal{A}$ y $\varphi \in \mathcal{LE}^C$. Tenemos lo siguiente:

- El operador modal K es en realidad un operador de necesidad en lógica multimodal (con una modalidad por agente):

$$K_i\varphi \equiv_{def} [i]\varphi$$

- Las fórmulas con el operador E son abreviaturas (en caso de que $\mathcal{B}$ sea finito):

$$E_A\varphi \equiv_{def} \bigwedge_{i\in\mathcal{B}} K_i\varphi$$

- El conocimiento común es una noción distinta:

$$C_{\mathcal{B}}\varphi \Rightarrow K_i\varphi \wedge K_iK_i\varphi \wedge \cdots \wedge K_jK_i\varphi \wedge \cdots \forall i,j \in \mathcal{B}$$

Definición 2.4.3 *(Semántica para lenguaje epistémico con conocimiento común). Consideremos la estructura acentuada* (M, w) *donde tenemos el siguiente marco de Kripke:* $M = \langle W, R, \mathcal{B}\rangle$ *con* $R \subseteq W \times \mathcal{B} \times W$*,* $\mathcal{B}$ *un conjunto de agentes,* $w \in W$*, y una función evaluación* $e : P_0 \times W \to \{V, F\}$*. Entonces*

$(M, w) \models p$ *sii* $e(p, w) = V \ \forall p \in \Phi$
$(M, w) \models K_i\varphi$ *sii* $\forall v \in W$*, si* $(w, v) \in R_i$ *entonces* $(M, w) \models \varphi$
$(M, w) \models E_{\mathcal{B}}\varphi$ *sii* $\forall i \in \mathcal{B}\ \forall v \in W$*, si* $(w, v) \in R_i$ *entonces* $(M, w) \models \varphi$
$(M, w) \models C_{\mathcal{B}}\varphi$ *sii* $\forall v \in W$*, si* $(w, v) \in R^*_{\mathcal{B}}$ *entonces* $(M, w) \models \varphi$.

La relación $R^*_{\mathcal{B}} \subseteq W \times W$ *se define inductivamente:*
$(w, w) \in R^0_{\mathcal{B}}$ *sii* $w \in W$
$(w, v) \in R^1_{\mathcal{B}}$ *sii* $\exists i \in \mathcal{B}$ *tal que* $(w, v) \in R_i$
$(w, v) \in R^{n+1}_{\mathcal{B}}$ *sii* $\exists u \in W$ *tal que* $(w, u) \in R^n_i \wedge (u, v) \in R^1_{\mathcal{B}}$
$(w, v) \in R^*_{\mathcal{B}}$ *sii* $\exists n \in \mathbb{N}$ *tal que* $(w, v) \in R^n_{\mathcal{B}}$.

Veamos ahora el sistema de demostración para la lógica epistémica.

Definición 2.4.4 *(Sistema de demostración para la lógica epistémica) Un sistema axiomático para la lógica epistémica* $\mathcal{LE}^C$ *contiene los axiomas de* $S5$ *correspondientes al operador* K *en lugar del operador de necesidad tradicional* $(\Box)$.

K $K_i(\varphi \to \psi) \to (K_i\varphi \to K_i\psi)$
T $K_i\varphi \to \varphi$
4 $K_i\varphi \to K_iK_i\varphi$
5 $\neg K_i\varphi \to K_i\neg K_i\varphi$.

2.5. Propiedades del conocimiento

Como es bien sabido, la noción de conocimiento ha sido ampliamente discutida por siglos sin llegar a un acuerdo definitivo. Si el mismo concepto no está claramente definido, es fácil darse cuenta de que cualquier propuesta de formalización va a ser susceptible de enfrentarse a problemas relacionados con aspectos que no logran ser capturados. Por lo anterior, es muy importante tener claro, dada una cierta formalización, qué nociones son las que logra capturar y cuáles no, así como las razones e implicaciones que tienen que ver con dichas nociones.

Ante la cuestión que acabamos de plantear, es natural hacerse la pregunta sobre qué tanto se apega la definición presentada en este capítulo a la idea intuitiva que tenemos de conocimiento. De acuerdo a la semántica presentada en la sección anterior, tenemos que para las fórmulas del tipo $K_i\varphi$, un agente conoce φ si y sólo si φ es verdadera en todos los mundos que el agente i considera posibles. Veamos dos propiedades que obtiene el conocimiento como consecuencia de la definición dada, revisando algunos esquemas válidos (def. 2.3.4).

- Axioma de Distribución (Axioma K).

$$(K) \ \models (K_i\varphi \wedge K_i(\varphi \to \psi)) \to K_i\psi$$

Notemos que esta forma del axioma K es equivalente a la presentada en la definición del sistema de demostración para $\mathcal{LE}^{\mathcal{C}}$.

Demostración. Sea K la fórmula $K_i(\varphi \to \psi) \to (K_i\varphi \to K_i\psi)$, y sea K' la fórmula $(K_i\varphi \wedge K_i(\varphi \to \psi)) \to K_i\psi$.

$K \to K'$

1. $K_i(\varphi \to \psi) \to (K_i\varphi \to K_i\psi)$ (hip 1)
2. $K_i\varphi \wedge K_i(\varphi \to \psi)$ (hip 2)
3. $K_i(\varphi \to \psi)$ (conj a 2)

4. $K_i\varphi \to K_i\psi$ (MP a 1 y 3)
5. $K_i\varphi$ (conj a 2)
6. $K_i\psi$ (MP a 4 y 5)

$\therefore (K_i\varphi \land K_i(\varphi \to \psi)) \to K_i\psi$

$K' \to K$

1. $(K_i\varphi \land K_i(\varphi \to \psi)) \to K_i\psi$ (hip 1)
2. $K_i(\varphi \to \psi) \land K_i\varphi$ (hip 2)
3. $K_i\psi$ (MP a 1 y 2)
4. $\neg K_i\varphi \lor K_i\psi$ (AD a 3)
5. $K_i\varphi \to K_i\psi$ (Equiv de la implicación)

$\therefore K_i(\varphi \to \psi) \to (K_i\varphi \to K_i\psi)$

■

La interpretación natural de este axioma es que todo agente conoce las consecuencias lógicas de su conocimiento. Si un agente sabe φ y también sabe $\varphi \to \psi$, entonces él sabe ψ. Esto es claro ya que si φ y $\varphi \to \psi$ son verdaderas en todos los mundos que el agente i considera posibles, entonces ψ debe de ser verdadera en cada uno de estos mundos.

- Regla de Generalización del conocimiento.

$$\text{Si } M \models \varphi \text{ entonces } M \models K_i\varphi$$

Esta propiedad expresa el hecho de que cada agente conoce toda fórmula válida en alguna estructura dada M. Recordemos que para que φ sea válida en una estructura $M = \langle W, R_i, V\rangle$, es necesario que $(M, w) \models \varphi$ para todo $w \in W$. Si φ es verdadera en todo en todo mundo w, entonces en particular es verdadera en todos los mundos que el agente i considera posibles por lo que el agente debe saberla. Ahora veamos qué podemos decir respecto a la relación de accesibilidad entre mundos posibles.

2.6. La relación de accesibilidad

La relación de accesibilidad en lógica modal puede tener diferentes propiedades. Por ejemplo, puede ser reflexiva ($(w, w) \in R_i, \forall w \in W$), transitiva (Si $(w, u), (u, v) \in R_i$ entonces $(w, v) \in R_i, \forall w, u, v \in W$), euclideana (si $(w, u), (w, v) \in R_i$ entonces $(u, v) \in R_i, \forall w, u, v \in W$), serial ($\forall u \in W\ \exists v \in W$ tal que $(u, v) \in R_i$), etc. Al restringir las estructuras a aquellas cuyas relaciones sean de cierto tipo, la definición de conocimiento adquiere nuevas propiedades. Veamos algunas de ellas.

a) Si la relación R_i es *reflexiva*, entonces el mundo actual w pertenece al conjunto de los mundos que cada agente i considera posibles. Esta propiedad es equivalente al axioma **T**; al que se le conoce también como *Axioma de conocimiento* o *Axioma de verdad*.

$$(\mathbf{T})\ K_i\varphi \rightarrow \varphi$$

Demostración. Si $K_i\varphi$ se cumple, significa que φ es cierto en todos los mundos que el agente considera posibles, en particular en w. Esto tiene consecuencia que todo aquello que el agente sabe debe de ser verdad. ■

b) Cuando la relación es *transitiva* entonces se cumple el *Axioma de introspección positiva*, denotado como **4**.

$$(\mathbf{4})\ K_i\varphi \rightarrow K_iK_i\varphi$$

Demostración. Supongamos que $(M, w) \models K_i\varphi$. Dado que $K_i\varphi$ es cierto en w, entonces φ es cierto en todo mundo accesible desde w. Es decir, $(M, u) \models \varphi$ para todo $u \in W$ tal que $(w, u) \in R_i$. Pero como R_i es transitiva, entonces $(w, v) \in R_i$ para todo $v \in W$ tal que $(u, v) \in R_i$. Como $(w, v) \in R_i$, tenemos también que $(M, v) \models \varphi$ para todo $v \in W$ tal que $(u, v) \in R_i$. Ya que φ es verdadera en cualquier mundo accesible desde u, tenemos que $(M, u) \models K_i\varphi$ para cualquier u accesible desde w. Por lo tanto, $(M, w) \models K_iK_i\varphi$. La interpretación natural de lo que dice este axioma es que el agente i es conciente de aquello que sabe. ■

c) Cuando la relación R_i es *simétrica* y *transitiva* entonces se cumple el *Axioma de introspección negativa*, denotado como **5**.

$$(\mathbf{5})\ \neg K_i\varphi \to K_i\neg K_i\varphi$$

Demostración. Supongamos que $(M, w) \models \neg K_i\varphi$. Entonces, $(M, u) \models \neg\varphi$ para un $u \in W$ tal que $(w, u) \in R_i$. Dado que R_i es simétrica, entonces $(v, w) \in R_i$, y como R_i es transitiva, entonces $(v, u) \in R_i$. Como $(M, u) \models \neg\varphi$ y u es accesible desde v entonces $(M, v) \models \neg K_i\varphi$. Esto es cierto para todo v tal que $(w, v) \in R_i$. Entonces $(M, w) \models K_i\neg K_i\varphi$. La interpretación natural de lo que dice este axioma es que el agente i es conciente de aquello que no sabe. ■

Los sistemas axiomáticos de lógica modal (incluyendo la lógica epistémica) en general incluyen el axioma **K**, así como las reglas **MP** y **N** (retomaremos esto en el capítulo 4). Por tradición, los sistemas se han nombrado de acuerdo a los axiomas que se cumplen en ellos. El sistema de demostración que hemos presentado hace que la lógica epistémica sea una lógica $S5$ (o $KT45$) pues se cumplen los axiomas **K, T, 4** y **5**. Una observación importante es que en una lógica $S5$ las relaciones de accesibilidad son de equivalencia, es decir, son reflexivas simétricas y transitivas.

Reglas de inferencia:

MP $\frac{\varphi \to \psi, \varphi}{\psi}$

K $K_i(\varphi \to \psi) \to (K_i\varphi \to K_i\psi)$

N $\frac{\varphi}{K_i\psi}$.

Veamos ahora, el planteamiento formal del sistema de revisión de creencias.

Capítulo 3

Revisión de creencias

Ahora que contamos con una herramienta que nos permite modelar el conocimiento para un cierto conjunto de agentes, podemos hacernos la pregunta de cómo la actitud epistémica de éstos cambia a través del tiempo. En el ejemplo 2.4.1 del capítulo anterior podemos ver que tales cambios son importantes; el estado de la información que cada agente posee cambia en función del anuncio de cualquiera de los agentes -notemos que incluso el primer anuncio de Ana ("Yo no conozco los números"$= \varphi$- cambia su estado de información: después de decir φ tenemos que Ana sabe que Beto sabe que φ). Es decir, tenemos que $K_aK_b\varphi$ (de hecho tenemos que $K_aK_bK_aK_b\varphi$, etc.).

Existen diferentes perspectivas para la revisión de creencias. Podríamos preguntarnos por ejemplo, si sólo ocurre que el agente actualiza sus creencias o su conocimiento acerca del mundo, o si puede ser también que dichas actualizaciones ocurran a partir de cambios en el mundo mismo. El estudio en el cual el cambio se debe únicamente a la modificación de la información que tenemos acerca del mundo (que es el caso del sudoku) es lo que en epistemología formal se suele llamar *revisión de creencias*. Cuando la razón por la que tenemos nueva información se debe a cambios en el mundo (como en el caso del dominó o del pókar) lo que tenemos es una *actualización*.

Podemos decir que los estudios formales de revisión de creencias son bastante recientes. Comienzan a partir de los trabajos de los filósofos William Harper (1976) e Isaac Levi (1967, 1977, 1980). Pero es hasta la década de los 80 que se establece como un área de investigación a partir del trabajo de Carlos Alchurrón, Peter Gärdenfors, y David Makinson (1985). Al modelo presentado por estos autores se le conoce como *El modelo AGM*. Una de las más fuertes influencias que recibieron los autores para esta teoría fue la de los trabajos realizados sobre lógicas condicionales, principalmente los resultados sobre condicionales contrafácticos presentados por Lewis (1973). El origen de este trabajo se da a partir la confluencia de los trabajos anteriores de los autores; Gärdenfors estaba interesado en establecer una semántica para los condicionales contrafácticos, pues quería definir reglas que permitieran evaluarlos sin compromiso ontológico (pero sí epistémico) (Gärdenfors 1979). Dichas reglas fueron expresadas en términos de postulados. Por su parte, Alchurrón y Makinson (1981, 1982) tenían como objetivo formalizar la dinámica de los códigos legales. Este trabajo dio origen a una nueva área de investigación y en los años siguientes se desarrollaron nuevos modelos constructivos para funciones de cambio, así como una semántica: el sistema de esferas de Grove (1988), inspirado fuertemente en la semántica de los contrafácticos de Lewis (1973).

Un aspecto muy interesante de esta área de estudio es que en ella participan, entre otros, filósofos que buscan un modelo para el comportamiento de un agente racional con respecto a su dinámica de creencias, o bien para analizar los cambios en las teorías científicas; lógicos interesados en los cambios en conjuntos de proposiciones en lenguajes formales; especialistas en derecho que buscan un modelo que represente la dinámica de las promulgaciones, enmiendas y derogaciones de leyes en los códigos; investigadores de ciencias de la computación interesados en un modelo que formalice la noción de actualizar bases de datos y bases de conocimiento, etc.

Antes de comenzar con un planteamiento formal sobre la revisión de creencias, veamos el siguiente ejemplo que es una variación del ejemplo presentado en Fermé (1999).

Consideremos el siguiente conjunto de proposiciones en lenguaje cotidiano: $p =$"María ha nacido en Bahía de Banderas", $q =$"Ana ha nacido en Bahía de Parita", $r =$ "Dos personas son compatriotas si han nacido en el mismo país". Supongamos que estas proposiciones son toda la información de la que disponemos actualmente sobre María y Ana, y supongamos que recibimos la siguiente nueva información: $s =$"Ana y María son compatriotas".

Si agregamos la nueva información a nuestro conjunto de creencias, obtendremos un nuevo conjunto de creencias que contiene las proposiciones p, q, r y s. Además podemos definir la operación adición como aquella cuyos argumentos son, una proposición y un conjunto, y que da como resultado el mínimo conjunto que incluya a ambos. Esta operación representa el modo más simple de cambiar un conjunto de creencias, pero este proceso no siempre resulta tan sencillo.

Por ejemplo, supongamos que consultando un mapa descubrimos que $t =$ Bahía de Banderas es parte de México, y $u =$ Bahía de Parita es parte de Panamá. Si agregamos t y u a nuestro conjunto original $\{p, q, r, s\}$, el resultado será un conjunto con información contradictoria: Ana y María son compatriotas, pero Bahía de Banderas y Bahía de Parita no están en el mismo país. Entonces, la adición no está reflejando la idea de *actualización consistente*. Si deseamos compatibilizar la información previa con la nueva, alguna parte de nuestro conjunto inicial debe ser abandonada o tal vez, la nueva información debe ser rechazada (total o parcialmente). En el ejemplo que acabamos de presentar, existen varias alternativas: podemos suponer que la información acerca de los lugares de nacimiento de Ana y María es errónea, o tal vez el mapa contuviera errores de impresión, quizá Ana y María no son compatriotas. Cualquiera de estas opciones nos permite resolver el problema de incompatibilidad entre las nuevas y las viejas creencias.

Consecuentemente, podemos definir una operación que toma como argumentos un conjunto y una proposición y arroja un nuevo conjunto consistente. Este nuevo conjunto incluye parcial o totalmente nuestras creencias originales y la nueva proposición (si es que hemos decidido aceptarla). El resultado de esta operación de actualización puede ser expresado como un subconjunto de la operación de adición. Notemos que esta operación de actualización está basada en dos nociones: consistencia y una selección entre los posibles caminos para realizar el cambio.

Otro tipo de cambio es el siguiente. Supongamos que descubrimos que r es incorrecta y, en consecuencia, deseamos eliminarla de nuestro conjunto de creencias, el resultado será un nuevo conjunto al que no pertenecerá r. Queremos que el hecho de que Ana y María sean compatriotas quede indeterminado. Esto no quiere decir que Ana y María no sean compatriotas, sólo quiere decir que ignoramos esa información.

A pesar de el éxito que han tenido los procesos de revisión de creencias en las últimas décadas, aún quedan abiertas muchas preguntas de interés filosófico. Por ejemplo, podríamos preguntarnos si los conjuntos de creencias se actualizan "espontáneamente" o requieren algún tipo de "estímulo.externo. Es decir, si un conjunto de creencias es internamente estable. Por supuesto, esta pregunta tiene respuesta en casos particulares; como mencionamos al principio del capítulo, en el sudoku sólo la información se actualiza, en el pókar la información se actualiza de acuerdo a los cambios que se van presentando en el mundo. Otra pregunta interesante es si los valores epistémicos asignados a las expresiones del lenguaje serán solamente aceptación, rechazo, indeterminación, o se deben considerar grados de aceptabilidad. Tampoco hay acuerdo en general acerca del hecho de que el conjunto deba, necesariamente, ser consistente, etc. En este trabajo no abordaremos estas cuestiones, pero todas ellas son consideraciones importantes que debemos hacer si es que decidimos tomar la vía formal para el estudio de fenómenos epistemológicos.

En la siguiente sección daremos el planteamiento formal de la revisión de creencias con algunas consideraciones acerca de las nociones que capturan, así como de las intuiciones que las originan.

3.1. El enfoque AGM

El actual paradigma en la revisión de creencias es justamente el enfoque AGM. Por esta razón es indispensable presentarlo si queremos tener un buen panorama. Para ello seguiremos el tratamiento de van Ditmarsch (2007), del cual omitimos algunas demostraciones que quedan fuera de los objetivos del presente trabajo. Una de las motivaciones para modelar una teoría de la revisión tiene que ver con la idea general de que las teorías (científicas) tienen la tendencia a que, cuando aparece un fenómeno del cual no pueden dar cuenta, deben realizarse únicamente los mínimos cambios necesarios tales que nos permitan explicar el nuevo fenómeno. Así pues, una de las motivaciones de la revisión de creencias es el *principio de cambio mínimo* (de una teoría).

Dado un conjunto de creencias $\mathcal{K}$ y alguna nueva información φ, la teoría de revisión de creencias AGM, distingue tres tipos de cambio:

- En una *expansión* de $\mathcal{K}$ con φ, el conjunto de creencias resultante (al que denotaremos $\mathcal{K} \oplus \varphi$) admite φ: la evidencia es añadida a las creencias, posiblemente generando inconsistencia.

- Una *contracción* de $\mathcal{K}$ con φ, denotada por $\mathcal{K} \ominus \varphi$, resulta en un conjunto de creencias del cual se elimina φ o, más precisamente, del cual, no se sigue φ.

- Una *revisión* de $\mathcal{K}$ con φ, denotada por $\mathcal{K} \circledast \varphi$, es el resultado de incorporar nueva información φ. El caso más interesante es cuando φ es contradictoria con $\mathcal{K}$. En este caso el resultado será un nuevo conjunto de creencias que contenga a φ y que por lo demás sea “muy similar” a $\mathcal{K}$. Utilizando lo que se conoce como la *identidad Levi*, la revisión puede ser expresada en términos de las dos operaciones anteriores:

Definición 3.1.1 *Identidad Levi*

$$\mathcal{K} \circledast \varphi \equiv (\mathcal{K} \ominus \neg\varphi) \oplus \varphi$$

En otras palabras, a fin de revisar un conjunto de creencias con nueva información (contradictoria) φ, primero "hacemos espacio"para φ quitando aquello con lo que se contradice (es decir, $\neg\varphi$), y entonces podemos agregar de manera *segura* la nueva información φ. A continuación presentaremos un lenguaje para la revisión AMG.

Definición 3.1.2 *Definimos el lenguaje $\mathcal{L}_0$ como el conjunto de fórmulas proposicionales generadas por algún conjunto de fórmulas atómicas Φ y las conectivas clásicas. Sea $C_n(\cdot)$ el operador clásico de consecuencia. Es decir, $C_n(\Sigma) = \{\sigma \in \L_0 | \Sigma \vdash \sigma\}$. Un conjunto de creencias $\mathcal{K}$ es un conjunto de fórmulas proposicionales cerrado bajo $C_n(\cdot)$. Es decir, $C_n(K) = \mathcal{K}$. Denotamos el conjunto único de creencias inconsistente (que consta de todas las fórmulas proposicionales) por $\mathcal{K}_\perp$. Más aún, a menos que se indique lo contrario, en este capítulo, φ, ψ y χ denotan fórmulas proposicionales que representan nueva información.*

Una vez presentado un lenguaje puramente proposicional para las creencias, notemos que los razonamientos acerca de los cambios se realizan en un meta-nivel. Para describir el comportamiento de cada tipo de cambio deberemos presentar algunos postulados. Estos postulados proporcionan las propiedades deseables para las operaciones, las cuales están inspiradas en intuiciones generales acerca del cambio de teorías o de información.

3.1.3. Expansión

Recordemos que la operación $\oplus$ modela la expansión: la idea de $\mathcal{K} \oplus \varphi$ es que éste es el conjunto resultante una vez que aceptamos φ como la nueva información agregada al conjunto de creencias original $\mathcal{K}$. El comportamiento de la expansión se dará de acuerdo a los postulados del cuadro 3.1.

El postulado ($\mathcal{K} \oplus 1$) garantiza que el resultado de una expansión es un conjunto de creencias: esto facilita la realización de expansiones repetidas, y ya que esta condición se aplica para todas las operaciones, esto garantizará que se cumplan propiedades como la identidad Levi (3.1.1) o la conmutatividad de $\oplus$. El segundo postulado para $\oplus$, *acierto*, expresa todo lo que

$(\mathcal{K} \oplus 1)$	$\mathcal{K} \oplus \varphi$ es un conjunto de creencias	tipo
$(\mathcal{K} \oplus 2)$	$\varphi \in \mathcal{K} \oplus \varphi$	acierto
$(\mathcal{K} \oplus 3)$	$\mathcal{K} \subset \mathcal{K} \oplus \varphi$	expansión
$(\mathcal{K} \oplus 4)$	Si $\varphi \in \mathcal{K}$ entonces $\mathcal{K} = \mathcal{K} \oplus \varphi$	acción mínima
$(\mathcal{K} \oplus 5)$	Para toda $\mathcal{H}$, si $\mathcal{K} \subset \mathcal{H}$, entonces $\mathcal{K} \oplus \varphi \subset \mathcal{H} \oplus \varphi$	monotonía
$(\mathcal{K} \oplus 6)$	$\mathcal{K} \oplus \varphi$ es el más pequeño conjunto que satisface $(\mathcal{K} \oplus 1) - (\mathcal{K} \oplus 5)$	cambio mínimo

Cuadro 3.1: Los postulados para la expansión

tiene que ver con la expansión. Una vez que hemos decidido aceptar φ como nueva información, ésta debe ser incorporada a nuestras creencias.

El resto de los postulados refieren a algún tipo de *cambio mínimo*, o *economía informacional*, o *racionalidad*. Para ser más precisos, $(\mathcal{K} \oplus 3)$ dice que, cuando decidimos agregar información nueva a nuestras creencias, no desechamos arbitrariamente otras creencias, por el contrario, intentamos mantener todo el viejo conjunto de creencias (esto podría considrarse un tipo de cambio mínimo: no removemos cualquier cosa de manera innecesaria). Los postulados $(\mathcal{K} \oplus 1) - (\mathcal{K} \oplus 3)$ garantizan, por lo menos, que nada debe ser desechado cuando realizamos una expansión. Es decir que,

$$Cn(\mathcal{K} \cup \{\varphi\}) \subseteq \mathcal{K} \oplus \varphi$$

El postulado de *acción mínima* $(\mathcal{K} \oplus 4)$, reduce al mínimo las circunstancias para cambiar el conjunto de creencias $\mathcal{K}$ una vez que aceptamos la nueva información φ; si φ ya pertenecía a $\mathcal{K}$ entonces el conjunto permanecerá igual. El postulado $\mathcal{K} \oplus 6$ es obviamente acerca del cambio mínimo: una vez que hemos aceptado que una expansión corresponde a la adición de elementos a un conjunto $(\mathcal{K} \oplus 3)$, no debemos agregar nada más que lo absolutamente necesario. Es fácil ver que si aceptamos el resto de los postulados, $(\mathcal{K} \oplus 6)$ es equivalente a

Si $\vdash \varphi \leftrightarrow \psi$ entonces $\mathcal{K} \oplus \varphi = \mathcal{K} \oplus \psi$	extensionalidad
$(\mathcal{K} \oplus \varphi) \oplus \psi = \mathcal{K} \oplus (\varphi \wedge \psi)$	conjunción
$(\mathcal{K} \oplus \varphi) \oplus \psi = (\mathcal{K} \oplus \psi) \oplus \varphi)$	conmutatividad
$(\mathcal{K} \cap \mathcal{H}) \oplus \varphi = (\mathcal{K} \oplus \varphi) \cap (\mathcal{H} \oplus \varphi)$	distributividad

Cuadro 3.2: Algunas propiedades derivadas de la expansión

$$\mathcal{K} \oplus \varphi \subseteq Cn(\mathcal{K} \cup \{\varphi\})$$

El postulado $(\mathcal{K} \oplus 5)$ es similar a el postulado $(\mathcal{K} \oplus 4)$ y se da por razones históricas y conceptuales (van Ditmarsch,& van der Hoek,& Kooi, Berteld, 2007 p.46). El significado que se ha presentado de cada uno de los postulados para la expansión se resume en el siguiente teorema.

Teorema 3.1.4 *Una función $\oplus$ satisface $(\mathcal{K} \oplus 1) - (\mathcal{K} \oplus 6)$ si y sólo si $\mathcal{K} \oplus \varphi = Cn(\mathcal{K} \cup \{\varphi\})$.*

El teorema 3.1.4 es un teorema de representación para la expansión: los postulados permiten dar una definición explícita de $\oplus$. Más adelante veremos de manera breve que éste no es el caso para la contracción ni para la revisión.

Definición 3.1.5 *Dado un conjunto de postulados $(\mathcal{K} \oplus i_1), \ldots, (\mathcal{K} \oplus i_k)$, definimos un postulado de cambio mínimo $(\mathcal{K} \oplus min)^{(i_1,\ldots i_k)}$, como sigue:*

$(\mathcal{K} \oplus min)^{(i_1,\ldots i_k)}$ *$\mathcal{K} \oplus \varphi$ es el mínimo conjunto que satisface*

$$(\mathcal{K} \oplus i_1), \ldots, (\mathcal{K} \oplus i_k)$$

3.1.6. Contracción

Denotaremos con $\mathcal{K} \ominus \varphi$ al conjunto que resulta cuando φ es eliminado como creencia de $\mathcal{K}$. El caso de la contracción es un poco más complejo que

$(\mathcal{K} \ominus 1)$ $\mathcal{K} \ominus \varphi$ es un conjunto de creencias	tipo
$(\mathcal{K} \ominus 2)$ $\mathcal{K} \ominus \varphi \subseteq \mathcal{K}$	contracción
$(\mathcal{K} \ominus 3)$ Si $\varphi \notin \mathcal{K}$ entonces $\mathcal{K} = \mathcal{K} \ominus \varphi$	acción mínima
$(\mathcal{K} \ominus 4)$ Si $\nvDash \varphi$ entonces $\varphi \notin \mathcal{K} \ominus \varphi$	aceptación
$(\mathcal{K} \ominus 5)$ Si $\varphi \in \mathcal{K}$ entonces $\mathcal{K} \subseteq (\mathcal{K} \ominus \varphi) \oplus \varphi$	recuperación
$(\mathcal{K} \ominus 6)$ Si $\vdash \varphi \leftrightarrow \psi$ entonces $\mathcal{K} \ominus \varphi = \mathcal{K} \ominus \psi$	extensionalidad
$(\mathcal{K} \ominus 7)$ $((\mathcal{K} \ominus \varphi) \cap (\mathcal{K} \ominus \psi)) \subseteq \mathcal{K} \ominus (\varphi \wedge \psi)$	conjunción mínima
$(\mathcal{K} \ominus 8)$ Si $\varphi \notin \mathcal{K} \ominus (\varphi \wedge \psi)$ entonces $\mathcal{K} \ominus (\varphi \wedge \psi) \subseteq \mathcal{K} \ominus \varphi$	conjunción máxima

Cuadro 3.3: Los postulados para la contracción

el de la expansión: Supongamos que el conjunto de creencias de un agente es $\mathcal{K} = Cn(\{p, q, r\})$. Supongamos también que este agente quiere contraer p de $\mathcal{K}$. Notemos que removiendo p, al tomar la cerradura bajo consecuencia lógica, el agente no obtiene el resultado deseado pues ¡$Cn(\mathcal{K} \setminus \{p\}) = \mathcal{K}$! (notemos que $(r \rightarrow p) \in \mathcal{K} \setminus \{p\}$ y por lo tanto $p \in Cn(\mathcal{K} \setminus \{p\})$). Más aún, supongamos que, dado $\mathcal{K}$, el agente necesita contraer $p \wedge q$. Obviamente, no parece haber razón para retirar r pero en cuanto a p y q, si lo que queremos es eliminar la conjunción (bajo el régimen de minimalidad) habrá que eliminar sólo una de las dos proposiciones atómicas, sin embargo, no es claro cuál. Aquí es donde la noción de *arraigo* entra en escena: en ocasiones el agente está arraigado a una creencia, es decir, está menos dispuesto a renunciar a ella. Entonces además de representar las creencias del agente, necesitamos no perder de vista las razones para cada una de sus creencias específicas.

Los supuestos subyacientes para para los postulados de $\ominus$ (ver cuadro 3.3) son, que el resultado de una contracción es único (por lo que, en efecto, en el ejemplo de la contracción de $p \wedge q$ tuvo que haber una elección), apegándonos como siempre, a un principio de economía informacional.

Los postulados $(\mathcal{K} \ominus 1) - (\mathcal{K} \ominus 6)$ forman un sistema básico para la contracción, y pueden ser motivados de una manera similar que los postulados para la expansión: $(\mathcal{K} \ominus 1)$ determina el tipo de $\ominus$: el resultado vuelve a ser un conjunto de creencias (no es el conjunto vacío, ni un error, ni un conjunto de conjuntos de creencias, etc.). El postulado $(\mathcal{K} \ominus 2)$ especifica que, cuando contraemos, nada debe de ser agregado al conjunto de creencias; $(\mathcal{K} \ominus 3)$ es similar a $(\mathcal{K} \oplus 4)$ y estipula que ninguna acción debe ser realizada si el resultado deseado ya se ha conseguido, y $(\mathcal{K} \ominus 6)$ dice que el resultado de una contracción no debe depender de la presentación sintáctica de la proposición que queremos contraer. Finalmente, el postulado de *recuperación*, o $(\mathcal{K} \ominus 5)$ nos da una conexión entre la expansión y la contracción: Supongamos que un agente cree φ, y también ψ. Entonces, si el agente retira la creencia φ puede ser que esté forzado a retirar ψ también, pero (por el principio del mínimo cambio) sólo si es necesario. Esto implica que tan pronto como el agente decida agregar φ nuevamente, ψ deberá ser aceptado también.

Desde luego, por $(\mathcal{K} \ominus 1)$ uno no puede contraer exitosamente una tautología, pues "el mínimo" conjunto de creencias es $Cn(\emptyset)$: el conjunto de todas las tautologías. El cual no es el conjunto vacío: no puede ser que un agente no tenga absolutamente ninguna creencia. Por lo tanto, puede haber sido más apropiado llamar a $(\mathcal{K} \ominus 4)$ *aceptación condicional*. Esto parece sugerir que tenemos la siguiente igualdad: $\mathcal{K} \ominus \top = \mathcal{K}$.

Los postulados $(\mathcal{K} \ominus 7)$ y $(\mathcal{K} \ominus 8)$ limitan el comportamiento de $\ominus$ cuando contraemos una conjunción. Una primera pregunta que podemos plantearnos es si cuando contraemos $\varphi \wedge \psi$ podemos también contraer ambos elementos de la conjunción de manera independiente. Es decir, si $\mathcal{K} \ominus (\varphi \wedge \psi)$ resulta en el mismo conjunto que obtenemos al realizar $\mathcal{K} \ominus \varphi$ y $\mathcal{K} \ominus \psi$, en términos de conjuntos, la pregunta es si la siguiente igualdad se cumple

$$\mathcal{K} \ominus (\varphi \wedge \psi) = (\mathcal{K} \ominus \varphi) \cap (\mathcal{K} \ominus \psi) \; (*)$$

Por el principio de minimalidad, $\subseteq$ no se cumple en general: si se tiene que remover $\varphi \wedge \psi$, no parece claro que necesariamente haya que remover tanto φ como ψ: Si yo creo que Juan (un mal estudiante) pasó un examen particular (llamemos a esta creencia φ), y también que Pedro (un buen estudiante) pasó este mismo examen (llamemos a esta creencia ψ), al enterarme de que no es el caso que ambos pasaron el examen es muy posible que me incline a sólo remover la creencia φ.

Notemos que la dirección contraria de la contención $(\supseteq)$ de $(*)$ es de hecho $(\mathcal{K} \ominus 7)$. Este postulado da un criterio minimal de la contención de

$\mathcal{K} \ominus (\varphi \wedge \psi)$, a saber, que todo lo que se conserva después de retirar φ y retirar ψ, debe pertenecer a $\mathcal{K} \ominus (\varphi \wedge \psi)$. Esto es nuevamente motivado por por un principio de minimalidad: Supongamos que una creencia particular χ se conserva (no importa si la contracción se hizo de φ o de ψ). Entonces χ se debe conservar también si *sólo* contraemos con respecto a la fórmula más fuerte $\varphi \wedge \psi$. La idea es que resulta más *sencillo* retirar una fórmula lógicamente más fuerte que retirar una más débil.

Sin embargo, lo anterior no significa que tengamos $\vdash \varphi \to \psi \Rightarrow \mathcal{K} \ominus \psi \subseteq \mathcal{K} \ominus \varphi$, de hecho, en general ni siquiera tenemos que

$$\mathcal{K} \ominus \varphi \subseteq \mathcal{K} \ominus (\varphi \wedge \psi) \ (**)$$

Por ejemplo, supongamos que yo creo que Pedro es ligeramente mejor estudiante que Juan. También tengo la creencia de que ambos pasaron el examen (utilicemos φ para Pedro, y ψ para Juan). Ahora supongamos además, que tengo que retirar φ. Es muy posible que yo me quede con mi creencia ψ, con lo cual tendríamos $(**)$, dado que $\psi \in \mathcal{K} \ominus (\varphi \wedge \psi)$, lo que es absurdo: Si tengo que admitir que no todos los estudiantes pasaron, empiezo eliminando al peor.

Finalmente, $(\mathcal{K} \ominus 8)$ da un criterio para la contención de $\mathcal{K} \ominus (\varphi \wedge \psi)$. Podemos motivar este postulado como sigue. Primero, podríamos esperar tener que: Si $\varphi \notin \mathcal{K} \ominus (\varphi \wedge \psi)$, entonces $\mathcal{K} \ominus (\varphi \wedge \psi) = \mathcal{K} \ominus \varphi$. Por que si φ es removida cuando ya se ha removido $(\varphi \wedge \psi)$, el postulado de *aceptación* se cumple y no hay necesidad alguna de remover más, o menos, para obtener $\mathcal{K} \ominus (\varphi \wedge \psi)$. Por ejemplo, si yo creo que Pedro es mejor estudiante que Juan, entonces remover la creencia de que ambos pasaron el examen es lo mismo que remover la creencia de que Juan lo pasó. Sin embargo, en lugar de eso, yo puedo tener la firma creencia de que Pedro y Juan son igualmente buenos estudiantes, en cuyo caso no tendríamos $\mathcal{K} \ominus (\varphi \wedge \psi) = \mathcal{K} \ominus \varphi$. Finalmente, presentemos el cambio por *revisión.*

3.1.7. Revisión

La forma más estudiada de cambio de creencias es justamente la revisión. La idea de la revisión $(\mathcal{K} \circledast \varphi)$ es que el agente, habiendo aceptado que φ sea incorporada en sus creencias, requiere que el cambio del conjunto anterior al nuevo sea de tal forma que (1) el conjunto de sus creencias permanezca consistente (si es que φ y $\mathcal{K}$ lo son), (2) que como resultado del cambio, el agente crea φ, (3) que el agente mantenga sus creencias originales tanto como sea posible, y (4) que las nuevas creencias que adopte sean las menos posibles. Los postulados de revisión son dados en el cuadro 3.4: recordemos que $\mathcal{K}_\perp$ es el conjunto de creencias inconsistente, es decir, el conjunto total del lenguaje. Al igual que en el caso de los postulados $(\mathcal{K}\ominus 1)-(\mathcal{K}\ominus 6)$ para la contracción , los postulados $(\mathcal{K}\circledast 1)-(\mathcal{K}\circledast 6)$ establecen un sistema básico para la revisión. Los postulados $(\mathcal{K}\circledast 1)$ y $(\mathcal{K}\circledast 2)$ tienen una motivación análoga a la que tienen para los otros operadores los postulados *tipo* y *aceptación*, al igual que para el postuldo de *extensionalidad*. El postulado $(\mathcal{K}\circledast 3)$ garantiza que al aplicar la *revisión* con φ no agregamos nada que no agregáramos al aplicar la *expansión* con φ, siendo ésta la "mínima manera"de agregar φ. De acuerdo a $(\mathcal{K}\circledast 4)$ y $(\mathcal{K}\circledast 3)$, si φ es consistente con $\mathcal{K}$, la revisión y la expansión son operaciones equivalentes. Aunque en principio no lo parece, el postulado $(\mathcal{K}\circledast 4)$ también induce minimalidad: dice que no podemos retirar fórmulas de $\mathcal{K}$ innecesariamente al aplicar la revisión. El postulado $(\mathcal{K}\circledast 5)$ garantiza que sólo estemos obligados a aceptar un conjunto inconsistente de creencias si la nueva información es inconsistente.

Si la nueva información φ es consistente con lo que se creía, la revisión no genera ningún problema. Supongamos que $\circledast$ satisface los postulados básicos $(\mathcal{K}\circledast 1)-(\mathcal{K}\circledast 6)$. Entonces, si $\neg\varphi\notin\mathcal{K}$ esto implica que $\mathcal{K}\circledast\varphi=\mathcal{K}\oplus\varphi$. Notemos que esta igualdad es inmediata: Por el postulado $\mathcal{K}\circledast 3$, tenemos que $\mathcal{K}\circledast\varphi\subseteq\mathcal{K}\oplus\varphi$ y por el postulado $\mathcal{K}\circledast 4$, tenemos la otra contención.

El principio de *preservación* para la revisión dice que si uno en principio, no cree el opuesto de la nueva información, entonces no necesita retirar nada cuando adopta la nueva información:

$$\text{Si } \neg\varphi\notin\mathcal{K} \text{ entonces } \mathcal{K}\subseteq\mathcal{K}\circledast\varphi$$

$(\mathcal{K}\circledast 7)$ y $(\mathcal{K}\circledast 8)$ son generalizaciones de $(\mathcal{K}\circledast 3)$ y $(\mathcal{K}\circledast 4)$ para el caso de la revisión iterada. Juntos implican lo siguiente:

Si $\psi\notin\mathcal{K}\circledast\varphi$, entonces la revisión con $\varphi\wedge\psi$ resulta el mismo conjunto que el obtenido con la revisión con φ y la expansión con ψ. Incluso $(\mathcal{K}\circledast 3)$ y $(\mathcal{K}\circledast 4)$ se siguen de $(\mathcal{K}\circledast 7)$ y $(\mathcal{K}\circledast 8)$ si hacemos la débil, y muy natural, suposición de que $\mathcal{K}\circledast\top=\mathcal{K}$.

Veamos ahora, la versión modal de revisión de creencias.

$(\mathcal{K} \circledast 1)$ $\mathcal{K} \circledast \varphi$ es un conjunto de creencias	tipo
$(\mathcal{K} \circledast 2)$ $\varphi \in \mathcal{K} \circledast \varphi$	aceptación
$(\mathcal{K} \circledast 3)$ $\mathcal{K} \circledast \varphi \subseteq \mathcal{K} \oplus \varphi$	frontera superior
$(\mathcal{K} \circledast 4)$ Si $\neg\varphi \notin \mathcal{K}$ entonces $\mathcal{K} \oplus \varphi \subseteq \mathcal{K} \circledast \varphi$	frontera inferior
$(\mathcal{K} \circledast 5)$ $\mathcal{K} \circledast \varphi = \mathcal{K}_\perp$ ssi $\vdash \neg\varphi$	trivialidad
$(\mathcal{K} \circledast 6)$ Si $\vdash \varphi \leftrightarrow \psi$ entonces $\mathcal{K} \circledast \varphi = \mathcal{K} \circledast \psi$	extensionalidad
$(\mathcal{K} \circledast 7)$ $\mathcal{K} \circledast (\varphi \wedge \psi) \subseteq (\mathcal{K} \circledast \varphi) \oplus \psi$	iteración $(\mathcal{K} \circledast 3)$
$(\mathcal{K} \circledast 8)$ Si $\neg\psi \notin \mathcal{K} \circledast \varphi$ entonces $(\mathcal{K} \circledast \varphi) \oplus \psi \subseteq \mathcal{K} \circledast (\varphi \wedge \psi)$	iteración $(\mathcal{K} \circledast 4)$

Cuadro 3.4: Los postulados para la revisión

3.2. Revisión de creencias con lógica modal

Como ya habíamos mencionado, con frecuencia en epistemología, la noción de creencia está dada en términos de mundos posibles. Por esta razón es que resulta pertinente presentar una semántica de revisión de creencias en términos de mundos posibles. Entonces tenemos que, una posible definición de verdad sobre una creencia es la siguiente:

$$(M, w) \models B_i\varphi \text{ si y sólo si } \forall v((w, v) \in R \rightarrow (M, v) \models \varphi)$$

Al igual que en lógica epistémica, el operador B_i puede definirse de distintas maneras ya que puede tener por objeto capturar distintas intuiciones acerca de la noción de creencia. En nuestro modelo del capítulo 4 utilizaremos un operador similar cuyo sentido es “considerar posible”, que no necesariamente es lo mismo que creer. En general, el principio rector detrás de una semántica de mundos posibles para la revisión de creencias es que, *en la medida en que menos cree, un agente considera más mundos posibles* y *vice versa*. Los casos extremos en el contexto de la creencia serían que: (i) el agente considera que *no hay mundos posibles*, esto modela el hecho de que él cree *todo*, incluyendo todas las proposiciones p y $\neg p$; (ii) el agente considera *un mundo posible*, en este caso él es omnisciente con respecto a este mundo: él cree todo lo que hay que decir acerca de este mundo (en particular él cree φ o $\neg\varphi$ para toda φ); y (iii) el agente no descarta ningún mundo, es decir, él considera *todos los mundos posibles*, en cuyo caso, no cree nada, excepto tautologías, como $\varphi \vee \neg\varphi$. En el mundo real los casos extremos son, en efecto, (i) y (iii) pues los conjuntos de creencias pueden ser inconsistentes.

En el caso en el que sólo las fórmulas proposicionales puedan ser creencias, podemos suponer que un mundo posible coincide con una evaluación, y sintácticamente representa cada mundo como un conjunto maximal consistente de fórmulas proposicionales. Dado w, escribimos w_0 para su correspondiente conjunto maximal consistente: $w_0 = \{\varphi \in \mathcal{L}_0 \mid w \models \varphi\}$. Llamaremos a w_0 la *teoría (proposicional) de* w, e inversamente, diremos que la evaluación w es el *(único) modelo para* w_0. Denotaremos con $[[\varphi]]_0$ la extensión proposicional de φ: $[[\varphi]]_0 = \{w \mid w \models \varphi\}$. La extensión de un conjunto de creencias $\mathcal{K}$ es $[[\mathcal{K}]]_0 = \bigcap_{\varphi\in\mathcal{K}}[[\varphi]]_0$. Finalmente, denotaremos el conjunto de todas las posibles evaluaciones por $\mathcal{V}$.

Dado un marco de Kripke M y un estado w, podemos pensar el conjunto de creencias $\mathcal{K}_w$ de un agente como:

$$\mathcal{K}_w = \{\varphi \in \mathcal{L}_0 \mid (M,w) \models B\varphi\} = \bigcap_{(w,v)\in R} v_0$$

De manera contraria, dado el conjunto de creencias de un agente $\mathcal{K}_w$ en w, podemos construir el conjunto de los mundos $R(w)$ que el aparentemente considera posibles desde w:

$$R(w)_{\mathcal{K}} = \{v \in \mathcal{V} \mid v \models \mathcal{K}_w\}$$

Por lo que vimos en 3.1.3, es obvio que la semántica de una expansión debería ser algo así: Como $\mathcal{K} \oplus \varphi$ es simplemente la adición de φ obteniendo $C_n(\mathcal{K} \cup \varphi)$, semánticamnte tenemos que debemos restringir el conjunto de mundos posibles (a este proceso se le conoce como *relativización*):

Definición 3.2.1 *Dado un marco de Kripke M, un mundo w y una relación de accesibilidad R, la relativización de R a $[[\varphi]]_0$ resulta un marco M', un mundo $w' = w$ y una relación R', donde $R' = R \cap (W \times [[\varphi]]_0)$. También escribimos $(M,w) \underline{\oplus}\ [[\varphi]]_0$ para denotar a (M', w').*

El siguiente lema justifica la definición anterior: éste dice que agregar una información φ a las creencias de un agente, semánticamente se reduce a relativizar los mundos que considera posibles en donde se satisface φ.

Lema 3.2.2 *Sea $\mathcal{K}_w$ el conjunto de creencias de un agente desde un mundo w. Entonces:*

$$\psi \in \mathcal{K}_w \oplus \varphi \text{ si y sólo si } (M, w) \underline{\oplus} [[\varphi]]_0 \models B\psi$$

Ahora vamos a definir la contracción en términos de mundos posibles. Haciendo una analogía respecto al lema 3.2.2, es de esperarse que un agente que quiere contraer respecto a φ debería agregar algunos mundos a aquellos que considera posibles, a saber, aquellos en los que φ es falsa. La pregunta, desde luego es, ¿cuáles son dichos mundos? Para referir a la contraparte semántica de la relación podemos utilizar un sistema de esferas, también llamado *modelo de repliegue* de teorías respecto a las teorías actuales del agente.

Para capturar esto, un sistema de esferas para $R(w)$ en un marco M con dominio W es un conjunto $Sphe(R(w))$ de esferas $Sp \supseteq R(w)$ con un orden lineal dado por $\supseteq$, con un máximo W y un mínimo $R(w)$, y, para toda φ, si alguna $S \in Sphe(R(w))$ intersecta a $[[\varphi]]_0$ entonces existe el más pequeño (respecto a $\supseteq$) $S' \in Sphe(R(W))$ que intersecta a $[[\varphi]]_0$. Intuitivamente, las esferas que són "más cercanas" a aquellos mundos $R(w)$ que el agente actualmente considera posibles, representan los conjuntos de mundos que él "replegaría"si tuviera que renunciar a algunas de sus creencias. Véase la figura 3.1.

Ahora la contracción puede ser fácilmente definida: dado el conjunto actual $R(w)$ de alternativas para nuestro agente, si el quiere dejar de creer φ, esto significa que $R(w) \cap [[\neg\varphi]]_0$ no es vacío: si lo es, el agente lo relega, es decir, agranda su conjunto de mundos accesibles a la primera esfera $S_p \supseteq R(w)$ que tiene una intersección no vacía con $[[\neg\varphi]]_0$. En la segunda parte de la figura el área sombreada está formada por los mundos que forman el conjunto expandido de $R(w)$, así que φ a dejado de ser una creencia.

Definición 3.2.3 *Dado un marco de Kripke M, un mundo w, una relación de accesibilidad R y un sistema de esferas $Sphe(R(w))$ alrededor de $R(w)$, la contracción semántica $(M, w) \underline{\ominus} [[\varphi]]_0$ es definida como M', w' con $w' = w$ y $M' = \langle W, R' \rangle$, donde $R'(w) = S$, con $S = \min\{S' \in Sphe(R(w)) \mid S' \cap [[\neg\varphi]]_0 \neq \emptyset\}$, y para toda $v \neq w$, $R'(v) = R(v)$.*

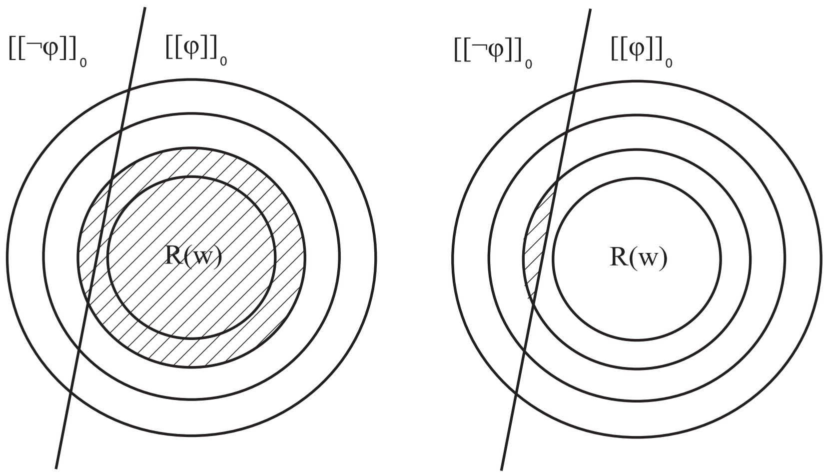

Figura 3.1: Semántica de mundos posibles para el cambio de información

La contraparte semántica de la revisión está dada por la identidad Levi (3.1.1): un agente que quiere revisar con respecto a φ primero "va a" la primera esfera S que intersecta a $[[\neg\varphi]]_0$, y entonces descarta de S todos esos mundos que aún satisfacen φ. En la primera imagen de la figura, el área marcada representa aquellos mundos que son considerados posibles después de la revisión respecto a $\neg\varphi$.

Con el fin de tener las tres definiciones de cambio de información en términos de mundos posibles, presentamos una definición semántica de $\underline{\circledast}$.

Definición 3.2.4 *Dado un marco de Kripke M, un mundo w, una relación de accesibilidad R y un sistema de esferas $Sphe(R(w))$ alrededor de $R(w)$, sea $\underline{\ominus}$ como se definió en 3.2.3. Entonces (M, w) $\underline{\circledast}$ $[[\varphi]]_0$ es definida como $\langle W, R''\rangle, w'$, donde $R''(w'') = R'(w') \cap [[\neg\varphi]]_0$ y (M, w) $\underline{\ominus}$ $[[\neg\varphi]]_0 = \langle W, R'\rangle, w'$ y $w'' = w'$.*

Una vez que hemos presentado un panorama general sobre cómo trabaja la epistemología formal, procederemos a construir nuestro propio marco con el cual, modelaremos el conocimiento en el sudoku.

Capítulo 4

El conocimiento en el sudoku

4.1. Creencias en el sudoku

El sudoku es un juego de tipo rompecabezas que se cree que fue creado hace poco más de tres décadas. Se compone de una tabla de 9×9 casillas, compuesta por subtablas de 3×3 a las que llamaremos subcuadrículas.

El objetivo del juego es rellenar las casillas vacías con un número en cada una de ellas, de tal forma que en cada columna, fila y subcuadrícula aparezcan los números 1,2,3,4,5,6,7,8 y 9 una y sólo una vez. Un juego bien planteado tiene una única solución. Al principio del juego, algunas casillas exhiben los números que les corresponden, a estos números les llamaremos *pistas iniciales*. Notemos que el conocimiento inicial puede ser descrito por proposiciones del tipo *El número 4 corresponde a la casilla ubicada en la fila h, columna v, subcuadrícula c.*

A partir del conocimiento que está dado por las pistas iniciales, el procedimiento mediante el cual se resuelve el sudoku, consiste en ir formando conjeturas (adquiriendo creencias) acerca de qué número corresponde a cada casilla. Cuando construimos una conjetura, si ésta está en lo correcto, en principio nuestra creencia acerca de la verdad de la conjetura se irá fortaleciendo conforme avance el juego. Por otro lado, cuando la conjetura es falsa, al ir avanzando en el juego sin violar las reglas, llegará un momento en el que no será posible seguir. Por ejemplo, si elaboramos la conjetura: *El número 5 corresponde a la casilla C_{ijk}* (donde C_{ijk} es la casilla ubicada en la fila i, en la columna j y en la subcuadrícula k), y ésta resulta ser verdadera, al avanzar el juego, la creencia se fortalecerá; digamos que hemos llenado la subcuadrícula

k sin generar ninguna contradicción con el resto de los números colocados sobre el tablero, este hecho nos proporciona buenas razones para fortalecer la creencia de que colocamos el número en la casilla correcta. Por el contrario, si la conjetura es falsa, quiere decir que el número 5 no corresponde a la casilla ubicada en la intersección de la fila i con la columna j y la subcuadrícula k. Pero este hecho nos da aún más información. Sabemos también que si éste es el caso, el número 5, necesariamente debe corresponder a alguna casilla en la fila i de las que no intersectan a la fila j. Análogamente, tiene que haber un número 5 en alguna de las casillas de la columna j que no intersectan a la fila i. Considerando la situación de la subcuadrícula, es muy probable que una vez que sabemos que el número 5 no corresponde a la casilla C_{ijk}, tengamos conocimiento acerca de la configuración de la misma subcuadrícula. Por ejemplo, podríamos tener algún tipo de conocimiento disyuntivo; digamos que, al saber que el número 5 no corresponde a la casilla C_{ljk}, esto nos podría conducir a la deducción de que a la casilla C_{ijk} le corresponde necesariamente o bien el número 4, o bien el número 3. Que coloquemos el 5 en una casilla equivocada implica que si seguimos las reglas del juego, después de colocar el número 5 en la casilla C_{ijk}, eventualmente tendremos dos veces al número 5 en el mismo renglón, o en la misma columna, o en la misma subcuadrícula. En este caso, debemos detenernos y regresar por el camino de conjeturas hasta descubrir cuál fue la falsa creencia que nos condujo a una conjetura errónea, y entonces eliminar la cadena que se construyó a partir de ella.

El sudoku es un juego que presenta un fenómeno interesante de interacción de creencias. Incluso podemos afirmar que el éxito del juego depende de la interacción entre dichas creencias; a diferencia de otros juegos cuya solución no está determinada o en los que interviene el azar. Así pues, proponemos un proyecto para modelar el fenómeno epistémico del sudoku dentro del campo de la epistemología formal. Este proyecto consta de tres etapas, de las cuales, la primera es la que se desarrolla en el presente trabajo y en particular en este capítulo. Las etapas del proyecto son las siguientes:

1. Separar los aspectos computacional y epistémico del sudoku. Esta etapa del proyecto es importante por que la distinción entre ambos aspectos es necesaria para poder aplicar las dos etapas siguientes. Por ejemplo, para la etapa 2, se propone utilizar herramientas formales como revisión de creencias y lógica dinámica epistémica. Estas herramientas son aplicables a las proposiciones epistémicas.

2. Definir un estilo de actualización de modelos. Esto quiere decir que con el fin de modelar la interacción entre las creencias, necesitamos que nuestro modelo sea capaz de actualizarese ante la aparición de nueva información (por ejemplo, cada vez que agregamos un número). Una propuesta para desarrollar esta etapa es utilizar revisión de creencias o lógica dinámica epistémica.

3. Construir un modelo para resolver de forma eficiente un sudoku considerando el aspecto epistémico del juego. Es decir, un modelo cuya función sea guíar el proceso de actualización hacia la solución del juego. En teoría de la computación se han desarrollado algoritmos para poder resolver un juego de sudoku. Sin embargo ninguno de ellos ha tomando en cuenta el aspecto espistémico. Este proyecto no sólo nos conduciría a la solución del juego, sino que lo haría tomando en cuenta que el juego es resuelto por un agente epistémico.

Además, estas etapas deben (o pueden) incluir herramienta formal que modele otros aspectos del fenómeno epistémico. A lo largo del desarrollo puede incluirse alguna teoría que ayude a determinar los grados de creencia de una proposición en un momento dado del juego. Esto podría ser de gran utilidad, por ejemplo, para la tercera etapa del proyecto.

Así pues, nuestro objetivo a continuación es presentar un modelo que distinga el aspecto epistémico del aspecto computacional en el juego del sudoku.

Definición 4.1.1 *Definimos las proposiciones atómicas del sudoku como*

$$\Phi' = \{N_{ijk} | N, i, j, k = 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

Donde N_{ijk} es una abreviatura de la proposición "*El número N corresponde a la casilla C_{ijk}*", y la casilla C_{ijk} está determinada por la intersección de la fila h_i y la columna v_j, la cual pertenece a la subcuadrícula s_k.

Las filas se numerarán de arriba hacia abajo, las columnas de izquierda a derecha y las subcuadrículas en el mismo orden que el de los números en el teclado de un teléfono. Notemos que un renglón h_i y una columna v_j bastan para determinar una casilla. Sin embargo es importante, dadas las reglas del juego, observar que

$$h_i \cap v_j \in s_1 \text{ cuando } 1 \leq i \leq 3 \wedge 1 \leq j \leq 3$$
$$h_i \cap v_j \in s_2 \text{ cuando } 1 \leq i \leq 3 \wedge 4 \leq j \leq 6$$

$$h_i \cap v_j \in s_3 \text{ cuando } 1 \leq i \leq 3 \wedge 7 \leq j \leq 9$$

$$h_i \cap v_j \in s_4 \text{ cuando } 4 \leq i \leq 6 \wedge 1 \leq j \leq 3$$

$$h_i \cap v_j \in s_5 \text{ cuando } 4 \leq i \leq 6 \wedge 4 \leq j \leq 6$$

$$h_i \cap v_j \in s_6 \text{ cuando } 4 \leq i \leq 6 \wedge 7 \leq j \leq 9$$

$$h_i \cap v_j \in s_7 \text{ cuando } 7 \leq i \leq 9 \wedge 1 \leq j \leq 3$$

$$h_i \cap v_j \in s_8 \text{ cuando } 7 \leq i \leq 9 \wedge 4 \leq j \leq 6$$

$$h_i \cap v_j \in s_9 \text{ cuando } 7 \leq i \leq 9 \wedge 7 \leq j \leq 9$$

La observación anterior muestra qué conjunto de casillas determinadas por las intersecciones entre filas y columnas corresponden a cada una de las subcuadrículas.

Mediante una conjunción de proposiciones atómicas es posible describir en cualquier momento la situación actual del juego. Sin considerar por supuesto, las creencias o la serie de pasos que derivaron en dicha situación. Por ejemplo, si tenemos la siguiente configuración en un juego de sudoku

		6		3	4			
9	4	8		5		3		
7							2	
	6		2	7				
					6		8	5
			6			4		
		5		8				
		7			3	5		6

queda descrita por la siguiente conjunción de proposiciones atómicas

$$6_{131} \wedge 3_{152} \wedge 4_{162} \wedge 9_{211} \wedge 4_{221} \wedge 8_{231} \wedge 5_{252} \wedge 3_{273} \wedge 7_{311} \wedge 2_{383} \wedge 6_{524} \wedge 2_{545} \wedge$$

$$7_{555} \wedge 6_{665} \wedge 8_{686} \wedge 5_{696} \wedge 6_{747} \wedge 4_{779} \wedge 5_{837} \wedge 8_{858} \wedge 7_{937} \wedge 3_{968} \wedge 5_{979} \wedge 6_{999}$$

El problema de esta proposición es que no nos permite distinguir las pistas iniciales de los números que han sido colocados durante la evolución del juego. Tampoco nos dice nada acerca de los criterios con los que fueron colocados los números que no forman parte de las pistas iniciales y por tanto, no nos da información alguna sobre el grado de certeza con el que podemos asegurar que cada uno de los números mostrados corresponde a la casilla en la que se encuentra.

Antes de dar el lenguaje que utilizaremos es importante hacer algunas observaciones. La primera de ellas es que, debido a que un juego bien planteado tiene una única solución, las proposiciones atómicas siempre tienen un valor de verdad. Es decir, una proposición del tipo $5_{142} =$"El número 5 corresponde a la casilla C_{142}"va a ser verdadera o falsa desde el principio hasta el último momento del juego. Lo que sí varía es nuestro conocimiento acerca del valor de verdad de la proposición; mientras la casilla C_{142} permanezca en blanco, es posible que desconozcamos el valor de verdad de 5_{142}. Podríamos tener conjeturas con mayor o menor grado de certeza al respecto, pero lo importante es notar que el valor de verdad de la proposición original no cambia, pues en el juego, los hechos siempre son los mismos, y lo único que va cambiando es la información que poseemos acerca de ellos conforme vamos "descubriendo"la correspondencia entre los dígitos y las casillas.

La observación presentada en el párrafo anterior es una muy buena motivación para utilizar lógica epistémica, ya que ésta permite capturar propiedades no sólo acerca de los hechos del mundo, sino también del conocimiento que poseemos de ellos. Otra motivación es que para poder tomar las decisiones no basta con conocer los números que en un determinado tiempo t se exhiben en el tablero. Es necesario saber cuándo sabemos que determinado número corresponde a una casilla y cuándo simplemente lo hemos conjeturado, incluso, es útil tener una noción del grado de certeza con el que están dispuestos los números en las casillas que se exhiben (en este trabajo no trataremos de resolver el problema de determinar el grado de certeza de las conjeturas; esto correspondería a la segunda parte del proyecto general). Por ejemplo, sería una pérdida de tiempo si al momento de tomar la decisión para la siguiente jugada, cuestionaramos de igual forma la verdad de una pista inicial que la verdad de una proposición que representa una conjetura cuyo fin es sólo el de ir eliminando posibilidades como estrategia para resolver el juego.

Así pues, para modelar el conocimiento durante el juego, necesitamos dar cuenta del valor de verdad de las proposiciones de manera relativa a lo que podemos saber (es decir, en función de los números que se exhiben) pues los mundos posibles se darán en función de la información con la que contamos. Notemos que si consideramos un mundo posible como un arreglo parcial del tablero en el que sólo se muestran los números correspondientes a ciertas casillas, habrá proposiciones sobre las que no podemos saber su valor de verdad. Esto plantea la necesidad de utilizar otro tipo de lógica (distinta a la habitual) para nuestra función evaluación; en este caso, vamos a utilizar una *lógica multimodal epistémica trivaluada.*

4.2. Lógica modal multivaluada

Las lógicas modales multivaluadas han sido estudiadas durante las últimas cuatro décadas (Segerberg 1967, Kripke 1975, Schotcg & Jense & Larsen & MacLellan 1978, Morikahua 1989, etc.), con el proposito de aumentar los grados de libertad no sólo en cuanto a las modalidades que una lógica puede tener, sino también en cuanto a los valores de verdad que ésta puede asignar. En particular en (Fitting, 1992), se presentan dos distintos enfoques para una lógica modal multivaluada: uno es utilizando marcos de Kripke que permiten a las fórmulas tomar valores en una lógica finita multivaluada en cada mundo posible, y el otro es una generalización del primero que permite también que la relación de accesibilidad entre mundos sea multivaluada. Para elaborar nuestro modelo basta con plantear el primer enfoque.

4.2.1. Sintaxis básica y semántica

Antes de abordar las cuestiones modales, es necesario hacer una breve contextualización en lo que son las lógicas multivaluadas. La lógica multivaluada puede ser considerada como una generalización de lógica clásica en la que los valores de verdad constituyen una lattice finita. Una lattice (o retículo) es un conjunto parcialmente ordenado (es decir un conjunto con una relación refexiva, antisimétrica y transitiva) en el que para cualesquiera dos elementos existe un supremo (una mínima cota superior). A partir de esto podemos observar que si bien, las nociones de conjunción y disyunción son las usuales, no necesariamente ocurre lo mismo con la negación. Por su parte, la implicación juega un papel intermedio: está definida y es interpretada bajo el orden parcial de valores de verdad de la lattice, sólo debemos tener presente que ésta en general no puede ser anidada (es decir, debemos tener cuidado con fórmulas del tipo $((a_1 \to a_2) \to a_3) \to a_4$).

Definición 4.2.2 *Definimos una lattice $\mathcal{T}$ de valores de verdad como el espacio constituido por el conjunto de estos valores y un orden parcial al que* *denotaremos $\leq$. También vamos a agregar los operadores $\wedge$ y $\vee$. Al nivel más bajo de $\mathcal{T}$ lo denotamos como falso y al más alto como verdadero. Suponemos que $falso \neq verdadero$.*

En lógica clásica las fórmulas $P \wedge \neg P$ y $P \vee \neg P$ sirven para definir dentro del lenguje contrapartes de *falso* y *verdadero*. En una lógica más general multivaluada como la que estamos considerando, no es suficiente con la maquinaria básica para definir contrapartes de todos los valores de verdad, por lo que las constantes proposicionales son agregadas explícitamente al lenguaje. Para ver esto de manera más simple, supongamos que los mienbros de $\mathcal{T}$ pueden ser usados ellos mismos como fórmulas atómicas.

Definición 4.2.3 *El lenguaje proposicional $L_{\mathcal{T}}$ está dado por:*

- **Fórmulas atómicas** *de $L_{\mathcal{T}}$: se dividen en variables proposicionales denotadas por P, Q, etc., y constantes proposicionales, las cuales son miembros de $\mathcal{T}$.*

- **Fórmulas** *de $L_{\mathcal{T}}$ que se construyen con las siguientes reglas: Las fórmulas atómicas de $L_{\mathcal{T}}$ son fórmulas de $L_{\mathcal{T}}$. Si A y B son fórmulas de $L_{\mathcal{T}}$, entonces $(A \wedge B)$ y $(A \vee B)$ también lo son.*

- **Implicaciones** *de $L_{\mathcal{T}}$: son expresiones de la forma $(A \to B)$ donde A y B son fórmulas.*

Para conveniencia del lector las *fórmulas* son denotadas con A, B, etc.; las *implicaciones* son denotadas con X, Y, etc., y los *conjuntos de implicaciones son denotados con* Γ, Δ, etc.

Definición 4.2.4 *Una evaluación es un mapeo (una función continua y suprayectiva) que asigna a cada una de las fórmulas de $L_{\mathcal{T}}$ un elemento de $\mathcal{T}$ y a cada elemento de $\mathcal{T}$ lo manda en él mismo (es decir, la evaluación restringida a $\mathcal{T}$ es la función identidad).*

Si v es una evaluación, su acción se extiende a todas las fórmulas de la manera habitual: $v(A \wedge B) = v(A) \wedge v(B)$ donde el operador $\wedge$ del lado derecho de la igualdad pertenece a $\mathcal{T}$. Similarmente para $\vee$. Entonces $v(A) \in \mathcal{T}$ si A es una fórmula de $L_{\mathcal{T}}$. Finalmente, las implicaciones de $L_{\mathcal{T}}$ son mapeadas a $\mathcal{T}$ de modo que $v(A \to B) = verdadero$ si y sólo si $v(A) \leq v(B)$. Si $v(A) \not\leq v(B)$, el valor exacto de $v(A \to B)$ no importará por ahora. Por supuesto, el siguiente paso más natural es presentar el sistema de demostración y mostrar que el sistema es completo. Sin embargo, omitiremos estos dos aspectos para pasar a la incorporación de la lógica multivaluada a una versión modal.

Ejemplo 4.2.5

Supongamos que $\langle \mathcal{G}, \mathcal{R} \rangle$ es un marco de Kripke. Para tener una estructura de mundos posibles (o modelo de Kripke) es necesario agregar una interpretación v, que mapea parejas ordenadas de mundos posibles y fórmulas atómicas en valores de verdad (normalmente en falso y verdadero). Entonces v es extendida a todas las fórmulas en la manera usual. En particular $v(w, \Box A) = verdadero$ (es decir, $w, v \models \Box A$) si y sólo si $v(w', A) = verdadero$ (es decir, $w', v \models A$)para toda $w' \in G$ tal que $(w, w') \in \mathcal{R}$. De aquí se sigue que $v(w, \Box A) = falso$ si $v(w', A) = falso$ para alguna $w' \in \mathcal{G}$ con $(w, w') \in \mathcal{R}$.

Ahora supongamos que la interpretación v es parcial; en cada mundo posible, v mapea algunas fórmulas atómicas a *verdadero*, otras a $falso$ y el resto están indefinidas. Necesitamos una forma natural de extender la información incompleta contenida en v a todas las fórmulas. Para las conectivas clásicas es simple: $A \wedge B$ es verdadera en un mundo si tanto A como B son verdaderas en él, y es falsa si alguno de los dos dos es falso. Esto deja sin valor de verdad a $A \wedge B$ en cualquiera de los otros casos; por ejemplo, si A es verdadera pero B carece de valor de verdad en un mundo determinado.

El caso modal puede ser tratado más o menos de la misma manera. Tenemos que $v(w, \Box A) = verdadero$ si $v(w', A) = verdadero$ para toda $w' \in \mathcal{G}$ tal que $(w, w') \in \mathcal{R}$. Tenemos que $v(w, \Box A) = falso$ si $v(w', A) = falso$ para algún $w' \in \mathcal{G}$ tal que $(w, w') \in \mathcal{R}$. En cualquier otro caso, $v(w, \Box A)$ está indefinido. Por ejemplo, si existe algún $w' \in \mathcal{G}$ tal que $(w, w') \in \mathcal{R}$ y que $v(w', A)$ esté indefinido (aún cuando A sea verdadero en todos los demás mundos accesibles desde w), entonces $v(w, \Box A)$ estará indefinido. Una formulación de este estilo de lógica modal parcial fue presentada en (Kripke, 1975), en donde él muestra que su teoría de la verdad puede ser extendida de manera natural.

Pero lo que Kripke presenta es una lógica parcial; hay proposiciones cuyo valor no está definido, y esto podría ser considerado por muchos, en contra del espíritu de la lógica y las matemáticas. Lo que sí viene a cuenta en este momento, es tomar la propuesta de lógica modal parcial de Kripke y reconsiderarla tratando de sustituir el caso del valor indefinido por un tercer valor de verdad (que es de hecho, lo que nos va a servir más adelante). En el caso de la lógica proposicional, éste proceso nos conduce a lo que se conoce como la *lógica trivaluada fuerte de Kleene* (Kleene, 1950).

Definición 4.2.6 *Lógica trivaluada fuerte de Kleene. Consideremos* $\mathcal{T} = \{falso, \perp, verdadero\}$ *con el orden* $falso \leq \perp \leq verdadero$*. Vamos a leer el tercer valor,* $\perp$*, como "desconocido"*o "indeterminado".

Mientras que en la lógica parcial de Kripke, la preocupación era definir los casos en los que las proposiciones son verdaderas y el papel de la negación no parecía relevante, en la lógica de Kleene, la negación juega un papel más importante. La operación negación intercambia $falso$ por $verdadero$ y deja fijo a $\perp$. Simplemente vamos a añadir un operador unario, $\neg$, vamos a pedir que el conjunto de fórmulas sea cerrado bajo este operador y vamos a agregar los siguientes axiomas:

1. $\neg verdadero \to falso$
2. $\neg\perp \to \perp$
3. $\neg falso \to verdadero$
4. $falso \to \neg verdadero$
5. $\perp \to \neg\perp$
6. $verdadero \to \neg falso$

El siguiente paso es combinar la lógica trivauada de Kleene con la propuesta de Kripke para obtener así una lógica modal trivaluada. Lo que Fitting hace en (Fitting, 1992) para introducir el operador modal, es presentar una versión trivaluada de la semántica parcial de Kripke.

Definición 4.2.7 *Definimos una evaluación, v, como un mapeo que va del producto cartesiano de mundos posibles y fórmulas atómicas al conjunto* $\{falso, \perp, verdadero\}$*. El mapeo se extiende a todas las fórmulas mediante la evaluación de los operadores* $\wedge$*,* $\vee$ *y* $\neg$ *en cada mundo posible de acuerdo a la lógica de Kleene y usando la siguiente igualdad para el caso modal:*

$$v(\Box A, w) = \bigwedge\{v(w', A) \mid (w, w') \in \mathcal{R}\}.$$

Esta igualdad ya había sido utilizada en (Schotch & Jensen & Larsen & MacLellan, 1978) y en (Thomason, 1978). Fitting presenta una semántica detallada que es la que utilizaremos más adelante.

Definición 4.2.8 *Definimos $L_{\mathcal{T}}[\Box]$ el lenguaje que es como $L_{\mathcal{T}}$, pero añadiendo la siguiente regla de formación de fórmulas: Si A es una fórmula del lenguaje, entonces $\Box A$ también lo es. De la misma manera, $L_{\mathcal{T}}[\Box, \rightarrow]$ permite la cerradura de fórmulas bajo $\Box$ y bajo $\rightarrow$.*

Como ya se había señalado antes, la semántica es una modificación directa del modelo convencional de mundos posibles.

Definición 4.2.9 *Un modelo trivaluado modal es una estructura $\langle \mathcal{G}, \mathcal{R}, v \rangle$ donde $\mathcal{G}$ es un conjunto no vacío de mundos posibles, $\mathcal{R}$ es una relación en $\mathcal{G}$, y v mapea mundos y fórmulas atómicas a la lattice de tres elementos $\mathcal{T}$ (conservando la condición de que los mienbros de $\mathcal{T}$ son mapeados a ellos mismos).*

El mapeo v se extiende al resto de las fórmulas en la forma usual, con la siguiente condición especial:

$$v(w, \Box A) = \bigwedge \{v(w', A) \mid (w, w') \in \mathcal{R}\}$$

Es fácil ver que si $\mathcal{T}$ es la lattice de dos valores de la lógica clásica, esta semántica se reduce a la semándica usual de la versión de Kripke. Así el operador necesidad ha sido introducido. Podríamos también dentro de esta línea, introducir un operados de posibilidad con la caracterización semántica:

$$v(w, \Diamond A) = \bigvee \{v(w', A) \mid (w, w') \in \mathcal{R}\}.$$

El caso del operados posibilidad no será discutido por el momento. A continuación presentaremos algunas propiedades que cumple esta concepción de lógica modal trivaluada.

Regla de necesitación trivaluada. Para cada $a, b, c \in \mathcal{T}$, y fórmulas A_1, A_2, B. Tenemos que

$$\frac{(a \rightarrow A_1, b \rightarrow A_2) \rightarrow (c \rightarrow B)}{(a \rightarrow \Box A_1, b \rightarrow \Box A_2) \rightarrow (c \rightarrow \Box B)}$$

Para dar una idea del comportamiento de esta lógica modal multivaluada, a continuación presentamos algunas propiedades que se derivan de la regla de necesitación trivaluada. Las demostraciones no se incluyen ya que no serán necesarias para el modelo.

Proposición 4.2.10

1. *Para cada* $t \in \mathcal{T}$, $t \to \Box t$
2. $\frac{A \to B}{\Box A \to \Box B}$
3. $(\Box A \wedge \Box B) \to \Box(A \wedge B)$
4. $\Box(A \wedge B) \to \Box A \wedge \Box B)$

Proposición 4.2.11 *Sea* $a \in \mathcal{T}$ *y* A *una fórmula arbitraria. Entonces:*

1. $(a \to \Box A) \to \Box(a \to A)$
2. $\Box(a \to A) \to (a \to \Box A)$

De esta manera procederemos a dar la semántica para nuestra lógica multimodal epistémica trivaluada.

4.3. Sintaxis para el modelo del Sudoku

Definición 4.3.1 *Definimos* $\mathcal{T}_S$ *como la lattice de valores de verdad para el sudoku conformada por los valores* $\{falso,\ desconocido,\ verdadero\}$*, con el orden* $falso \leq deconocido \leq verdadero$*.*

Definición 4.3.2 *El lenguaje lógico proposicional para el sudoku* $\mathcal{L}_{T_S}$ *está dado por:*

1. $\top \in \mathcal{L}_{T_S}$.
2. *Si* $p \in \Phi$ *(donde* $\Phi = \Phi' \cup \mathcal{T}_S$*) entonces* $p \in \mathcal{L}_{T_S}$.
3. *Si* $\varphi \in \mathcal{L}_{T_S}$ *entonces* $\neg\varphi \in \mathcal{L}_{T_S}$ *(la negación se define como en 4.2.6).*
4. *Si* $\varphi \in \mathcal{L}_{T_S}$ *y* $\psi \in \mathcal{L}_{T_S}$ *entonces* $\varphi \vee \psi \in \mathcal{L}_{T_S}$ *y* $\varphi \wedge \psi \in \mathcal{L}_{T_S}$.
5. *Si* $\varphi \in \mathcal{L}_{T_S}$ *y* $\psi \in \mathcal{L}_{T_S}$ *entonces las expresiones de la forma* $(\varphi \to \psi)$ *son llamadas implicaciones y pertenecen a* $\mathcal{L}_{T_S}$.

Es importante recordar que la implicación no puede ser dada de la forma tradicional pues la negación se define de manera diferente en una lógica multivaluada.

En notación BNF, las fórmulas $\varphi \in \mathcal{L}_{T_S}$ se definen con las siguientes reglas:

$$\varphi ::= \top | p | \neg\varphi | (\varphi \vee \psi) | (\varphi \wedge \psi) | \varphi \rightarrow \psi$$

Donde $p \in \Phi$, $\mathcal{T_S} \subset \Phi$, y $\varphi, \psi \in L_{T_S}$.

La fórmula $\top$ representa la fórmula que siempre es verdadera, $\neg\varphi$ es como en 4.2.6, $(\varphi \vee \psi)$ representa la disyunción de φ y ψ, $(\varphi \wedge \psi)$ representa la conjunción de φ y ψ, y las expresiones del tipo $(\varphi \rightarrow \psi)$ son llamadas implicaciones.

La situación actual del juego tiene una fuerte relación con nuestras creencias. Por un lado, a partir de las conjeturas que fueron determinando el juego, hasta llegar al estado actual. Por el otro, esta situación será a partir de la cual se construyan todas las conjeturas posteriores cuya finalidad es resolver el sudoku.

De esta manera, lo que necesitamos es definir operadores modales con los que podamos construir fórmulas que nos permitan describir el conocimiento que tenemos sobre el juego.

Definición 4.3.3 *Definimos el lenguaje epistémico para el sudoku como sigue. Las fórmulas del lenguaje lógico epistémico proposicional para el sudoku $\mathcal{LE}_{T_S}$ se construyen a partir del conjunto de fórmulas atómicas Φ siguiendo las siguientes reglas.*

1. $\top \in \mathcal{LE}_{T_S}$
2. *Si* $p \in \Phi$ *(donde* $\mathcal{T}_S \subset \Phi$*) entonces* $p \in \mathcal{LE}_{T_S}$
3. *Si* $\varphi \in \mathcal{LE}_{T_S}$ *y* $\psi \in \mathcal{LE}_{T_S}$ *entonces* $\varphi \vee \psi \in \mathcal{LE}_{T_S}$
4. *Si* $\varphi \in \mathcal{LE}_{T_S}$ *y* $\psi \in \mathcal{LE}_{T_S}$ *entonces* $\varphi \wedge \psi \in \mathcal{LE}_{T_S}$
5. *Si* $\varphi \in \mathcal{LE}_{T_S}$ *entonces* $\neg\varphi \in \mathcal{LE}_{T_S}$ *(con la negación definida como en 4.2.6).*
6. *Si* $\varphi \in \mathcal{LE}_{T_S}$ *y* $\psi \in \mathcal{LE}_{T_S}$ *entonces* $(\varphi \rightarrow \psi) \in \mathcal{LE}_{T_S}$
7. *Si* $\varphi \in \mathcal{LE}_{T_S}$ *entonces* $K\varphi \in \mathcal{LE}_{T_S}$

8. Si $\varphi \in \mathcal{LE}_{T_S}$ *entonces* $P\varphi \in \mathcal{LE}_{T_S}$

En notación BNF, las fórmulas $\varphi \in \mathcal{LE}_{T_S}$ *se definen con las siguientes reglas:*

$$\varphi ::= \top | p | \neg\varphi | (\varphi \vee \psi) | (\varphi \wedge \psi) | (\varphi \rightarrow \psi) | K\varphi | P\varphi$$

Donde $p \in \Phi$*,* $\mathcal{T}_S \subset \Phi$ *y* $\varphi, \psi \in \mathcal{LE}_{T_S}$*.*

La fórmula $\top$ *representa la fórmula que siempre es verdadera,* $\neg\varphi$ *representa la negación de* φ *(definida como en 4.2.6),* $(\varphi \vee \psi)$ *representa la disyunción de* φ *y* ψ*,* $(\varphi \wedge \psi)$ *representa la conjunción de* φ *y* ψ*, a* $(\varphi \rightarrow \psi)$ *le denominamos implicación,* $K\varphi$ *expresa que el jugador sabe* φ *y* $P\varphi$ *expresa que el jugador considera posible* φ *(donde "considerar posible"quiere decir que el jugador sabe que* φ *no violaría las reglas del juego). Notemos que el operador* P *es el dual de* K*, esto es por que el agente considera posible que un número ocupe una casilla si y sólo si él no sabe con certeza que esa casilla no puede ser ocupada por ese número. Es decir:*

$$P\varphi \equiv \neg K \neg\varphi$$

En el capítulo 2 mencionamos algo sobre lógica multimodal epistémica. La definición clásica de lógica modal epistémica es multimodal: con una modalidad por cada agente miembro del conjunto sobre el que se construye el modelo. Nuestro modelo para el conocimiento del sudoku será multimodal, aunque ya que en el sudoku sólo hay un agente, la multimodalidad será en un sentido diferente al clásico.

Las multimodalidades son un recurso que ha sido muy utilizado en epistemología formal. En 1970, cuando la lógica epistémica era aún bastante incipiente, Scott señaló que uno de los grandes errores que se cometían en lógica modal era concentrarse en sistemas con sólo un operador (Scott, 1970). La razón de este tipo de afirmaciones es que las multimodalidades nos permiten capturar de manera más precisa cuestiones más complejas. Por ejemplo, en el caso del modelo que estamos construyendo surge la necesidad de capturar el hecho de que las creencias (el conocimiento) que obtenemos se desarrollan en función de las jugadas. Esto en algunos juegos se resuelve utilizando lógicas dinámicas con operadores de acción. Sin embargo, el sudoku no necesita tanto pues como ya hemos mencionado, mientras que en otros juegos como el dominó, el estado del mundo va cambiando según las jugadas, en el sudoku todo permanece estático; lo único que se altera es la configuración de los números que se exhiben.

Definición 4.3.4 *Definimos el lenguaje epistémico con compatibilidad para el sudoku como sigue. Las fórmulas del lenguaje lógico epistémico proposicional con compatibilidad para el sudoku* $\mathcal{LEC}_{T_S}$ *se construyen a partir del conjunto de fórmulas atómicas* Φ *siguiendo las siguientes reglas.*

1. $\top \in \mathcal{LEC}_{T_S}$
2. *Si* $p \in \Phi$ *(donde* $\mathcal{T}_S \subset \Phi$*) entonces* $p \in \mathcal{LEC}_{T_S}$
3. *Si* $\varphi \in \mathcal{LEC}_{T_S}$ *y* $\psi \in \mathcal{LEC}_{T_S}$ *entonces* $\varphi \vee \psi \in \mathcal{LE}_{T_S}$
4. *Si* $\varphi \in \mathcal{LEC}_{T_S}$ *y* $\psi \in \mathcal{LEC}_{T_S}$ *entonces* $\varphi \wedge \psi \in \mathcal{LEC}_{T_S}$
5. *Si* $\varphi \in \mathcal{LEC}_{T_S}$ *entonces* $\neg\varphi \in \mathcal{LEC}_{T_S}$ *(con la negación definida como en 4.2.6).*
6. *Si* $\varphi \in \mathcal{LEC}_{T_S}$ *y* $\psi \in \mathcal{LEC}_{T_S}$ *entonces* $(\varphi \rightarrow \psi) \in \mathcal{LEC}_{T_S}$
7. *Si* $\varphi \in \mathcal{LEC}_{T_S}$ *entonces* $K\varphi \in \mathcal{LEC}_{T_S}$
8. *Si* $\varphi \in \mathcal{LEC}_{T_S}$ *entonces* $P\varphi \in \mathcal{LEC}_{T_S}$
9. *Si* $\varphi \in \mathcal{LEC}_{T_S}$ *entonces* $\Box_c\varphi \in \mathcal{LEC}_{T_S}$
10. *Si* $\varphi \in \mathcal{LEC}_{T_S}$ *entonces* $\Diamond_c\varphi \in \mathcal{LEC}_{T_S}$

La fórmula $\top$ *representa la fórmula que siempre es verdadera,* $\neg\varphi$ *representa la negación de* φ *(definida como en 4.2.6),* $(\varphi \vee \psi)$ *representa la disyunción de* φ *y* ψ*,* $(\varphi \wedge \psi)$ *representa la conjunción de* φ *y* ψ*, a* $(\varphi \rightarrow \psi)$ *le denominamos implicación,* $K\varphi$ *expresa que el jugador sabe* φ *y* $P\varphi$ *expresa que el jugador considera posible* φ*,* $\Box_c\varphi$ *expresa que* φ *es compatible con el resto de las proposiciones (es decir, que al realizar la acción expresada por* φ *no se contradicen las reglas del juego), y* $\Diamond_c\varphi$ *expresa que es posible que* φ *sea compatible con el resto de las proposiciones.*

Una vez que establecimos nuestro lenguaje, abordaremos ahora las cuestiones semánticas del modelo.

4.4. Semántica

En un juego bien planteado la solución es única, esto implica que pensar los mundos posibles como estados o situaciones de correspondencia entre números y casillas puede darnos muchos más mundos de los que necesitamos considerar. Por esta razón, lo más conveniente es pensarlos como estados o situaciones de los números que se encuentran exhibidos en un determinado tiempo t (notemos que este enfoque nos permite obtener conclusiones acerca de nuestro conocimiento del juego). Tampoco hay manera de que lleguemos a estados en los que aparezca dos veces el mismo número en el mismo renglón, columna o subcuadrícula, ya que si bien, podríamos colocar un número en una casilla sin estar completamente seguros de que ese es el sitio que le corresponde, lo que no va a ocurrir (si seguimos las reglas) es que una vez que estemos seguros de que un número corresponde a una casilla (tal vez por que es pista inicial) intencionalmente coloquemos ese numero en la misma fila, renglón o casilla. Así pues, los mundos posibles serán arreglos parciales en los que se exhiben ciertos números que no contradicen las reglas del juego. Nuestro interés es conocer el valor de verdad de las proposiciones que describen la correspondencia entre estos números y sus casillas. Por ejemplo, la solución del juego es un mundo posible (llamémosle $\mathcal{S}$). Éste puede ser descrito como la conjunción de todas las proposiciones verdaderas del tipo $9_{111}, 6_{121}, 8_{131}, 2_{211}, 7_{221}, 4_{231}, 1_{311}, 5_{321}, 3_{331}, \ldots, 2_{999}$, las cuales describen los números que son exhibidos en las casillas (que en este caso corresponden a todo el tablero).

9	6	8	4	7	3	1	2	5
2	7	4	9	1	5	8	6	3
1	5	3	2	8	6	7	9	4
8	9	5	1	3	2	4	7	6
7	4	2	8	6	9	5	3	1
3	1	6	5	4	7	2	8	9
6	2	1	3	5	8	9	4	7
5	3	9	7	2	4	6	1	8
4	8	7	6	9	1	3	5	2

$\mathcal{S}$= Sudoku resuelto

Definición 4.4.1 *(Relación de accesibilidad para compatibilidad). Sea w_i un subarreglo en el tablero del sudoku. Decimos que w_j es accesible (bajo la relación de compatibilidad) a w_i si w_j exhibe un número más que w_i y éste no genera contradicción con las reglas del juego; es decir, si el número que se agrega en w_j no aparecía ya en el mismo renglón, columna o subcuadrícula en w_i. De esta manera tenemos que*

$$\mathcal{R}_1 = \{(w_i, w_j) \mid w_i < w_j \text{ en sólo un número sin contradecir las reglas}\}$$

W será el conjunto de todos los w_i (posibles subarreglos) y $\mathcal{R}_1 \subset W \times W$ la relación de compatibilidad entre mundos posibles.

Si para algún $w \in W$ se tiene que no existe w_i tal que $(w, w_i) \in R_1$ sólo tenemos dos opciones:

a) $w = S$ (donde S es el arreglo solución).

b) Existe $\varphi \in w$ tal que su valor de verdad es distinto a su valor de verdad en S.

En otras palabras, w_j es accesible a w_i si podemos llegar de w_i a w_j en una sola jugada siguiendo las reglas del juego. Notemos que nuestro conocimiento en w_j es compatible (no genera contradicciones) con nuestro conocimiento en w_i. Una vez que no podemos avanzar más, sólo tenemos dos opciones: o hemos terminado exitosamente el juego, o hemos realizado alguna jugada incorrecta, lo que implica que no podemos tener conocimiento del total de la correspondencia número-casilla que se exhibe en w.

Definición 4.4.2 *(Relación de accesibilidad para conocimiento). Sea w_i un subarreglo en el tablero del sudoku (w_i está conformado por un conjunto de números exhibidos en un subconjunto de casillas del tablero). Decimos que w_j es accesible (bajo la relación de conocimiento) a w_i si el conjunto de números exhibidos en w_j es compatible con el de los exhibidos en w_i. Esto lo único que quiere decir es que no ocurre que en w_i haya un número en una cierta casilla, digamos C_{ijk}, y que en w_j aparezca el mismo número en la fila i, en la columna j o en la subcuadrícula k. De esta manera tenemos que*

$$\mathcal{R}_2 = \{(w_i, w_j) | \text{la exhibición de los números de } w_i \text{ no contradice a la de } w_j\}$$

W será el conjunto de todos los w_i (posibles subarreglos) y $\mathcal{R}_2 \subset W \times W$ la relación de accesibilidad de conocimiento entre mundos posibles.

Como se ha venido expresando, nuestro principal interés al desarrollar un modelo del conocimiento en el sudoku es observar qué aspectos epistémicos logran ser capturados por el modelo, cuáles no (ya sea por limitaciones particulares o generales) y qué problemas se pueden resuelven a partir del modelo. Para no desviarnos del objetivo, es importante no perdernos en definiciones, propiedades y demostraciones formales, y detenernos constantemente a reflexionar sobre el significado que cada elemento del aparato formal tiene en términos espistémicos. En este punto de la construcción tenemos que las relaciones de accesibilidad que darán significado a las dos modalidades, cumplen con ciertas propiedades. Por ejemplo, que $\mathcal{R}_2$ sea reflexiva quiere decir que nuestro conocimiento en un momento dado del juego, es compatible con nuestro conocimiento en ese mismo momento. Esto, más allá de ser trivial, es un hecho indiscutible. Cada una de las propiedades que cumplen las relaciones de accesibilidad dicen algo acerca del comportamiento del juego y de nuestras creencias dentro de él.

4.5. Satisfacción y verdad

Dada una proposición, digamos N_{ijk}, y un mundo posible w_i pueden ocurrir tres cosas:

Caso 1 Que sepamos que N_{ijk} es verdadera en w_i; por ejemplo, si N es una pista inicial en la casilla ijk, o si no hay manera de que N no corresponda a la casilla ijk.

Caso 2 Que sepamos que N_{ijk} es falsa en w_i; por ejemplo, si $M \neq N$ y M es una pista inicial en la casilla ijk, o si no hay manera de que M no corresponda a la casilla ijk.

Caso 3 Que no sepamos el valor de verdad de N_{ijk} al pararnos en el mundo w_i; por ejemplo, si la casilla ijk se encuentra en blanco, o si hemos conjeturado que N corresponde a la casilla ijk con algún grado de incertidumbre.

Como en cualquier lógica modal, la verdad de las proposiciones está dada respecto a cada mundo posible. Por las observaciones realizadas en el párrafo anterior, es necesario tener en mente las consideraciones sobre lógica multivaluada que se presentaron en 4.3 antes de dar las nociones de verdad y validez.

Definición 4.5.1 *Sea $\mathcal{F}_S = \langle W, \mathcal{R} \rangle$ el marco de Kripke para el modelo del conocimiento en el sudoku y Φ el conjunto de proposiciones atómicas (con $\Phi = \Phi' \cup \mathcal{T}_S$). Tenemos la función evaluación (trivalor) $e : \Phi \times W \to \mathcal{T}_S$ que asigna valores de la siguiente manera:*

- *Si $\varphi \in \mathcal{T}_S$, $e(\varphi, w) = \varphi$. Esto para toda $w \in W$.*
- *Sea $\varphi \in \Phi'$. Entonces φ es de la forma "El número n corresponde a la casilla C_{ijk}". Tenemos que $e(\varphi, w) = V$ si en el subarreglo w, el número n efectivamente se encuentra en la casilla ubicada en la fila i y el renglón j (cuya intersección se debe ubicar en la subcuadrícula k).*
- *Sea $\varphi \in \Phi'$. Entonces φ es de la forma "El número n corresponde a la casilla C_{ijk}". Tenemos que $e(\varphi, w) = F$ si en el subarreglo w, existe m con $m \neq n$, y m se encuentra en la casilla ubicada en la fila i y el renglón j.*
- *Sea $\varphi \in \Phi'$. Entonces φ es de la forma "El número n corresponde a la casilla C_{ijk}". Tenemos que $e(\varphi, w) = \perp$ si en el subarreglo w, la casilla C_{ijk} se encuentra en blanco.*

Definición 4.5.2 *Extendemos la función e a una función $e' : \mathcal{LEC}_{T_S} \times W \to \mathcal{T}_S$ como sigue.*

1. *Valores de verdad para la conjunción: $e'(\varphi \wedge \psi, w) = V$ en el caso en el que $e(\varphi, w) = V$ y $e(\psi, w) = V$, $e'(\varphi \wedge \psi, w) = F$ si $e(\varphi, w) = F$ o $e(\psi, w) = F$. $e(\varphi \wedge \psi, w) = \perp$ en cualquier otro caso.*
2. *Valores de verdad para la disyunción: $e'(\varphi \vee \psi, w) = V$ en el caso en el que $e(\varphi, w) = V$ o $e(\psi, w) = V$, $e'(\varphi \vee \psi, w) = F$ si $e(\varphi, w) = F$ y $e(\psi, w) = F$, $e'(\varphi \vee \psi, w) = \perp$ en cualquier otro caso.*
3. *Valores de verdad para la negación: $e'(\varphi, w) = V$ sii $e(\neg\varphi, w) = F$, $e'(\varphi, w) = F$ sii $e(\neg\varphi, w) = V$, $e'(\varphi, w) = \perp$ sii $e(\neg\varphi, w) = \perp$.*

Definición 4.5.3 *Definimos la validez para los operadores modales como sigue.*

1. *Decimos que $\Box_c\varphi$ es válida en el marco $\mathcal{F}_S$, bajo la evaluación e en el mundo $w \in W$. Es decir, que $\mathcal{F}_S, e, w \models \Box_c\varphi$ si y sólo si $e(\varphi, w') = V$ para todo w' tal que $(w, w') \in \mathcal{R}_1$*

2. *Decimos que $\Diamond_c\varphi$ es válida en el marco $\mathcal{F}_S$, bajo la evaluación e en el mundo $w \in W$. Es decir, que $\mathcal{F}_S, e, w \models \Diamond_c\varphi$ si y sólo si $\exists w'$ con $(w, w') \in \mathcal{R}_1$ tal que $e(\varphi, w') = V$.*

3. *Decimos que $K\varphi$ es válida en el marco $\mathcal{F}_S$, bajo la evaluación e en el mundo $w \in W$. Es decir, que $\mathcal{F}_S, e, w \models K\varphi$ si y sólo si $e(\varphi, w') = V$ para todo w' tal que $(w, w') \in \mathcal{R}_2$*

4. *Decimos que $P\varphi$ es válida en el marco $\mathcal{F}_S$, bajo la evaluación e en el mundo $w \in W$. Es decir, que $\mathcal{F}_S, e, w \models K\varphi$ si y sólo si si y sólo si $\exists w'$ con $(w, w') \in \mathcal{R}_2$ tal que $e(\varphi, w') = V$.*

Las definiciones anteriores nos da una idea bastante buena de cómo se comporta el juego.

4.6. Propiedades de las relaciones de accesibilidad

Recordemos que lo que la relación de compatibilidad $\mathcal{R}_1$ nos dice, es que un subarreglo w_1 del sudoku es accesible a otro w si del estado del juego w podemos llegar al estado del juego w_1 en un sólo paso sin romper las reglas del juego.

Tenemos que $\mathcal{R}_1$ no es reflexiva, pues no hay manera de que al agregar un número a una casilla el juego permanezca igual. Tampoco es simétrica ni transitiva. No es serial, pues existe el arreglo S que describe la solución y que no puede acceder a ningún mundo bajo $\mathcal{R}_1$. De hecho, la relación $\mathcal{R}_1$ es muy limitada y no satisface ninguna de las propiedades que se mencionan en la sección 2.6. Esto quiere decir que tampoco vamos a poder dar como válidos los axiomas del sistema S_5 para los operadores de compatibilidad.

Por otro lado, dado que trabajamos con dos modalidades, aún nos falta revisar la relación de accesibilidad de conocimiento $\mathcal{R}_2$. Ésta la revisaremos con más detalle.

A partir de la definición (4.4.2) podemos observar que $\mathcal{R}_2$ es una relación:

1. **Reflexiva**; $\forall w.(w, w) \in \mathcal{R}_2$. Los números que se exhiben en w no contradicen el conocimiento que poseemos en w.

2. **Serial**; $\forall w.\exists w_1.(w, w_1) \in \mathcal{R}_2$. Si una relación es reflexiva, entonces es serial. Pues siempre existe un mundo (el mismo w) que es accesible desde w.

3. **Simétrica**; $\forall w, w_1.(w, w_1) \in \mathcal{R}_2$ implica $(w_1, w) \in \mathcal{R}_2$. Si los números exhibidos en w_1 no generan contradicción con los números exhibidos en w, entonces los números exhibidos en w no generan contradicción con los exhibidos en w_1.

4. **Transitiva**; $\forall w, w_1, w_2$. Si $(w, w_1) \in \mathcal{R}_2 \wedge (w_1, w_2) \in \mathcal{R}_2$, entonces $(w, w_2) \in \mathcal{R}_2$. Si los números de un subarreglo no generan contradicción con los de otro, y éste a su vez, no genera contradicción con un tercero, entonces, el primer subarreglo tampoco genera contradicción con el tercero.

Es muy importante notar que nuestra relación de conocimiento es de equivalencia. Esto va a garantizar que sean válidos nuestros axiomas. Una vez que tengamos los axiomas, podemos afirmar que el modelo está bien construido y observar las virtudes que este nos brinda para el estudio del fenómeno (el conocimiento en el sudoku).

4.7. Axiomas del modelo

Los axiomas pueden tener diferentes funciones dentro del modelo. Algunos por ejemplo, establecen las “reglas”que determinan cómo proceder, otros son consecuencia natural de cómo construimos el lenguaje y la semántica. La validez de los axiomas muestra que hay al menos un sentido en el que nuestro modelo está “bien construido”. En el caso del modelo para el conocimiento del sudoku proponemos tres grupos de axiomas. Cada uno de estos grupos tiene un papel distinto en el modelo, pero como observaremos, todos están relacionados entre sí.

4.7.1. Axiomas del juego

Estos axiomas son los que expresan las reglas. Sabemos que en el sudoku está prohibido colocar dos veces el mismo número en la misma fila, en la misma columna o en la misma subcuadrícula. Esta idea debe ser capturada por los axiomas del modelo.

(J1) $N_{ijk} \Rightarrow \neg M_{ijk}$, si $N \neq M$

(J2) $N_{ijk} \Rightarrow \neg N_{lmn}$, si $(i = l) \vee (j = m) \vee (k = n)$

El axioma J1 dice que si un número corresponde a una casilla, entonces ningún otro puede corresponder a esa casilla (recordemos que la solución es única). El axioma J2 dice que si un número corresponde a una casilla, entonces éste no puede ocupar otra casilla en el mismo renglón, o en la misma columna, o en la misma subcuadrícula.

Al principio del capítulo mencionamos que cada casilla está determinada por la intersección de un renglón y una columna. Aquí podemos ver la importancia de indicar también mediante un subíndice la subcuadrícula; un número podría no repetirse ni en el mismo renglón ni en la misma columna y a pesar de ello repetirse en la misma subcuadrícula, lo cual también va en contra de las reglas del juego.

4.7.2. Axiomas de compatibilidad

Una observación importante es que la posibilidad epistémica está condicionada a la posibilidad de acción; en el sudoku sólo puedo *conocer* aquello que puedo *realizar* (jugar). Esta característica está siendo perfectamente capturada por nuestros axiomas.

Consideremos la proposición $\varphi =$ *El número n corresponde a la casilla* C_{ijk}. Es claro que si n no es una pista inicial, no hay manera de que yo tenga el conocimiento de que n corresponde a la casilla C_{ijk} si no es por que la acción de colocar n en C_{ijk} está permitida por las reglas del juego. O sea, una condición necesaria para que φ sea conocimiento en w es poder realizar la jugada que describe φ. Entonces, φ debe ser compatible con el estado en w, es decir que $\mathcal{F}_S, e', w \models \Box_c\varphi$ (para todo $w \in W$). Esto lo podemos expresar con el siguiente axioma.

(C1) $K\varphi \Rightarrow \Box_c\varphi$

Dos equivalencias de este axioma que parecen ir muy bien con nuestra intuición son las siguientes.

- $\neg\Box_c\varphi \Rightarrow P\neg\varphi$
- $\neg\Box_c\varphi \Rightarrow \neg K\varphi$

La primera equivalencia se obtiene aplicando el dual del operador K y usando la contrapositiva de la implicación, y lo que nos dice es que si no podemos realizar la jugada expresada por φ por que no es compatible, entonces consideramos posible que φ no sea el caso.

La segunda es la contrapositiva del axioma original, y lo que nos dice es que si no puedo realizar una jugada (φ), no puedo saber el contenido de la proposición φ. Por ejemplo, supongamos que hay un 5 (pista inicial) en la casilla C_{ilm}, entonces no puedo poner un 5 en la casilla C_{ijk} (axioma 2 del juego), esto implica que la proposición "*El 5 corresponde a la casilla ijk.*es falsa y por lo tanto, no puede ser conocimiento.

Bajo el mismo razonamiento, tenemos el siguiente axioma.

(C2) $P\varphi \Rightarrow \Diamond_c\varphi$

Equivalentemente (aplicando el dual y contrapositiva):

- $\neg\Box_c\varphi \Rightarrow K\neg\varphi$

Este axioma dice que una condición necesaria para considerar posible φ es que de hecho, sea posible realizar la jugada expresada por φ. En otras palabras, si φ es incompatible con el juego, entonces sabemos que no puede ser el caso que φ.

Ahora sería natural preguntarse si la implicación contraria de (C2) se cumple. Es decir, si tenemos que $\Box_c\varphi \Rightarrow K\varphi$. Y la respuesta es que no. Sería contraintuitivo (incluso falso) pensar que para que una jugada pueda realizarse es necesario que yo sepa que la proposición que la expresa es verdadera. ¿Qué ocurriría entonces en un juego bien determinado pero cuya dificultad no nos permite seguir "averiguando"las proposiciones verdaderas que nos conducen a la solución? Esto sólo nos dice que las posibilidades de acción en el juego no están condicionadas necesariamente a las posibilidades de conocimiento.

4.7.3. Axiomas de conocimiento

Para ver que los axiomas de conocimiento son válidos para nuestro modelo necesitamos algunas nociones generales. Primero, recordemos las reglas de inferencia de una lógica modal (utilizando los operadores de conocimiento para evitar confusiones).

(MP) $\frac{\varphi \to \psi, \varphi}{\psi}$

($\mathcal{K}$) $K(\varphi \to \psi) \to (K\varphi \to K\psi)$ (Axioma de distribución)

(N) $\frac{\varphi}{K\varphi}$

(EN) $\frac{\varphi \to \psi}{K\varphi \to K\psi}, \frac{\varphi \to \psi}{P\varphi \to P\psi}$

Todos los sistemas axiomáticos para lógica modal incluyen el axioma $\mathcal{K}$ y las reglas MP y N. La regla EN puede derivarse a partir de $\mathcal{K}$, MP y N. Además de estos axiomas, existen varios esquemas modales. Algunos de los más utilizados son:

(D(φ)) $K\varphi \to P\varphi$

(T(φ)) $K\varphi \to \varphi$ (Axioma de conocimiento)

(B(φ)) $\varphi \to KP\varphi$

(4(φ)) $K\varphi \to KK\varphi$ (Axioma de introspección positiva)

(5(φ)) $P\varphi \to KP\varphi$ (Axioma de introspección negativa)

Definición 4.7.4 *Llamamos S_5 a un sistema axiomatico cuando cumple con los esquemas $\mathcal{K}T5$.*

Tradicionalmente la lógica modal epistémica cumple con los axiomas de S_5. En principio, estos axiomas han funcionado bastante bien para capturar nuestras intuiciones acerca de lo que creemos que el conocimiento debería cumplir. Por esta razón es que el siguiente paso natural en la presentación de nuestro modelo es ver que se satisfacen los esquemas de S_5. Lo que haremos será demostrar que si tenemos $\mathcal{K}$DB4 entonces tenemos S_5. Después veremos que nuestro modelo del conocimiento del sudoku ($\mathcal{M}_S$) satisface $\mathcal{K}$DB4. Con esto quedará demostrado que el sistema S_5 es válido para nuestro modelo.

Sean S_α y S_β sistemas axiomaticos. Utilizaremos la notación $S_\alpha \leq S_\beta$ para expresar que todo todo teorema de S_α es teorema de S_β.

Lema 4.7.5 $\mathcal{K}DB5 \leq \mathcal{K}DB4$

Demostración.

1. $P\neg\varphi \to KP\neg\varphi$ $\quad 5(\neg\varphi)$
2. $\neg K\varphi \to \neg P\neg P\neg\varphi$ Interdefinibilidad de los operadores
3. $P\neg P\neg\varphi \to K\varphi$ Contrapuesta de la implicación
4. $PK\varphi \to K\varphi$ Interdefinibilidad en el antecedente
5. $KPK\varphi \to KK\varphi$ EN
6. $K\varphi \to KPK\varphi$ $\quad B(K\varphi)$

$\therefore$ $K\varphi \to KK\varphi$ Transitividad de la implicación a 6 y 5.

Hemos probado que el esquema 4 es teorema de $\mathcal{K}DB5$.

■

Lema 4.7.6 $\mathcal{K}DB4 \leq \mathcal{K}DB5$

Demostración.

1. $K\neg\varphi \to KK\neg\varphi$ $\quad 4(\neg\varphi)$
2. $\neg P\varphi \to \neg P\neg K\neg\varphi$ Interdefinibilidad de los operadores
3. $PP\varphi \to P\varphi$ Contrapuesta de la implicación
4. $KPP\varphi \to KP\varphi$ EN
5. $P\varphi \to KPP\varphi$ $\quad B(P\varphi)$

$\therefore$ $P\varphi \to KP\varphi$ Transitividad de la implicación a 5 Y 4.

Esto prueba que el esquema 5 es teorema de $\mathcal{K}DB5$.

■

Teorema 4.7.7 $\mathcal{K}DB4 \equiv \mathcal{K}DB5$

Demostración. Por el lema 4.7.5 tenemos que todo teorema de $\mathcal{K}DB5$ es teorema de $\mathcal{K}DB4$. Es decir que el conjunto de teoremas de $\mathcal{K}DB5$ está contenido en el conjunto de teoremas de $\mathcal{K}DB4$. Análogamente, por el lema 4.7.6 tenemos que el conjunto de teoremas de $\mathcal{K}DB4$ está contenido en el conjunto de teoremas de $\mathcal{K}DB5$. Por lo tanto, ambos conjuntos son iguales y $\mathcal{K}DB4 \equiv \mathcal{K}DB5$. ■

Lema 4.7.8 *T es teorema de $\mathcal{K}DB4$.*

Demostración.

1. $K\varphi \to KK\varphi$ $4(\varphi)$
2. $KK\varphi \to PK\varphi$ $D(K\varphi)$
3. $K\varphi \to PK\varphi$ Transitividad de la implicación de 1 y 2.
4. $\neg\varphi \to KP\neg\varphi$ $B(\neg\varphi)$
5. $\neg\varphi \to \neg PK\varphi$ Interdefinibilidad de los operadores.
6. $PK\varphi \to \varphi$ Contrapuesta de 5.

$\therefore\ K\varphi \to \varphi$ Transitividad de la implicación a 3 Y 6.

■

Teorema 4.7.9 $\mathcal{K}DB4 \equiv S_5$

Demostración. Como $\mathcal{K}DB4$ es equivalente a $\mathcal{K}DB5$ y T es teorema de $\mathcal{K}DB4$, entonces $\mathcal{K}DB4$ implica que tenemos el axioma 5 y el axioma T. Es decir, tenemos

$$KT5 = S_5$$

■

Por el teorema anterior, basta con ver que el modelo que construimos para el conocimiento del sudoku $\mathcal{M}_S = \langle W, \mathcal{R}_2, e\rangle$ satisface los axiomas de $\mathcal{K}DB4$. Pero antes de verificar los axiomas, vale la pena detenernos a ver cómo es que estos axiomas parecen capturar bien algunas de nuestras intuiciones acerca del juego.

(D) Lo que el axioma D está expresando es que si yo sé que un número n corresponde a una casilla C_{ijk}, entonces considero posible que el número n corresponda a la casilla C_{ijk}.

(B) Si es el caso que n corresponde a la casilla C_{ijk}, entonces debo saber que puedo considerar la posibilidad de que el número n corresponda a la casilla C_{ijk}.

(4) Si sé que un número n corresponde a una casilla C_{ijk}, entonces sé que lo sé. Esto es lo que en el sudoku nos permite distinguir entre una pista inicial o una casilla ocupada por un número cuya correspondencia no podría ser de otra manera, y un número que coloqué (quizá por estrategia) y que es susceptible de ser borrado.

Proposición 4.7.10 *El axioma D es válido en* $\mathcal{M}_S = \langle W, \mathcal{R}_2, e\rangle$.

Demostración. Supongamos que $\models_{\mathcal{M}_S} K\varphi$, esto quiere decir que $\mathcal{F}_S, e, w \models K\varphi$ para todo marco $\mathcal{F}_S$ y todo mundo $w \in W$, entonces $e(\varphi, w') = V$ para todo w' tal que $(w, w') \in \mathcal{R}_2$, como $\mathcal{R}_2$ es serial (4.6), existe al menos un w_1 tal que $(w, w_1) \in \mathcal{R}_2$, por lo anterior, $e(\varphi, w_1) = V$ y entonces $\mathcal{F}_S, e, w \models P\varphi$ para todo marco $\mathcal{F}_S$ y todo mundo $w \in W$. Por lo tanto $\models_{\mathcal{M}_S} P\varphi$. Entonces $K\varphi \to P\varphi$. ■

Proposición 4.7.11 *El axioma B es válido en* $\mathcal{M}_S = \langle W, \mathcal{R}_2, e\rangle$.

Demostración. Supongamos que $\models_{\mathcal{M}_S} \varphi$. Sea $w \in W$ y w' tal que $(w, w') \in \mathcal{R}_2$ (w' existe por que $\mathcal{R}_2$ es serial). Como $\models_{\mathcal{M}_S} \varphi$, en particular $e(\varphi, w') = V$ entonces $\mathcal{F}_S, e, w \models P\varphi$. Como $\mathcal{R}_2$ es simétrica tenemos que $(w', w)\mathcal{R}_2$ (para toda $w \in W$). Entonces $e(P\varphi, x) = V$ para toda $x \in W$ (en particular para todos los mundos accesibles desde cualquier mundo w). Por lo tanto, $\mathcal{F}_S, e, w \models KP\varphi$ para toda $w \in W$ y todo marco de Kripke, es decir,$\models_{\mathcal{M}_S} KP\varphi$. Entonces $\varphi \to KP\varphi$. ■

Proposición 4.7.12 *El axioma 4 es válido en* $\mathcal{M}_S = \langle W, \mathcal{R}_2, e\rangle$.

Demostración. Supongamos que $\models_{\mathcal{M}_S} K\varphi$. Dado que $K\varphi$ es cierto en w, entonces φ es cierto en todo mundo accesible desde w. Es decir, $e(\varphi, u) = V$ para todo $u \in W$ tal que $(w, u) \in \mathcal{R}_2$. Pero como $\mathcal{R}_2$ es transitiva (4.6), entonces $(w, v) \in \mathcal{R}_2$ para todo $v \in W$ tal que $(u, v) \in \mathcal{R}_2$. Como $(w, v) \in \mathcal{R}_2$, tenemos también que $e(\varphi, v) = V$ para todo $v \in W$ tal que $(u, v) \in \mathcal{R}_2$. Ya que φ es verdadera en cualquier mundo accesible desde u, tenemos que $e(K\varphi, u) = V$ para cualquier u accesible desde w. Por lo tanto, $\models_{\mathcal{M}_S} KK\varphi$. Entonces $K\varphi \to KK\varphi$. ■

Hemos probado entonces que es sistema $\mathcal{K}DB4$ es válido en $\mathcal{M}_S$, y por el teorema 4.7.9, tenemos que $\mathcal{M}_S$ es S_5.

Ejemplo 4.7.13 *Para concluir y concretar ideas, presentamos el siguiente ejemplo. Consideremos a w el mundo inicial del siguiente sudoku de 4×4:*

		1	
		2	
1	3	4	2
4	2	3	

Sea x_1 el mundo (subarreglo) resultante de agregar 3 en la casilla C_{242}. Tenemos que $(w, x_1) \in R_1$

		1	
		2	3
1	3	4	2
4	2	3	

x_1

Sean x_2 el mundo resultante de agregar 1 en la casilla C_{221}, x_3 el mundo resultante de agregar 4 en la casilla C_{142}, x_4 el mundo resultante de agregar 1 en la casilla C_{444} y x_5 el mundo resultante de agregar 3 en la casilla C_{111}.

		1	
	1	2	3
1	3	4	2
4	2	3	

x_2

		1	4
	1	2	3
1	3	4	2
4	2	3	

x_3

		1	4
	1	2	3
1	3	4	2
4	2	3	1

x_4

3		1	4
	1	2	3
1	3	4	2
4	2	3	1

x_5

Ahora las únicas opciones para la siguiente jugada son colocar 2 o 4 en la casilla C_{121}, o colocar 2 o 4 en la casilla C_{211}. El problema es que ningún *mundo resultante de cualquiera de estas jugadas es R_1-accesible desde x_5. Notemos que*

$$\neg(x_5 \models \Box 4_{211} \wedge 2_{211}) \vee \neg(x_5 \models \Box 4_{121} \wedge 2_{121})$$

Al no haber ningún mundo que sea R_1-compatible con x_5 tenemos que, o x_5 es la solución del juego, o ha habido un error al formular las creencias sobre la correspondencia entre número y casilla. Una característica de la solución s es que para cada casilla C_{ijk} $(i, j, k \in 1, 2, 3, 4)$ del tablero, existe un número $n \in 1, 2, 3, 4$ tal que $e(n_{ijk}, s) = V$. En este caso, para toda $n \in 1, 2, 3, 4$ tenemos que $e(n_{121}, x_5) = \bot$ (al igual que para C_{211}). Por lo tanto, x_5 no es la solución del juego y alguna de las conjeturas que nos llevó a este estado es incorrecta. Nuestro modelo nos permite derivar la siguiente conclusión:

$$\neg K(3_{242} \wedge 1_{221} \wedge 4_{142} \wedge 1_{444} \wedge 3_{111})$$

Es decir,

$$\neg K 3_{242} \vee \neg K 1_{221} \vee \neg K 4_{142} \vee \neg K 1_{444} \vee \neg K 3_{111}$$

La idea es aplicar (como parte de la segunda etapa del proyecto de modelado del fenómeno epistémico), por ejemplo, un proceso de revisión de creencias que determine cuál es, o cuáles son las creencias incorrectas. Esta acción queda pendiente para un trabajo posterior.

En este caso particular la creencia incorrecta es "3 corresponde a la casilla C_{242}". A nosotros nos gustaría que el modelo no nos permitiera saber este tipo de cosas, entonces queremos que $w \not\models K3_{242}$. Es decir, debe existir un mundo y_1 accesible a w tal quee$(3_{242}, y_1)$ no es verdadera. Vamos a exhibir uno para ver que es el caso.

Consideremos a w como antes (el sudoku original) y a z, y como a continuación.

		1	
		2	
1	3	4	2
4	2	3	

w

		1	
3		2	
1	3	4	2
4	2	3	

z

		1	
	4	2	
1	3	4	2
4	2	3	

y

Notemos que $(w, z) \in R_2$ y $(w, y) \in R_2$ Consideremos ahora w', y' y z' a los mundos resultantes de agregar el número 3 en la casilla C_{242} en w, z y y respectivamente. Tenemos entonces

		1	
		2	3
1	3	4	2
4	2	3	

w'

		1	
3		2	
1	3	4	2
4	2	3	

z'

		1	
	4	2	3
1	3	4	2
4	2	3	

y'

Notemos que $e(3_{142}, w') = V$*,* $e(3_{142}, y') = V$ *pero* $e(3_{142}, z') = \perp$ *pues colocar el* 3 *en la casilla* C_{142} *resultaría en un arreglo prohibido ya que contradice los axiomas del juego. Entonces tenemos que* $(w, z') \in R_2$ *y* 3_{142} *no es verdadera en* z'*. Por lo tanto,* $w \nvDash K3_{242}$*.*

Ejemplo 4.7.14 *Consideremos nuevamente el juego dado* w*.*

		1	
		2	
1	3	4	2
4	2	3	

De manera informal sabemos que el número 1 *corresponde a la casilla* C_{444}*. Queremos ver que* $w' \models K1_{444}$ *donde* w' *es el resultado de considerar a* w *con el* 1 *en la casilla* C_{444}*.*

		1	
		2	
1	3	4	2
4	2	3	1

w'

Vamos a revisar todos los mundos posibles $z_i \in W$ *tales que* $(w, z_i) \in R_2$*. Según las pretensiones de nuestro modelo, debe ocurrir que* $e(1_{444}, z_i) = V$ *para toda* z_i*.*

En este juego los mundos posibles sólo pueden ser resultado de agregar a alguna casilla los números $1, 2, 3$ *o* 4*. Así que podemos proceder por casos. Observemos los mundos resultantes de agregar un número* 1*.*

Los mundos resultantes de agregar 1 *en las casillas* $C_{ij2} \vee C_{1jk} \vee C_{i1k}$*no están dentro de los arreglos posibles pues generarían contradicción con las reglas. Entonces, sin considerar la casilla* C_{444} *(que es en la que queremos evaluar al número* 1*), sólo hay un mundo accesible desde* w*. Éste es el resultado de colocar al número* 1 *en la casilla* C_{221}*. A este mundo lo llamaremos* z_1*.*

		1	
	1	2	
1	3	4	2
4	2	3	

z_1

		1	
	1	2	
1	3	4	2
4	2	3	1

z_1'

El mundo w es R_1-accesible a z_1, es decir, $(w, z_1) \in R_1$. Si consideramos z_1' el mundo z_1 ya con el número 1 en la casilla C_{444}. Podemos observar que $e(1_{444}, z_1') = V$. Vamos a considerar ahora los mundos resultantes de agregar un número 2.

Los mundos resultantes de agregar 2 en las casillas $C_{i4k} \vee C_{12k} \vee C_{2jk}$ no están dentro de los arreglos posibles pues generarían contradicción con las reglas. Entonces, sólo hay un mundo accesible desde w. Éste es el resultado de colocar al número 2 en la casilla C_{111}. A este mundo lo llamaremos z_2.

2		1	
		2	
1	3	4	2
4	2	3	

z_2

2		1	
		2	
1	3	4	2
4	2	3	1

z_2'

El mundo w es R_1-accesible a z_2, es decir, $(w, z_2) \in R_1$. Si consideramos z_2' el mundo z_2 ya con el número 1 en la casilla C_{444}. Podemos observar que $e(1_{444}, z_2') = V$. Consideremos los mundos posibles resultantes de agregar el número 3.

Los subarreglos resultantes de agregar 3 en las casillas de la forma C_{i2k} no son mundos posibles de nuestro modelo ya que contradicen las reglas (violan el axioma J_2). Esto implica que los mundos R_1-accesibles a w que son resultado de agregar un número 3 son z_3, z_4, z_5 y z_6.

3		1	
		2	
1	3	4	2
4	2	3	

z_3

		1	
3		2	
1	3	4	2
4	2	3	

z_4

		1	3
		2	
1	3	4	2
4	2	3	

z_5

		1	
		2	3
1	3	4	2
4	2	3	

z_6

Consideramos z'_3, z'_4, z'_5 y z'_6 los mundos correspondientes agregando 1 *en la casilla C_{444}. Vamos a evaluar nuestra proposición en estos mundos.*

3		1	
		2	
1	3	4	2
4	2	3	1

z'_3

		1	
3		2	
1	3	4	2
4	2	3	1

z'_4

		1	3
		2	
1	3	4	2
4	2	3	1

z'_5

		1	
		2	3
1	3	4	2
4	2	3	1

z'_6

Como podemos ver $e(1_{444}, z'_3) = V$, $e(1_{444}, z'_4) = V$, $e(1_{444}, z'_5) = V$ y $e(1_{444}, z'_6) = V$. Sólo faltan por considerar los mundos posibles que resultan de agregar el número 4.

Podemos ver que los subarreglos resultantes de agregar 4 *en las casillas de la forma C_{i1k} no son mundos posibles de nuestro modelo (por la misma razón que en los casos anteriores). Así pues, los mundos R_1-accesibles (vía agregar el número* 4*) desde w son los siguientes.*

	4	1	
		2	
1	3	4	2
4	2	3	

z_7

		1	
	4	2	
1	3	4	2
4	2	3	

z_8

		1	4
		2	
1	3	4	2
4	2	3	

z_9

		1	
		2	4
1	3	4	2
4	2	3	

z_{10}

Consideramos z'_7, z'_8, z'_9 y z'_{10} los subarreglos correspondientes resultado de agregar 1 *en la casilla C_{444}. Evaluemos la proposición 1_{444} en estos mundos.*

	4	1	
		2	
1	3	4	2
4	2	3	1

z'_7

		1	
	4	2	
1	3	4	2
4	2	3	1

z'_8

		1	4
		2	
1	3	4	2
4	2	3	1

z'_9

		1	
		2	4
1	3	4	2
4	2	3	1

z'_{10}

Tenemos una vez más que $e(1_{444}, z'_7) = V$, $e(1_{444}, z'_8) = V$, $e(1_{444}, z'_9) = V$ y $e(1_{444}, z'_{10}) = V$.

Los mundos R_2- compatibles con w' son los $z'_i, i \in 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$, (o combinaciones compatibles entre estos mondos). Una combinación compatible sería por ejemplo el mundo $z'_{6,9}$, al que podemos pensar como la "combinación"de los mundos z'_6 y z'_9.

		1	*4*
		2	*3*
1	*3*	*4*	*2*
4	*2*	*3*	*1*

$z'_{6,9}$

Es claro que como nos fijamos en todos los mundos R_1-accesibles a w', en particular las "combinaciones"de estos mundos van a ser R_2-compatibles (y además son todos). En particular tenemos que $e(1_{444}, z) = V$ para todo z tal que $(w', z) \in R_2$. Por lo tanto $w' \models K1_{444}$ que era lo que queríamos.

Con esto terminamos el modelo que pretendíamos construir. Por supuesto, quedan pendientes aspectos acerca del conocimiento en el sudoku; como el grado de certeza con el que conjeturamos una posible jugada, o la manera en la cual nos desplazamos en un árbol de desiciones cada vez que colocamos un número en una casilla. Estas cuestiones van más allá de los alcances del presente trabajo y forman parte de las etapas 2 y 3 mencionadas al principio del capítulo.

Capítulo 5

Conclusiones

A lo largo de este trabajo hemos utilizado nociones epistemológicas tanto tradicionales como formales. Uno de nuestros objetivos era señalar puntos de coincidencia entre ambos enfoques. La idea era dar una respuesta (al menos parcial) a la pregunta sobre qué podemos decir de los estudios en epistemología a partir del gran desarrollo de herramientas formales que ha habido durante las últimas décadas. Así pues, hemos visto que podemos valernos de métodos como la revisión de creencias y la semántica de mundos posibles para dar una respuesta al escéptico que cita la posibilidad de error como el argumento más devastador contra la afirmación de conocimiento. La idea es delimitar el conjunto de posibilidades sobre las cuales el agente en cuestión tiene éxito: si el agente puede tener éxito dentro del conjunto de posibilidades relevantes, entonces el agente puede decir que tiene conocimiento, incluso si comete graves errores dentro de las posibilidades irrelevantes.

Por otro lado, herramientas formales como la lógica epistémica nos brindan una poderosa fuente de justificación. Para Bonjour (1976) una teoría epistemológica adecuada debe establecer una conexión entre cómo da cuenta de la verdad y cómo da cuenta de la justificación. En una epistemología basada en lógica formal, las condiciones de verdad de una proposición (creencia) brindan dicha conexión. Un método que, mediante la aplicación de reglas de formación de fórmulas y una semántica, determine cómo es que se llegó a proposiciones verdaderas (o cómo se justificaron las verdades postuladas), es un método fiable de formación de creencias (en el sentido de Goldman). En principio, ésta parece ser una virtud epistémica que podemos atribuirle a casi cualquier tratamiento formal.

Finalmente, cabe mancionar diferencias entre un tratamiento epistemológico de tipo *tradicional* y uno de tipo *formal*. Para Hendricks (2006), una manera de describir superficialmente cómo se realizan los estudios contemporáneos en epistemología sería: i) de forma convencional o informal: en gran parte mediante análisis conceptuales y concentrándose en ejemplos *pintorescos* y, a veces, extravagantes o en contraejemplos especulativos, o ii) en una manera *formal*; mediante la aplicación de herramientas y métodos de lógica, teoría de la decisión o de la probabilidad, a la teoría del conocimiento.

El modelo del conocimiento en el sudoku parece respaldar lo mencionado en los párrafos anteriores. Por ejemplo, dado un juego de sudoku, si yo tengo la creencia de que el número 5 corresponde a la casilla ubicada en la intersección de la fila 2 y la columna 8, el modelo me proporciona herramientas que me permiten, en un momento específico del juego determinar no sólo si la proposición “$5_{283}.^{\text{es}}$ verdadera, (lo que justificaría mi creencia de que el número 5 corresponde a la casilla ubicada en la intersección de la fila 2 y la columna 8), sino que también obtengo condiciones para poder afirmar que yo *tengo conocimiento* de que el número 5 corresponde a la casilla ubicada en la intersección de la fila 2 y la columna 8. Esto, por que el modelo proporciona condiciones de verdad para las proposiciones del tipo $K5_{283} =$“Yo sé que el número 5 corresponde a la casilla 283”. No cualquier proposición del tipo n_{ijk} puede ser evaluada en todo momento del juego, pero los axiomas brindan criterios que me *guían* hacia las condiciones que permitirán evaluar a cada proposición. Es decir, los axiomas (en particular el de compatibilidad), exhiben cómo en el caso del sudoku la posibilidad de conocimiento está constreñida a la posibilidad de acción. Otra forma de decirlo es que no puedo *conocer* la correspondencia entre una casilla y un número si no es posible realizar la jugada (regulada por los axiomas del juego) que llevaría a dicha correspondencia.

El presente trabajo deja varias preguntas abiertas a futuras investigaciones. Claramente, una tarea pendiente es seguir trabajando en el desarrollo de herramientas formales. Por ejemplo, al combinar diferentes trabajos en lógica, aquí hemos presentado una lógica multimodal epistémica trivaluada. Las necesidades que presentan los problemas epistemológicos dan pie para crear nuevos procesos, así como para afinar los existentes. Otra vertiente sería un trabajo complementario que modelara el conocimiento en el sudoku utilizando procesos de revisión de creencias. La idea sería modelar el conocimiento dentro del juego ante la aparición de nueva información (conforme vamos determinando la correspondencia entre números y casillas). Finalmente, podríamos preguntarnos sobre las similitudes entre el sudoku y otros contextos del conocimiento que nos permitan generalizar los resultados aquí presentados.

Bibliografía

[1] Alchourrón, Carlos & Gärdenfors, Peter & Makinson, David. (1985). "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions". Journal of Symbolic Logic, 50: 510-530.

[2] Ayer, Alfred. (comp.) (1959). *Logical positivism*, Free Pres, New York.

[3] Bonjour, L. (1976). "The Coherence Theory of Empirical Knowledge", Philosophical Studies, 30: 281-312.

[4] Carnap, Rudolf. (1946). "Modalities and quantification", The Journal of Symbolic Logic, 11(2): 33-64.

[5] Dancy, J. (1993). "Conocimiento", en *Introducción a la epistemología contemporánea*, Tecnos, Madrid, 39-67.

[6] Dalal, M. (1988). "Investigations into a Theory of Knowledge Base revision: Preliminary Report", Seventh National Conference on Artificial Intelligence, 88: 475-479.

[7] Fermé, Eduardo (1999). *Revising the AGM Postulates.* Tesis Doctoral. Universidad de Buenos Aires.

[8] Fitting, Melvin C. (1992). "Many-Valued Modal Logics", Fundamenta Informaticae.

[9] Foley, Richard (1992). "The Epistemology of Belief and the Epistemology of Degrees of Belief". American Philosophical Quaterly 29, 111-124.

[10] Gärdenfors, Peter (1978). "Conditionals and changes of belief". Acta Philosophica Fennica, 30: 381-404.

[11] Gettier, Edmund L. (1963). "Is Justified True Belief Knowledge?". Analysis 23, 121-123.

[12] Goldman, Alvin (1979). "What is Justified True Belief?Ïn G.S. Pappas (ed.), Justification and Knowledge. Dordrecht: D. Reidel: 1-23.

[13] Goldman, Alvin (1986). *Epistemology and Cognition.* Harvard University Press.

[14] Grove, A. (1988). "Two Modellings for Theory Change", The Journal of Philosophical Logic, 17: 157-170.

[15] Hanson, Norwood R (1958). "The logic of discovery", Journal of Philosophy, Vol. LV, N° 25, p.1081.

[16] Harper, William (1976). "Ramsey test conditionals and iterated belief change". En Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science, Vol I, Harper, W.L. y Hooker, C.A. eds. , Reidel, Dordrecht.

[17] Hawthrone, James & Bovens, Luc (1999). *The Preface, The lottery, and the Logic of Belief.* Mind 108, 241-264.

[18] Hempel, Carl G.; Oppenheim, Paul (Apr. 1948). *Studies in the Logic of Explanation*, Philosophy of Science, Vol. 15, No. 2, pp. 135-175.

[19] Hendricks, Vincent F. (2001). *The Convergence of the Scientific Knowledge: A View from the Limit.* Trends in Logic: Studia Logica Library Series.

[20] Hendricks, Vincent F. (2006). *Mainstream and formal epistemology*, New York: Cambridge University Press.

[21] Hintikka, Jakko (1957). "Modality as referential multiplicity", Ajatus, 20: 49-64.

[22] Hintikka, Jakko (1962). *Knowledge and Belief.* Ithaca: Cornell University Press.

[23] Hintikka, Jakko (2003). "A Second Generation Epistemic Logicand Its General Significance". En Hendricks et al. 2003.

[24] Hubert, Franz (2008). "Assessing Theories, Bayes Style". Synthese 161, 89-118.

[25] Kuhn, Thomas S. (1962). *The Structure of Scientific Revolutions*, 1st. ed., Chicago: Univ. of Chicago Pr.

[26] Kripke, Saul (1963). "Semantical analysis of modal logic", Zeitschrift für Mathematische Logik und Grundlagen der Mathematik, 9: 67-96.

[27] Kripke, Saul (1975). "Outline of a theory of truth", The Journal of Philosophy, 72: 690-716.

[28] Kleene, Stephen C. (1950). *Introduction to Metamathematics*, D. Van Nostrand, Princeton, NJ, 1950.

[29] Levi, Isaac (1967). "Gambling with Truth: an essay on induction and the aims of science". New York: Knopf. Reprint, Cambridge, MIT Press, 1973.

[30] Levi, Isaac (1977). "Subjunctives, Dispositions and Chances", Synthese , 34: 423-455.

[31] Levi, Isaac (1980). *The Enterprise of Knowledge.* MIT Press, Cambridge, MA.

[32] Lewis, David (1973). *Counterfactuals*, Harvard University Press; impresión revisada Blackwell 1986.

[33] Nozick, Robert (1981). *Philosophical Explanations.* Cambridge, MA: Harvard University Press.

[34] Ruben David-Hillel (1990). "Arguments Laws and Explanation", del libro *Explaining Explanation*, New York, Routledge.

[35] Salmon, Wesley (1984). *Scientific Explanation and the Causal Structure of the World* , Princeton University Press.

[36] Sankey, H. (1999). "A Realist Approach to Method and Truth". Lectura entregada al Departament of Philosophy and Science Studies, Roskilde, University, Denmark.

[37] Sccot, D. (1970). "Advice on Modal Logic". En K. Lambert (ed.), Philosophical Problems in Logic. Dordrecht: D. Reidel: 143-173.

[38] Segerberg, Krister (1967). "Some Modal Logics Based on a Three-Valued Logic", Theoría.

[39] Stalnaker, R. C. (1968). *A theory of conditionals*. En N. Rescher (Ed.), Studies in logical theory. Oxford: Basil Blackwell.

[40] Thomason, S. K.(1978). "Possible worlds and many tuth values". Studia Logica, 37: 195-204.

[41] van Benthem, Johan (2007). "Dynamic Logic for Belief Revision". Journal of Applied Non-Classical Logics 17, 129-155.

[42] van Ditmarsch, Hans & van der Hoek, Wiebe & Kooi, Berteld (2007). *Dynamic Epistemic Logic*. Dordrecht: Springer.

[43] von Wright, G. H. (1951). "An Essay in Modal Logic". North Holland, Amsterdam.

[44] Zalta, Edward N. (1995). *Basic Concepts in Modal Logic*. Notes: Center for the Study of Language and Information, Stanford University.