



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
**POSGRADO EN FILOSOFÍA DE LA CIENCIA**  
(DGDC) DIRECCIÓN GENERAL DE DIVULGACIÓN DE LA CIENCIA  
(FFYL) FACULTAD DE FILOSOFÍA Y LETRAS  
(FC) FACULTAD DE CIENCIAS  
(IIF) INSTITUTO DE INVESTIGACIONES FILOSÓFICAS

ANÁLISIS EPISTÉMICO DEL SISTEMA DISTRIBUIDO LISTENER-  
TELLER: UNA INSTANCIA DE OMNISCENCIA LÓGICA EN  
INTELIGENCIA ARTIFICIAL Y UNA APLICACIÓN DE LA LÓGICA DE  
JUSTIFICACIONES.

TESIS  
QUE PARA OPTAR POR EL GRADO DE:  
MAESTRO EN FILOSOFÍA DE LA CIENCIA

PRESENTA:  
RODRIGO MEDINA VEGA  
TUTOR:  
DR. FRANCISCO HERNÁNDEZ QUIROZ  
FACULTAD DE CIENCIAS, UNAM

CIUDAD DE MÉXICO, SEPTIEMBRE DEL AÑO 2016



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## Índice

### 0. Introducción. (p. 1)

### 1. Contexto y Fundamentos: sobre el análisis formal de propiedades epistémicas en Sistemas Distribuidos (p. 5)

- 1.1 ¿Qué es la Inteligencia Artificial? (p. 5)
- 1.2 ¿Cuál es la relación entre Filosofía e Inteligencia Artificial? (p. 6)
- 1.3 ¿Cuál es la importancia del estudio filosófico del conocimiento en agentes con inteligencia artificial? (p. 8)
- 1.4 ¿Qué son los Sistemas Distribuidos y por qué son relevantes en el estudio de propiedades epistémicas en Inteligencia Artificial? (p. 10)
- 1.5 ¿Qué es una base de datos y qué papel juega en un SD? (p. 12)
- 1.6 ¿Qué se requiere para estudiar propiedades epistémicas de SDs? (p. 13)
- 1.7 ¿Qué sistema formal y herramientas lógicas se utilizarán para analizar SDs? (p. 14)
  - + 1.7.1 *Conocimiento y creencia en lógica modal epistémica y doxástica* (p. 15)
  - + 1.7.2 *Conocimiento y creencia con justificaciones* (p. 17)
- 1.8 ¿Cómo se modelan SDs con el aparato formal de SI? (p. 18)

### 2. Análisis epistémico en un Sistema Distribuido de Mensajería (p. 25)

- 2.1 ¿Cuáles son las lógicas  $S_5$  y  $KD_{45}$ ? (p. 25)
  - + 2.1.1 *Lógica modal epistémica ( $S_5$  ó  $KT_{45}$ )*. (p. 25)
    - 2.1.1.1 *Sintaxis* (p. 25)
    - 2.1.1.2 *Condiciones semánticas de verdad* (p. 26)
    - 2.1.1.3 *Axiomas* (p. 26)
  - + 2.1.2 *Lógica modal doxástica ( $KD_{45}$ )* (p. 27)
    - 2.1.2.1 *Sintaxis* (p. 27)
    - 2.1.2.2 *Condiciones semánticas de verdad* (p. 28)
    - 2.1.2.3 *Axiomas* (p. 28)
- 2.2 ¿Qué importancia tienen estos axiomas en agentes epistémicos? (p. 29)
- 2.3 ¿Cuál es la conexión de los SI con relación a  $S_5$  y  $KD_{45}$ ? (p. 31)
- 2.4 ¿Qué son los Sistema de Mensajería? (p. 35)

- 2.5 ¿Que otras propiedades son necesarias para el análisis epistémico de los SDM? (p. 37)
- 2.6 *Listener-Teller*. Un SDM con Protocolo de Comunicación (p. 37)
  - + 2.6.1 *Primera Fase*. (p. 39)
    - 2.6.1.1  $R_1$  (p. 42)
    - 2.6.1.2  $R_2$  (p. 49)
  - + 2.6.2 *Segunda Fase*,  $R_3$  (p. 51)

### **3. Omnisciencia Lógica. Un problema de representación epistémica y sus consecuencias en Listener (p. 56)**

- 3.1 ¿Qué es la Omnisciencia Lógica? (p. 56)
- 3.2 ¿Porqué los agentes de  $S_5$  y  $KD_{45}$  son omniscientes? (p. 60)
- 3.3 ¿Por qué es Listener una agente omnisciente? (p. 62)
  - + 3.3.1 *Omnisciencia lógica con relación al uso de axiomas* (p. 62)
  - + 3.3.2 *Omnisciencia lógica con relación a la noción de memoria* (p. 66)
- 3.4 ¿Qué consecuencias tiene la Omnisciencia Lógica en Listener? (p. 66)

### **4. La Lógica de Justificaciones. Una propuesta alterna para enfrentar la Omnisciencia Lógica (p. 72)**

- 4.1 ¿Qué tipo de propuestas existen para enfrentar la Omnisciencia Lógica? (p. 72)
  - + 4.1.1 (a) *Propuestas conceptuales* (p. 73)
  - + 4.1.2 (b) *Condiciones semánticas alternas del conocimiento* (p. 77)
    - 4.1.2.1 *Semántica de Vecindad* (p. 77)
    - 4.1.2.2 *Enfoque Sintáctico* (p. 79)
    - 4.1.2.3 *Estados Imposibles* (p. 80)
    - 4.1.2.4 *Semántica de Kripke con Lógica no Estándar* (p. 80)
    - 4.1.2.5 *Función de Consciencia Lógica* (p. 82)
  - + 4.1.3 (c) *Conocimiento algorítmico* (p. 83)
- 4.2 ¿Qué es la lógica de justificaciones? (p. 87)
  - + 4.2.1 *Sintaxis* (p. 88)
  - + 4.2.2 *Semántica* (p. 88)
  - + 4.2.3 *Axiomatización* (p. 90)
- 4.3 ¿Cómo se expresa el conocimiento y las creencias en esta lógica? (p. 92)

- 4.4 ¿Qué ventajas tiene la lógica de justificaciones frente a  $S_5$  y  $KD_{45}$ ? (p. 97)
  - + 4.4.1 *Pruebas de Omnisciencia Lógica* (p. 97)
  - + 4.4.2 *Conocimiento como creencia verdadera justificada* (p. 99)
    - 4.4.2.1 *Granero Rojo* (p. 101)
      - \* 4.4.2.1.1 *Análisis con lógica epistémica* (p. 102)
      - \* 4.4.2.1.2 *Análisis con lógica de justificaciones* (p. 103)
  - + 4.4.3 *Conocimiento potencial VS Conocimiento Actual.* (p. 104)

## **5. Primeros pasos hacia un análisis de Listener-Teller bajo una perspectiva de la Lógica de Justificaciones (p. 107)**

- 5.1 ¿Qué se sugiere para llevar a cabo un análisis de *Listener-Teller* bajo una perspectiva de agentes no omniscientes? (p. 107)
  - + 5.1.1 *¿Qué relación podría haber entre la evidencia y la memoria en Listener?* (p. 112)
- 5.2. ¿Cómo podrían abordarse los dos problemas identificados en el modelo de *Listener-Teller* con Omnisciencia Lógica bajo la perspectiva que ofrece un modelo de justificaciones? (p. 114)
- 5.3 ¿Qué retos tiene un modelo de justificaciones respecto al análisis y representación de *Listener-Teller* y que vías de investigación se sugieren para enfrentarlos? (p. 117)
  - + 5.3.1 *Tiempo epistémico y tiempo del sistema* (p. 118)
  - + 5.3.2 *Otras acciones de carácter epistémico* (p. 119)
  - + 5.3.3 *Evidencia implícita y explícita* (p. 121)

## **6. Anexo (p. 125)**

6.1 t-MEL en un modelo de justificaciones (p. 125)

## **Bibliografía (p. 130)**



## 0. Introducción

Esta tesis surge en el contexto de la Filosofía de la Información y la Inteligencia Artificial. La Filosofía de la Información se encarga de estudiar aspectos filosóficos sobre la caracterización de la noción de *información* y su dinámica entre agentes comunicativos, mientras que la Inteligencia Artificial se centra en el diseño de agentes cuya capacidad de procesar información se traduce en una capacidad de modificar sus acciones para cumplir los objetivos para los que son construidos.

Dependiendo de los aspectos de la información que quieran ser estudiados, diversas líneas de investigación emergen en la intersección de ambas áreas. El texto presente se ubica en el estudio de aspectos epistémicos de las capacidades y procesos inferenciales respecto a la información en agentes artificiales, con base en el empleo de lenguajes formales.

De acuerdo con este marco de trabajo se defenderá la siguiente hipótesis: *A diferencia de las lógicas modales estándar epistémica y doxástica, la lógica de justificaciones estándar ofrece herramientas más adecuadas para analizar con mayor detalle las capacidades y procesos inferenciales de los agentes con inteligencia artificial que forman parte del Sistema Distribuido Listener-Teller.*

Esta hipótesis se inscribe en el campo de la Filosofía de la Ciencia, específicamente en la rama de Filosofía de las Matemáticas y Lógica de la Ciencia. La hipótesis es relevante para tal rama porque en ella se enfatiza la importancia que tienen los lenguajes formales de tradición filosófica para reflexionar y discutir aspectos específicos de las ciencias, en este caso de un área perteneciente a las Ciencias Computacionales, que es, la Inteligencia Artificial.

La estrategia para defender la hipótesis consiste en estudiar las capacidades y procesos inferenciales de los agentes del sistema *Listener-Teller* bajo dos perspectivas formales. Primero, su estudio se llevará a cabo con las lógicas epistémica y doxástica, con la finalidad de identificar las virtudes y problemas que presentan estos lenguajes como herramientas de análisis sobre los aspectos epistémicos de estos agentes artificiales (especialmente de la agente Listener). Posteriormente se presentará la lógica de justificaciones, y se propondrán estrategias con las que podrían abordarse, con esta lógica, los problemas que surgen con las lógicas epistémica y doxástica.

Se verá que las insatisfacciones surgidas en el análisis de *Listener-Teller*, bajo la perspectiva epistémica y doxástica, se originan a causa de una propiedad que satisfacen todos los agentes cuyas capacidades y procesos inferenciales son modeladas con estos lenguajes, la Omnisciencia Lógica. Un agente que satisface esta propiedad conoce todas las consecuencias lógicas de la información que posee, y conoce todas las expresiones que son válidas en estas lógicas. La Omnisciencia Lógica se considera un problema cuando es un impedimento para modelar agentes epistémicos sin capacidades inferenciales idealizadas.

En este trabajo no se afirmará que la lógica de justificaciones es la herramienta más adecuada que existe para analizar aspectos epistémicos de *Listener-Teller*, ni que supone una solución, y mucho menos definitiva, a la Omnisciencia Lógica, cuando esta se considera un problema. Tampoco se abogará por un reemplazo de las lógicas epistémica doxástica, a cambio de la lógica de justificaciones en este contexto. Más bien, se propondrá que esta es una lógica con herramientas útiles para dar los primeros pasos hacia la representación y estudio de agentes no omniscientes, y una visión más

detallada de sus capacidades y procesos inferenciales, entre los cuales se encuentran los agentes con inteligencia artificial como *Listener* y *Teller*.

A causa de los límites de esta investigación no será posible explorar otras lógicas de corte epistémico que, sin lugar a dudas, podrían resultar muy benéficas en el estudio de agentes no omniscientes. Durante el análisis de *Listener-Teller* tampoco se abordarán versiones más complejas de las lógicas a utilizar, por ejemplo la lógica dinámica epistémica con anuncios públicos, o la lógica de justificaciones con órdenes de plausibilidad. Lo que sí es factible será sugerir estos recientes desarrollos como herramientas en futuras líneas de investigación sobre el tema.

También es pertinente mencionar que hay importantes omisiones en el modelo de *Listener-Teller* que se ha formalizado. Por ejemplo, se encuentra fuera de los límites de este trabajo el abordaje de aspectos temporales y probabilísticos del sistema. Un tratamiento completo del mismo, que incluya todos los aspectos relevantes a estudiar, resulta en una ardua tarea que es imposible incluir en un trabajo como este. Finalmente, considérese que los modelos propuestos con lógica de justificaciones tienen la finalidad de ilustrar una propuesta de análisis epistémico, pero que de ninguna manera se presentan como modelos con una axiomatización completa y correcta .

A continuación, en el Capítulo 1 se presentará el contexto en el que se inscribe este trabajo, y se introducirá un aparato formal para modelar Sistemas Distribuidos. En el Capítulo 2 se introducen las lógicas epistémica y doxástica, y se procede a analizar *Listener-Teller* con ambas. Las virtudes y problemas que surgen en este análisis, a causa de la Omnisciencia Lógica, serán tratadas en el Capítulo 3. En el Capítulo 4 se presenta la lógica de justificaciones y la utilidad de sus herramientas para modelar

agentes no omniscientes. Por último, en el Capítulo 5 se propone el empleo de este lenguaje para analizar *Listener-Teller*, indicando las posibles ventajas de un modelo como este, así como los retos que enfrentaría, y sus respectivas líneas de investigación.

# **1. Contexto y Fundamentos. Sobre el análisis formal de propiedades epistémicas en Sistemas Distribuidos**

Un proyecto importante de la Inteligencia Artificial es el estudio del comportamiento de agentes artificialmente inteligentes. Un agente artificial, i.e. máquina o computadora, se comporta inteligentemente cuando razona sobre su ambiente para inferir de manera explícita la información que anteriormente se encontraba implícita en su memoria, en función de cumplir con los objetivos para los que fue diseñado. Como el comportamiento inteligente puede entenderse en términos de conocimiento y razonamiento, uno de los intereses de la Inteligencia Artificial es el estudio propiedades relativas al conocimiento y manipulación de la información en agentes inteligentes.

Una clase de estos agentes es la de aquellos que forman parte de Sistemas Distribuidos. Estos sistemas serán abordados en el presente capítulo bajo una perspectiva filosófica de la Inteligencia Artificial. Dado que en este trabajo se defiende un tipo de lenguaje formal sobre otro para modelar las capacidades y procesos inferenciales de agentes en un Sistema Distribuido en particular, también se introducirá un marco formal para adaptarlo a ambos tipos de lenguaje.

## **1.1 ¿Qué es la Inteligencia Artificial?**

La Inteligencia Artificial (IA) es la ciencia e ingeniería de diseñar agentes artificialmente inteligentes (p. 2, McCarthy<sup>2</sup>, 2007). De acuerdo con Wooldridge (p. 29, 1999) un agente artificial es una máquina o programa computacional situado en un ambiente, que es capaz de actuar de manera autónoma en el mismo. Se le llama autonomía a la capacidad de actuar sin intervención humana o de otros agentes. Un ejemplo de

agente autónomo es un termostato, que al detectar la temperatura en un espacio por medio de un sensor, activará o desactivará el calor según la temperatura que registre.

A un agente autónomo se le considera inteligente cuando tiene la habilidad de lograr objetivos en su ambiente al comportarse flexiblemente (p. 3, McCarthy<sup>2</sup>). Que un agente inteligente se comporte con flexibilidad significa que actúe de tres maneras distintas, de acuerdo con sus necesidades en cada momento (p. 32, Wooldridge):

- *reactividad*: los agentes responden a los cambios que ocurren en su ambiente.
- *pro-actividad*: los agentes muestran un comportamiento dirigido a un objetivo, tomando la iniciativa para cumplirlo.
- *habilidad social*: los agentes interactúan con otros agentes para satisfacer objetivos en común.

No es recomendable que el agente siempre sea reactivo, pues en algunos momentos se requiere su pro-actividad para alcanzar su objetivo. Por otro lado, ser pro-activo en todo momento podría llevarlo a seguir su objetivo ciegamente sin prestar atención al ambiente; se requiere que algunas veces reaccione a los cambios del mismo (p. 33, Wooldridge). Finalmente el aspecto social es importante, ya que algunas veces se requiere cooperar o negociar con otros agentes, o intercambiar información, para lograr objetivos en colectividad (p. 34, Wooldridge)<sup>1</sup>.

## 1.2 ¿Cuál es la relación entre Filosofía e Inteligencia Artificial?

---

<sup>1</sup> Por ejemplo, un programa anti-virus que solo registra (reactivamente) las infecciones que sufre una computadora no podrá cumplir con el objetivo para el que fue diseñado, a saber, el de mantenerla libre de virus. Si el programa llega a un punto en el que lo único que hace es ejecutar (pro-activamente) comandos para eliminar virus, no podrá eliminar nuevos virus dado que no los habrá registrado. Ergo, el programa requiere combinar ambos tipos de actos para actuar adecuadamente. Ahora bien, el programa será capaz de reconocer una mayor cantidad de virus en la computadora si su base de datos es actualizada periódicamente, lo cual supone trabajar colectivamente con otro(s) programa(s) que le proporcione(n) esta información.

La Filosofía es relevante para el diseño de agentes autónomos e inteligentes. El punto de contacto más importante entre ambos es el estudio de la información. En la *Filosofía de la Información* se estudia el concepto de *información* y las teorías que tratan el procesamiento, administración y uso de la misma. Estas teorías son implementadas por la IA para diseñar agentes que manipulan y administran información para cumplir sus objetivos en su ambiente (p. 542, Schiaffonati, 2003).

La relevancia de la Filosofía de la Información sucede en dos niveles: *fundacional* y *metodológico* (p. 543, Schiaffonati). Fundacionalmente, esta contribuye al proceso de la fundación de la IA. La IA parte de algunos problemas tradicionales de la Filosofía de la Información, como el de la conexión entre conocimiento y acción, y además le aporta conceptos básicos, como la investigación filosófica en torno a la noción de inteligencia. Metodológicamente, la Filosofía de la Información ofrece herramientas a la IA, como las teorías sobre conocimiento, creencias, intenciones, etc., para diseñar lenguajes de comunicación entre agentes.

El papel de la IA en la Filosofía consiste en el uso de métodos y herramientas de la primera para abordar problemas tradicionales y novedosos de la segunda, desde una nueva perspectiva (pp. 543-544, Schiaffonati). En un nivel fundacional, las teorías filosóficas se expresan en términos computacionales para alcanzar un rigor formal más elevado. Por ejemplo, del clásico estudio filosófico de la mente a la investigación de un diseño computacional de la misma para analizar sus mecanismos, capacidades y evolución. En un nivel metodológico, algunos programas y herramientas de la IA se utilizan como medios prácticos para estudiar problemas filosóficos. Por ejemplo, el uso de redes neuronales para evaluar el nivel de coherencia en una explicación científica.

### 1.3 ¿Cuál es la importancia del estudio filosófico del conocimiento en agentes con inteligencia artificial?

Este trabajo parte de la importancia que tienen los trabajos filosóficos de la información para el diseño de agentes artificialmente inteligentes, en particular, el estudio formal del conocimiento proposicional. De este estudio se encarga la *Epistemología*, cuyos fines son especificar las condiciones que deben satisfacerse para que un agente A conozca un enunciado  $\Phi$ , así como los criterios para decidir cuándo las creencias están justificadas, o si las inferencias de una pieza de evidencia a una conclusión son válidas, etc. (p. 345, Lacey & Lee, 2003; p. 1, Sloman, 1979).

El estudio de estas condiciones supone un revisión conceptual de términos como *conocimiento*, *creencia*, *verdad*, *justificación*, etc. Así pues, el trabajo en IA, dirigido al diseño de entidades inteligentes, incluye necesariamente un estudio del conocimiento (p. 1, Sloman)<sup>2</sup>.

La relación entre agentes inteligentes y los estudios epistémicos de la Filosofía de la Información se encuentra en el análisis de propiedades epistémicas de los primeros a través de teorías formales del conocimiento provenientes de la segunda. En la medida que un agente entra en conexión con su ambiente obtiene nueva información de éste. Tal información satisface, o no, ciertas *propiedades epistémicas*, las cuales

---

<sup>2</sup> En la epistemología tradicional el concepto de conocimiento se ha concebido bajo la noción tripartita de *creencia verdadera justificada*. Esta caracterización ha sido muy discutida desde la aparición de trabajos como el de Gettier (1963), donde ilustra dos situaciones en las que el conocimiento no puede ser definido a partir de esta noción. Los contra-argumentos hacia la noción tripartita apuntan a la necesidad de introducir otras nociones para definir adecuadamente el conocimiento, por ejemplo, hay casos de creencia verdadera justificada sin conocimiento, en los que parece que la noción de *evidencia* requiere un estudio más detallado, identificando entre evidencia relevante y no relevante para las creencias.

Así mismo, este problema conceptual es relevante para agentes artificiales, cuando se estudian sus capacidades inferenciales en un contexto epistémico formal que pretende modelar nociones como las de creencia, justificación y verdad. La discusión sobre la noción tripartita será abordada con mayor detalle en el cuarto capítulo (sección 4.4.2), en el contexto de la lógica de justificaciones.

describen características del conocimiento de esta información. Algunas de estas propiedades son las siguientes (pp. 339-340, Lacey & Lee):

- *Precisión*: la información que recibe el agente puede ser inexacta. Por ello requiere de algún mecanismo para decidir la verdad o falsedad de sus creencias.
- *Compleitud*: el agente podría no recibir toda la información que necesita del ambiente, entonces deberá derivar nueva información con base en la que tiene.
- *Consistencia*: para tener un cuerpo de conocimiento consistente el agente debe eliminar la información que entra en conflicto con el resto.

En esta tesis se le llamará *análisis epistémico* a la identificación y discusión de propiedades epistémicas como las anteriores<sup>3</sup>.

Para hablar de conocimiento en IA en un sentido no trivial se requiere el uso de técnicas que permitan a un agente cognitivamente acotado expresar conocimiento de un vasto ambiente sin límites (p. 339, Lacey & Lee). Por ejemplo, no tendría sentido hablar de inconsistencia o imprecisión en un agente omnisciente y omnipotente respecto a sus capacidades cognitivas.

En IA se emplean significados específicos de *conocimiento* y *razonamiento* en el análisis epistémico de agentes (pp. 2-4, KRR)<sup>4</sup>. *Conocimiento* es el contenido expresado a través de un enunciado; un agente *conoce* cuando posee y entiende el significado de enunciados relevantes a su ambiente. *Razonamiento*, por otro lado, se entiende como la manipulación de símbolos de un lenguaje simbólico, con el objetivo

---

<sup>3</sup> Así mismo, un *análisis doxástico* refiere al estudio de propiedades relativas a las creencias. Sin embargo, al mencionar *análisis epistémico* o *propiedades epistémicas* la intención también será referirse a las creencias.

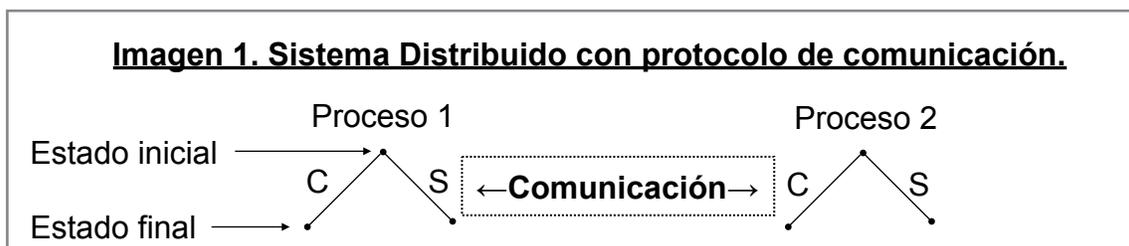
<sup>4</sup> En algunas citas se menciona el nombre del texto en cuestión por medio de abreviaturas. En la bibliografía puede verse a qué texto corresponde cada abreviatura encontrada en este trabajo.

de crear nuevos enunciados con ellos. *Un agente inteligente infiere nuevo conocimiento al razonar sobre la información que posee de su entorno.*

Considerando estas definiciones, se dice que el comportamiento de un agente se determina de acuerdo a los enunciados que conoce. Por ejemplo, cuando sus actos son reactivos, su capacidad para adaptarse a los cambios del entorno donde se desenvuelve depende de que actualice el conocimiento que tiene sobre el mismo al inferir nuevos enunciados acerca de su ambiente (p. 234, Russell & Norving, 2010).

Este trabajo se centra en un análisis epistémico y doxástico para un tipo particular de agentes inteligentes, los *Sistemas Distribuidos*. Las definiciones anteriores de *conocimiento* y *razonamiento* serán suficientes para presentar estos sistemas (secciones 1.4, 1.5 y 1.6), pero se establecerán unas más específicas (1.7) cuando se presenten las herramientas lógico-filosóficas que se emplearán para el análisis.

#### 1.4 ¿Qué son los Sistemas Distribuidos y porqué son relevantes en el estudio de propiedades epistémicas en Inteligencia Artificial?



Se entiende por *Sistema Distribuido* (SD) una colección de *procesos* o *sub-sistemas* que operan en conjunto para alcanzar una meta común. Los procesos están conectados por una red de comunicación, por lo que pueden conocer el estado en el que se encuentran otros procesos a partir de la transmisión de información.

El objetivo común de los SDs es ejecutar *protocolos*. Los protocolos son reglas, o programas, que prescriben, en cada estado del sistema, las acciones que los procesos tienen disponibles. Los protocolos constriñen el funcionamiento de Sistemas Distribuidos, ya que la ejecución de los primeros induce al conjunto de las secuencias de acciones que son admisibles para los segundos. Así pues, se entenderá por *ejecución de protocolo* al hecho que los procesos lleven a cabo secuencias de acciones a las que están condicionados.

Hay distintos tipos de protocolos y se ejecutan con diversos fines. Por ejemplo, se lleva a cabo un protocolo de comunicación para que un proceso transmita su información a otro (Imagen 1<sup>5</sup>). O bien, se emplean protocolos de verificación para conocer el valor de verdad de una fórmula.

Los SDs son relevantes para estudiar propiedades epistémicas de entidades artificiales, pues en ellos se observa el cambio que sufre la información en los procesos durante la ejecución de protocolos. Esta información está compuesta por enunciados del tipo *el valor de la variable x es 0*, que en un lenguaje natural equivale a enunciados como *está lloviendo en San Francisco* (pp. 1 y 3, Fagin & Halpern, 1994).

El estudio en torno al conocimiento es aún más interesante cuando se introduce el factor de incertidumbre. De acuerdo con Halpern y Moses, en un ambiente de esta naturaleza un proceso desconoce, entre otras cosas, el estado actual de otros procesos (p. 38, Meyer & Hoek, 1995). El proceso puede mantener creencias respecto al estado del resto, y estas se transforman durante la elección del protocolo en curso, por ejemplo, cuando otro proceso le envía información sobre su estado presente. Para

---

<sup>5</sup> En este ejemplo se ilustra un SD donde dos procesos lanzan una moneda simultáneamente. Al finalizar, ambos se transmiten el resultado que obtuvieron (*Cara* o *Sello*). En el estado inicial ambos pueden mantener creencias respecto al resultado que obtendrán y al que obtendrá el otro proceso.

expresar el conocimiento de procesos en este tipo de ambientes generalmente se utilizan distribuciones de probabilidad, o una lógica para expresar conocimiento falible.

### 1.5 ¿Qué es una base de datos y qué papel juega en un SD?

Los procesos suelen carecer de memoria, esto es, se informan sobre los hechos que suceden en su estado presente pero lo olvidan en el siguiente estado. Sin embargo, es posible definir una clase procesos que cumple con la propiedad de memoria perfecta (*perfect recall*). En términos informales, un proceso con esta propiedad siempre recuerda el historial de sus estados anteriores. A tal proceso se le llama *base de datos*.

Una base de datos almacena enunciados, los cuales constituyen su conocimiento. A este conocimiento se le llama explícito (p. 25, KRR), o declarado (p. 234, Russell & Norving). Solamente los enunciados que se encuentran en la base de datos son conocimiento explícito para el proceso (p. 32, Nilsson).

Las bases de datos tienen la capacidad de añadir enunciados a su conocimiento cuando se les proporciona nueva información, o cuando infieren aquellos que se encuentran implícitos. Un enunciado implícito se torna explícito si la base lo infiere a partir de los enunciados que conoce explícitamente. Por ejemplo, supóngase que los siguientes enunciados están contenidos en la base de datos X<sup>6</sup>:

- p1: Gaia es un perro.
- p2: Los perros son mamíferos.

Hay un enunciado implícito que X puede derivar,

- p3: Gaia es un mamífero.

---

<sup>6</sup> El ejemplo a continuación es similar a uno de *Knowledge Representation and Reasoning* [KRR] (p. 24).

Cuando X infiere p3 a partir de p1 y p2, en IA se dice que p3 es el nuevo conocimiento que ha adquirido, pues este enunciado se ha vuelto explícito. Puede decirse además que X clasifica a Gaia dentro del conjunto de mamíferos porque conoce el enunciado p3<sup>7</sup>. Así mismo puede explicarse esta inferencia con base en que conoce p2 y p1.

## 1.6 ¿Qué se requiere para estudiar propiedades epistémicas de SDs?

Para esta tarea se requieren dos componentes formales:

- (1) Un formalismo de carácter neutral y general para modelar SDs, de manera que sea posible analizar propiedades de procesos bajo diversos lenguajes formales.
- 2) Un tipo de lógica simbólica (modal, de primer orden, etc.) para expresar adscripciones proposicionales de conocimiento y creencias de procesos, en un modelo formal de SD<sup>8</sup>. Lo que sabe o cree un agente sobre su ambiente, los hechos de la situación en particular en la que debe actuar, y sus objetivos, son representados con fórmulas. El agente utiliza reglas deductivas de inferencia para derivar nuevas fórmulas, y con base en ellas elige las acciones que son apropiadas para lograr sus objetivos (p. 8, McCarthy<sup>2</sup>).

---

<sup>7</sup> De acuerdo con *Artificial Intelligence: foundations of computational agents* (Cambridge, pp. 549-550, 2010), un sistema reconoce enunciados como p1 y p2, e infiere aquellos como p3, porque su base de datos contiene *conocimiento ontológico*. Los diseñadores de sistemas añaden conocimientos sobre clasificaciones y jerarquías de individuos, y esto es lo que habilita que sistemas como X *entiendan* el significado de “perro” y de “mamífero”, así como al jerarquía entre ambos términos (i.e. que “perros” es elemento del conjunto denotado por “mamíferos”).

En la obra citada se dedica una sección al diseño y estudio de sistemas con información ontológica (pp. 563-576). El tema también es brevemente abordado en *Real World Reasoning [RWR]* (p. 25, 2011). Se asume además que el conocimiento ontológico del sistema depende de la conceptualización que su diseñador tenga del mundo (p. 36, Nilsson, 1991). Aunque es interesante, no trataré este tema durante el presente trabajo.

<sup>8</sup> Una de las incógnitas básicas respecto al enfoque epistémico de agentes artificiales es la elección de una teoría formal de la información para expresar sus creencias o conocimiento (p. 2, Sloman). En este trabajo se emplean distintas lógicas simbólicas con este fin. La ventaja de emplear una lógica es que se trata de lenguajes que son comprensibles para los agentes así como para su diseñador (p. 33, Nilsson).

## 1.7 ¿Qué sistema formal y herramientas lógicas se utilizarán para analizar SDs?

Acerca del primer requisito (1), en este trabajo se formalizarán SDs mediante *Sistemas Interpretados* (SI). Un SI se trata de un marco formal para razonar sobre propiedades epistémicas, entre otras, en ambientes con múltiples agentes. Bajo los SI, un SD se define como un ambiente dinámico donde hay agentes (procesos) que ejecutan acciones (pp. 109-110, Halpern et.al., 2003).

La idea detrás de los SI es que existe “*un numero de posibles estados de cosas, además de los estados verdaderos*”<sup>9</sup> (p. 18, Wan, 2014, trad.). En el contexto de los SDs esto significa que, además de los estados espacio-temporales que se suceden en la ejecución de un protocolo, los procesos pueden mantener creencias respecto a futuras transiciones a otros estados, y creencias sobre el estado actual de otros procesos<sup>10</sup>. Se definirán los SDs con base en el formalismo de SI en la sección 1.8.

Para expresar conocimiento y creencias de procesos (2º requisito) se utilizará *lógica modal epistémica y lógica modal doxástica*, y posteriormente la *lógica de*

---

<sup>9</sup> “...based on a number of other possible states of affairs besides the true state of affairs.” (cita original).

<sup>10</sup> Otro de los marcos formales importantes para modelar SDs son los *Procesos de Decisión Markovianos* (*Markovian Decision Process*, MDPs). Los MDPs son formalismos para modelar problemas de decisión donde los agentes transitan a diferentes estados del mismo a través de sus acciones. Shoham y Leyton-Brown (p. 465, 2010) definen los MDPs como juegos estocásticos, pues los agentes mantienen distribuciones de probabilidad sobre la transición a otros estados.

Bajo esta visión un SD se define como un ambiente donde los procesos eligen las acciones que con mayor probabilidad les permiten transitar a estados más favorables respecto a la ejecución del protocolo en curso. Con base en esta idea se diseñan estrategias óptimas para procesos, con el objetivo de regular la selección de sus acciones disponibles en cada estado (Becutti, et.al., 2015, pp. 5-7).

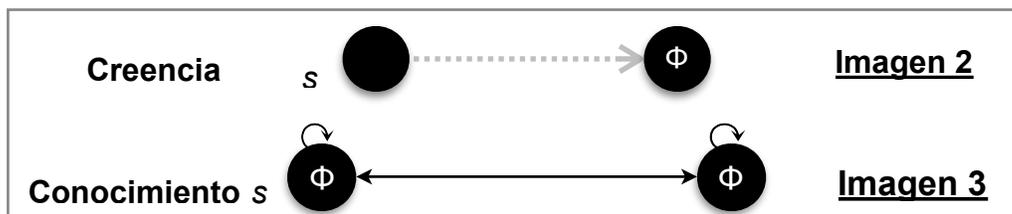
Hay dos razones por las que se podría preferir el formalismo SI sobre el de MDP. La primera es que en la mayoría de los modelos basados en MDPs se representan exclusivamente propiedades cuantitativas de SDs. El numero de estos trabajos se reduce más si solo se toman en cuenta los modelos que emplean lógicas de tradición filosófica respecto al conocimiento (como la lógica epistémica).

Esto no quiere decir que no existan trabajos interesantes al respecto. Por ejemplo, en *Action representation and partially observable planning using epistemic logic* (2002), Lang, et.al. proponen un aparato formal basado en lógica epistémica para procesos que generan planes en ambientes de decisión parcialmente observables (*Partially Observable MDP*, o POMDP). También, en *Verification of Epistemic Properties in Probabilistic Multi-Agent Systems* (2009), Delgado y Benevides proponen una lógica epistémica, temporal y probabilística para Cadenas Markovianas.

La segunda razón es que SI parece ser un formalismo que se adecua con mayor naturalidad a lógicas con operadores epistémicos. Esta razón será más clara durante el siguiente capítulo.

*justificaciones*. Con la primera pueden formularse enunciados del tipo *el agente sabe que Gaia es un perro*, y con la segunda, enunciados del tipo *el agente cree que Gaia es un perro*. Con lógica de justificaciones es posible expresar adscripciones tanto de creencia como de conocimiento justificado, por ejemplo, *el agente está justificado en creer (o conocer el hecho que) que Gaia es un perro dada la evidencia e*.

### 1.7.1 Conocimiento y creencia en lógica modal epistémica y doxástica



En lógica epistémica y doxástica el conocimiento y las creencias se definen con relación a estados posibles de cosas, también llamados mundos posibles. La condición tanto para la creencia, así como para el conocimiento de una fórmula  $\Phi$ , en un estado  $s$  en particular, es una sola, que  $\Phi$  se cumpla en todos los escenarios que son compatibles con  $s$ . En términos más formales,  $\Phi$  debe satisfacerse en todos los estados a los que se tiene acceso desde  $s$ .

Para distinguir la creencia del conocimiento se atiende a las propiedades que cumplen las relaciones entre estos estados (sección 2.1). Se verá que, mientras las relaciones entre estados para modelar el conocimiento requieren que los estados sean auto-accesibles por una relación reflexiva, no es el caso para las relaciones respectivas a las creencias (Imágenes 2 y 3). Esta diferencia en las relaciones refleja un hecho intuitivo: se conocen hechos que son verdaderos, es decir, se trata de creencias que

son verdaderas, y para que lo sean, deben cumplirse en el estado actual, siendo esta una condición que no requieren las creencias de la lógica doxástica.

Estas serán las definiciones de conocimiento y creencia que se adoptarán para los Capítulos 2 y 3. En el siguiente capítulo estas se explicarán con mayor detalle, y se revisarán virtudes específicas que poseen los axiomas de estas lógicas para la reflexión filosófica y análisis epistémico de agentes.

Ambas lógicas son relevantes para el análisis de SDs por dos razones, una histórica y otra metodológica. Para defender la razón histórica, en *Logic and Artificial Intelligence* (p. 851, 2009) Thomason estudia los artículos publicados en la revista *The Journal of Philosophical Logic*, y los compara con trabajos sobre análisis epistémico en agentes inteligentes. El resultado de tal investigación es que estos lenguajes, así como otros en torno al conocimiento práctico y a la revisión de creencias, ha sido de interés común para los trabajos desarrollados en ambas áreas.

La razón metodológica es que las lógicas epistémica y doxástica han resultado de gran utilidad en el contexto de los SI<sup>11</sup>. Fagin y Halpern, pioneros en este tipo de formalismos, construyen en 1994 el primer SI para razonar sobre adscripciones epistémicas de procesos en SDs bajo ambientes de incertidumbre. Los autores combinan axiomas de lógica epistémica y axiomas de teoría de probabilidades para expresar enunciados del tipo *de acuerdo con agente i, el evento  $\Phi$  se mantiene con probabilidad mayor a 'b'* (p. 14).

---

<sup>11</sup> Otros autores como McCarthy (1990, McCarthy1) prefieren la lógica de predicados. La motivación de tal elección es que ésta posee mayor poder expresivo en comparación a otras. Con predicados y funciones, dice McCarthy, es posible expresar una diversidad de hechos relacionados con el conocimiento (pp. 16-17). Esta situación es controversial, pues al ganar expresividad con la lógica de predicados se paga un alto precio a cambio: la falta de efectividad para determinar el valor de verdad de las fórmulas. Si bien este es un tema relacionado, no será abordado a lo largo del presente trabajo.

Existen diversas propuestas contemporáneas a esta, como la de Wei Wan (pp. 44-47, 2014), quien ofrece una lógica para razonar sobre propiedades epistémicas, probabilísticas y temporales. Este lenguaje surge de una extensa familia de lógicas modales en las que es posible formular adscripciones de conocimiento probabilístico que evolucionan en el transcurso del tiempo<sup>12</sup>.

La discusión de estas lógicas no será posible en este trabajo por razones de espacio. Dado que la presente investigación se enfoca a la reflexión de propiedades epistémicas, el campo de trabajo se acotará al análisis con lógicas epistémica y doxástica estándar (Capítulos 2 y 3).

La utilidad de este tipo de lógicas se extiende a otras aplicaciones relacionadas, como el diseño de protocolos de comunicación para procesos de SDs, con la finalidad de garantizar la transmisión de información libre de errores (p. 40, Meyer & Hoek; p. 420, Shoham & Leyton-Brown). Por ejemplo, pueden formularse instrucciones del tipo *el agente x responde el mensaje  $\mu$  del agente z solo si x conoce  $\mu$* .

### **1.7.2 Conocimiento y creencias con justificaciones**

En el Capítulo 3 se abordará un problema particular de estas lógicas, el de la *Omnisciencia Lógica*. En términos generales, a un agente omnisciente se le adscribe conocimiento y creencias sobre información de la cual no tiene evidencia para justificarla. Esto significa que no es posible determinar si un agente omnisciente sabe realmente la información que se le adscribe, o si solamente se trata de algo que podría conocer (creer) en teoría.

---

<sup>12</sup> Hendricks y Symons (p. 160, 2006) llaman a estos lenguajes *los frutos de la segunda generación de lógica epistémica*.

Entre las diversas propuestas exploradas para abordar este problema, existe un mecanismo formal, basado en pruebas sintácticas deductivas, cuyo objetivo es detectar si las inferencias que se le adscriben a un agente se encuentran dentro de sus capacidades inferenciales y por lo tanto están justificadas, o si por el contrario, no hay algo que respalde estas inferencias. Este mecanismo es mejor conocido como la lógica de justificaciones y será presentada durante el 4o Capítulo.

La lógica de justificaciones es relevante en el contexto de este trabajo, pues con ella la IA tiene un mayor acercamiento a la Epistemología, al establecer la relación entre justificación epistémica y *revisión de creencias*. La revisión de creencias en IA consiste en diseñar estrategias para corregir y expandir información de agentes con bases de datos, es decir, cómo la nueva información puede incorporarse a un cuerpo de conocimiento (p. 351, Lacey & Lee).

Hay dos grandes enfoques para la revisión de creencias, ambos provenientes de la Epistemología: el *fundacional* y el *coherentista* (p. 348-349, Lacey & Lee). El enfoque fundacional distingue creencias justificadas por otras creencias, y aquellas que no se justifican. En el enfoque de coherencia las creencias no requieren justificación, siempre y cuando sean consistentes con el resto. Hay fuertes motivos para considerar a la lógica de justificaciones un enfoque fundacionista como método para la revisión de creencias, que serán evidentes durante la presentación de esta lógica.

## **1.8 ¿Cómo se modelan SDs con el aparato formal de SI?<sup>13</sup>**

---

<sup>13</sup> El formalismo de esta sección pertenece a Halpern, et.al. (pp. 117-121). Algunos cambios en la notación se han hecho para mantener consistencia con la notación de otras secciones del trabajo.

En un modelo de Sistemas Interpretados (SI), un Sistema Distribuido (SD) se compone de una colección de procesos  $(1, \dots, n) \in \mathbb{N}$ , ahora llamados agentes, y un conjunto de estados globales, tal que para cada estado global  $S$  hay un estado local para cada agente  $(s_1, \dots, s_n)$ , donde  $1, \dots, n$  son agentes. El conjunto de los estados locales actuales de los agentes forman una descripción del estado global en el que se encuentra el SD en el presente, por lo que puede concebirse un estado global como una visión general de lo que sucede con relación a cada agente, en un momento particular del sistema.

Para analizar el flujo de la información de agentes en estados globales de un SD, son necesarios otros elementos formales de los SI. Estos elementos consisten en una función de valuación para enunciados primitivos ( $V$ ), un conjunto de corridas ( $\mathbf{R}$ ), y un conjunto de relaciones de acceso de equivalencia a estados ( $R_i$ ) para cada agente  $i$ .

Mediante la función  $V$  se les asignan enunciados primitivos ( $\Phi, \Psi$ , etc.) a los estados globales, los cuales describen lo que sucede durante cada estado. Por ejemplo, la expresión  $(SI^*, S) \models \Phi$  se interpreta como *el enunciado  $\Phi$  se satisface en el estado global  $S$  de un sistema particular  $SI^*$* , donde  $\Phi$  puede representar hechos como *el agente 1 envía el mensaje "x" al agente 2*.

Los hechos representados por enunciados primitivos, como se observa arriba, pueden tratarse de acciones, tales como el envío o recepción mensajes entre agentes. Dado que en el siguiente capítulo se presentará un SD donde los agentes efectúan acciones de este tipo, es importante aclarar la noción de *acción* que se estará empleando. En el contexto de SI se entiende por acción un hecho o evento, representado por un enunciado primitivo  $\Phi$ , susceptible de satisfacerse en estados del

sistema, es decir, el valor de verdad de estos enunciados depende de que se cumplan, o no, en los estados globales del sistema (p. 117, Halpern et.al.).

Ahora bien, las acciones no son eventos aislados, pues están íntimamente ligadas a la noción formal de *corrida*. Al conjunto de estados globales que describe el estado del sistema, en cada momento del tiempo durante la ejecución de un protocolo, se le llama *corrida* ( $r$ ,  $run$ ). En otros términos, una corrida es una secuencia que describe las acciones ejecutadas por los agentes (i.e. eventos expresados mediante enunciados primitivos) durante cada estado.

Los estados a los que transitan los agentes, después de haberse ejecutado las acciones en los estados anteriores, contienen la información necesaria para que decidan las acciones que deben tomar. Si un agente lanza una moneda y resulta en Cara, pero el protocolo le pide lanzarla hasta que resulte en Sello, entonces, con base en la satisfacción del evento *la moneda lanzada resulto en Cara* en el estado presente del agente, éste sabrá que debe lanzarla una vez más.

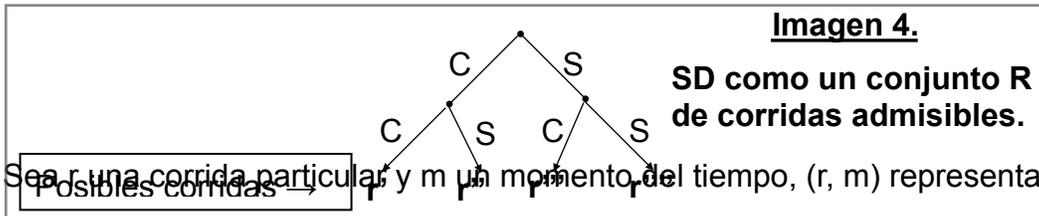
Dado que hay distintas maneras en cómo puede llevarse a término un protocolo, y en cómo no puede llevarse a cabo, hay corridas admitidas por el protocolo, y otras que no lo son. Por tal razón, un SD se define e identifica con el conjunto de sus posibles corridas admisibles para cada protocolo:  $\mathbf{R} = (r, r', r'', \dots)$ .

Véase la Imagen 4<sup>14</sup> para una descripción gráfica de la relación entre corridas y acciones. En la ilustración se observa que acciones del tipo *tirar una moneda que*

---

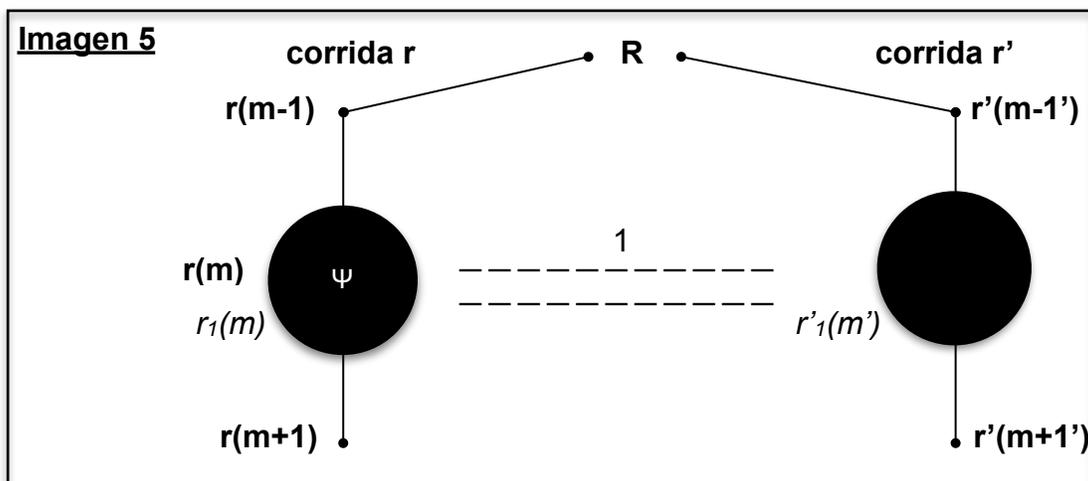
<sup>14</sup> En este ejemplo se presentan las cuatro posibles corridas que el Agente 1 puede obtener al lanzar dos veces una moneda. Así como en la Imagen 1, este puede comunicar sus resultados a otros agentes, y mantener creencias sobre los posibles resultados en cada uno de los estados. En este caso, el agente 1 puede lanzar una moneda en cada estado, con dos posibles resultados (*Cara* o *Sello*).

resulta en cara (o sello) son susceptibles de satisfacerse en cada estado del sistema, los cuales se representan con puntos (Imagen4):



Sea  $r$  una corrida particular, y  $m$  un momento de tiempo,  $(r, m)$  representa un punto del SD, y  $r(m)$  denota el estado global del SD durante la corrida  $r$  en el momento  $m$ ; ya que *punto* y *estado* pueden entenderse similarmente, es el caso que  $(r, m) = r(m)$ , para todo punto y estado. Así mismo, para toda  $r, m$  y agente  $i$ , hay un estado local para  $i$  tal que  $r_i(m)$ . Dado que símbolos como “S” y “s<sub>i</sub>” no son lo suficientemente expresivos para representar estados como momentos particulares de corridas, estos serán reemplazados por símbolos del tipo  $r(m)$  y  $r_i(m)$  respectivamente.

Las relaciones de acceso ( $R_i$ ), por otro lado, juegan un papel de gran importancia para representar en el modelo el punto de vista que tienen los agentes durante cada estado global del sistema, es decir, sus estados locales. Supóngase que el agente 1 no sabe  $\Psi$ , donde  $\Psi$  representa el hecho que *el agente 2 envía el mensaje x al agente 3*. Esta situación puede ilustrarse así (considerando solamente al agente 1):



Donde ( $\equiv$ ) es una relación para el agente 1 que denota equivalencia entre estados tal que son indistinguibles para éste, y  $r, r' \in \mathbf{R}$ . Nótese que el estado global  $r(m)$  es también el estado local presente del agente 1, esto es,  $r_1(m)$ . La diferencia reside en que en  $r_1(m)$  el agente 1 tiene acceso a un estado alternativo al presente, es decir de una corrida y tiempo alternos  $r'_1(m')$ , donde la proposición  $\Psi$  no se satisface. Intuitivamente, el agente considera posible un estado en el que el agente 2 no envía el mensaje  $x$  al agente 3. Por esta razón sucede que en el estado global  $r(m)$ , desde el punto de vista del agente 1, hay incertidumbre respecto al envío del mensaje  $x$ .

Que la relación ( $\equiv$ ) denote equivalencia significa ésta que satisface tal propiedad, formalmente, la relación que hay entre los estados  $r(m)$  y  $r'(m')$  cumple con las propiedades de reflexividad, simetría y transitividad (o alternativamente, euclidianismo). La satisfacción de estas características en las relaciones de los modelos SI no es un hecho trivial, ya que éstas inducen a reflexiones interesantes sobre la información que los procesos mantienen en sus estados locales. Que las relaciones de acceso de un agente entre estados satisfagan equivalencia significa, en el caso particular de la Imagen 5, que  $r(m)$  y  $r'(m')$  son para el agente 1 estados indistinguibles en términos de información<sup>15</sup>.

Para verificar lo anterior basta con observar que, en ambos estados, el agente tiene acceso a un estado donde  $\Psi$  se satisface y otro donde no se satisface. Es decir,  $r(m)$  y  $r'(m')$  le son indistinguibles dado que en ambos tiene incertidumbre sobre el envío del mensaje  $x$ . Finalmente, hay un aspecto relevante que debe enfatizarse sobre la situación ilustrada en la Imagen 5; es perfectamente posible que en los estados

---

<sup>15</sup> En el siguiente capítulo se presentará con detalle el papel que juegan las relaciones de equivalencia y cómo es posible reflexionar sobre el conocimiento de agentes por medio de las mismas. Así mismo se implementarán relaciones *doxásticas* para reflexionar sobre creencias en estos modelos.

anterior y posterior a  $r(m)$ , el agente 1 tenga acceso a estados alternos a  $r(m-1)$  y  $r(m+2)$  donde su información sea totalmente consistente con el estado global en el que se encuentra el sistema en esos momentos del tiempo.

A continuación, se ofrece una definición sistemática de los elementos anteriores  $\{(L_i, N_i)_{i \in N}, P, EG, \Sigma, V\}$ <sup>16</sup> donde,

- $L_i$  es el conjunto de estados locales para el agente  $i \in N$ , tal que  $\forall r_i(m), r_i(m) \in L_i$ .
- Si  $P$  es un protocolo que prescribe la ejecución de las acciones  $\Phi, \Psi$  para el estado  $r^*(m)$  y  $r^*(m+1)$  respectivamente, tal que  $r^* \in \mathbf{R}$ , entonces la corrida  $r^*$  se define como la secuencia de eventos  $(\Phi, \Psi)$  hasta el momento  $m+1$ .
- El producto cartesiano de  $L_1 \times \dots \times L_n$  equivale a un conjunto de posibles tuplas  $(r_1(m), \dots, r_n(m))$ , donde cada tupla conforma un estado global (EG).
- $\Sigma$  es un conjunto de estados globales iniciales del sistema, tal que  $\Sigma \subseteq EG$ . Es importante mencionar que para cada  $P$ , existe una corrida por cada estado inicial  $r(s)$  admisible, donde  $r(s) \in \Sigma$ .
- Si  $AP$  es un conjunto de enunciados atómicos que describen estados,  $V \subseteq EG \times AP$  es una función de valuación que relaciona EGs con enunciados atómicos.

En este capítulo se situó al lector en el contexto del estudio filosófico sobre aspectos epistémicos de agentes con inteligencia artificial. La estrategia para hacerlo consistió en explicar la intersección que hay entre la Inteligencia Artificial y la Filosofía de la Información (en particular la Epistemología). Posteriormente se presentó la clase de agentes artificiales que forman parte de SDs, así como la importancia de estudiar, en

---

<sup>16</sup> La notación a continuación pertenece a Wan (pp. 18-19), salvo un par de cambios, con el objetivo de mantener consistencia en la notación para secciones posteriores del trabajo.

términos epistémicos, el flujo de información que ocurre entre ellos. Finalmente se introdujo el marco formal SI, útil para estudiar SDs bajo distintos lenguajes formales, y el cual, será de gran utilidad para analizar en el siguiente capítulo el SD *Listener-Teller* con lógicas epistémica y doxástica, lenguajes que serán presentados a continuación.

## 2. Análisis epistémico en un Sistema Distribuido de Mensajería

Los Sistemas de Mensajería son una clase de Sistemas Distribuidos (SDs) donde se ejecutan protocolos de comunicación. En ellos, los agentes envían y reciben información con el objetivo de conocer más sobre su ambiente. En este capítulo se analizan propiedades epistémicas y doxásticas en un sistema de este tipo, en el que un agente (Teller) envía información a otra (Listener) durante cada estado del protocolo.

Primero se introducen los elementos necesarios para llevar a cabo este análisis. Estas son las lógicas epistémica ( $S_5$ ) y doxástica ( $KD_{45}$ )<sup>17</sup>. Posteriormente se introduce la relación formal entre estas y los Sistemas Interpretados (SI). Después se presentan los sistemas de mensajería, y por último, algunas propiedades temporales interesantes desde el punto de vista del conocimiento.

### 2.1 ¿Cuáles son las lógicas $S_5$ y $KD_{45}$ ?

#### 2.1.1 Lógica modal epistémica ( $S_5$ ó $KT_{45}$ )

##### 2.1.1.1 Sintaxis

$$\Phi, \Psi ::= p \mid \sim\Phi \mid \Phi \vee \Psi \mid K\Phi \mid K_i\Phi$$

Donde  $\Phi$  es una fórmula que puede ser atómica o parte de una fórmula compuesta con las conectivas de lógica proposicional. Nótese que con la negación ( $\sim$ ) y disyunción ( $\vee$ ) se definen el resto de las conectivas, por ejemplo,  $(\sim p \vee q) \leftrightarrow (p \rightarrow q)$ .  $K$  es un operador que denota conocimiento, y  $K\Phi$  es una fórmula epistémica cuya interpretación es *se conoce que  $\Phi$* . Finalmente,  $K_i\Phi$  es una fórmula que adscribe el conocimiento  $\Phi$ , específicamente al agente  $i$ .

---

<sup>17</sup> La presentación de ambas lógicas esta basada en la de van Benthem en *Modal Logic for Open Minds* (pp. 135-139, y 147-149, 2010).

### 2.1.1.2 Condiciones semánticas de verdad

En la semántica de  $S_5$  se emplean modelos de Kripke ( $M$ ) para evaluar fórmulas epistémicas y de lógica proposicional. Estos modelos se componen de un conjunto  $W$  de estados (mundos) posibles, relaciones de acceso del tipo " $\kappa \rightarrow$ " para estados, y una función ( $V$ ) de valuación que asigna valores de verdad a fórmulas en cada estado del modelo, en términos formales  $M = (W, \kappa \rightarrow, V)$ . Si  $s$  y  $t$  son estados pertenecientes a  $W$ , entonces se satisfacen las siguientes condiciones:

- $(M, s) \models p$  sii la función  $V$  asigna el valor verdadero a  $p$  en  $s$ .
- $(M, s) \models \sim\Phi$  sii  $(M, s) \not\models \Phi$ , es decir,  $(M, s) \models \Phi$  no se cumple.
- $(M, s) \models \Phi \vee \Psi$  sii  $(M, s) \models \Phi$  ó  $(M, s) \models \Psi$
- $(M, s) \models \Phi \& \Psi$  sii  $(M, s) \models \Phi$  y  $(M, s) \models \Psi$
- $(M, s) \models \Phi \rightarrow \Psi$  sii cuando  $(M, s) \models \Phi$  es verdadero entonces  $(M, s) \models \Psi$  también.
- $(M, s) \models \Phi \leftrightarrow \Psi$  sii cuando  $(M, s) \models \Phi$  es verdadero entonces  $(M, s) \models \Psi$  también, y si  $(M, s) \models \Psi$  es verdadero entonces  $(M, s) \models \Phi$  también.
- $(M, s) \models K\Phi$  sii  $\forall t$  si  $s \kappa \rightarrow t$  entonces  $(M, t) \models \Phi$
- $(M, s) \models K_i\Phi$  sii  $\forall t$  si  $s \kappa \rightarrow_i t$  entonces  $(M, t) \models \Phi$ , donde  $\rightarrow_i$  es relación de acceso para el agente  $i$ .

### 2.1.1.3 Axiomas

Los estados con relaciones de acceso conforman *marcos de Kripke*. Las propiedades que satisfacen estas relaciones, en un marco, determinan cuáles son los axiomas válidos e inválidos en el mismo. Por ejemplo, en marcos donde las relaciones satisfacen la propiedad  $X$ , serán válidos los axiomas que requieren relaciones del tipo  $X$  para cumplirse en todos los mundos del mismo, independientemente de la asignación

de verdad de la función V. Véase a continuación la condición de validez para cada axioma de  $S_5$ <sup>18</sup>. Sean  $x, y, z$  estados:

- Todas las tautologías de lógica proposicional.
- (K)  $K(\Phi \rightarrow \Psi) \rightarrow (K\Phi \rightarrow K\Psi)$ . Es válido en todos los marcos de Kripke.
- (T)  $K\Phi \rightarrow \Phi$  (*veracidad*). Es válido en marcos reflexivos:  $\forall x. (x \kappa \rightarrow x)$
- (4)  $K\Phi \rightarrow KK\Phi$  (*introspección positiva*).

Es válido en marcos transitivos:  $\forall xyz. [(x \kappa \rightarrow y) \& (y \kappa \rightarrow z)] \rightarrow (x \kappa \rightarrow z)$

- (5)  $\sim K\Phi \rightarrow K\sim K\Phi$  (*introspección negativa*). Es válido en marcos euclidianos:  $\forall xyz. [(x \kappa \rightarrow y) \& (x \kappa \rightarrow z)] \rightarrow (y \kappa \rightarrow z)$
- Adicionalmente, (G) *si  $\Phi$  es válida entonces  $K\Phi$  también lo es*, se trata de una regla que se cumple en todos los marcos.

En los marcos  $S_5$  de Kripke las relaciones a estados posibles satisfacen reflexividad, transitividad y euclidianismo, y por lo tanto, en este tipo de marco se cumplen los axiomas (T), (4) y (5). La satisfacción de estas tres propiedades significa que las relaciones en marcos de  $S_5$  son de *equivalencia*.

### 2.1.2 Lógica modal doxástica ( $KD_{45}$ )

#### 2.1.2.1 Sintaxis

$$\Phi, \Psi ::= p \mid \sim\Phi \mid \Phi \vee \Psi \mid K\Phi^{19} \mid K_i\Phi \mid B\Phi \mid B_i\Phi$$

La construcción sintáctica en  $KD_{45}$  es muy similar a la de  $S_5$ . La diferencia es que se introduce el operador B, que denota creencia.  $B\Phi$  significa *se cree que  $\Phi$* , y  $B_i\Phi$  adscribe al agente i la creencia  $\Phi$ .

<sup>18</sup> Así mismo, la validez de los axiomas de  $KD_{45}$  y lógica de justificaciones también se definirá por medio de las propiedades que satisfacen las relaciones de acceso de los marcos.

<sup>19</sup> Se reiteran formulas del tipo  $K\Phi$ , ya que algunos axiomas doxásticos los incluyen.

### 2.1.2.2 Condiciones semánticas de verdad

Las mismas que en  $S_5$ , pero la relación de acceso es para estados posibles doxásticos, por lo cual, se empleará " $B \rightarrow$ " para distinguirla de las relaciones tipo " $K \rightarrow$ ", además:

- $(M, s) \models B\Phi$  sii  $\forall t$  si  $s \rightarrow t$  entonces  $(M, t) \models \Phi$ , siendo un caso análogo para  $B_i\Phi$

### 2.1.2.3 Axiomas

- Todas las tautologías de lógica proposicional.
- (K)  $B(\Phi \rightarrow \Psi) \rightarrow (B\Phi \rightarrow B\Psi)$ . Es válido en todos los marcos de Kripke
- (D)  $B\Phi \rightarrow \sim B\sim\Phi$ . Es válido en marcos seriales:  $\forall x\exists y. (x \rightarrow y)$ <sup>20</sup>.
- (4)  $B\Phi \rightarrow BB\Phi$  o en su versión fuerte. Valido en marcos transitivos.
- (5)  $\sim B\Phi \rightarrow B\sim B\Phi$  en su versión fuerte. Válido en marco euclidianos<sup>21</sup>.

Los marcos doxásticos no son reflexivos, pues es un hecho intuitivo que las creencias pueden ser falsas.

Adicionalmente, en el caso de las creencias se introducirá el operador modal de posibilidad " $\diamond$ " para construir fórmulas del tipo  $B_i\diamond\Phi$ , con el objetivo de expresar enunciados como *el agente i cree que  $\Phi$  es posible*. La condición semántica de  $\diamond$  es la siguiente, sean  $s$  y  $s'$  estados:

- $s \models \diamond\Phi$  sii  $\exists s'. (s \rightarrow s') \ \& \ (s' \models \Phi)$ .

---

<sup>20</sup> Es importante notar que  $KT_{45} \models D$ , ya que todos los marcos reflexivos son trivialmente seriales, entonces,  $S_5$ ,  $KT_{45}$  y  $KDT_{45}$  son lógicas equivalentes, en el sentido que las relaciones de los marcos donde se satisfacen estos tres conjuntos de axiomas cumplen con las mismas propiedades. Nótese que las tres son lógicas epistémicas, dado que contienen el axioma (T).

<sup>21</sup> Para los axiomas (4) y (5) hay versiones fuertes, a saber, la creencia (no creencia) de  $\Phi$  implica el conocimiento sobre tal creencia (falta de creencia):

(4')  $B\Phi \rightarrow KB\Phi$

(5')  $\sim B\Phi \rightarrow K\sim B\Phi$

Los axiomas son válidos en marcos transitivos y euclidianos respectivamente, pero en unos donde se cumpla  $B \rightarrow \subseteq K \rightarrow$ .

Es importante indicar que el operador “ $\diamond$ ” se interpretará como una etiqueta existencial para la fórmula que antecede. Así pues, el significado de fórmulas como  $s \models B_i \diamond \Phi$  será: *para todo estado doxásticamente accesible desde  $s$ , el enunciado “existe al menos un estado al que puede accederse doxásticamente donde  $\Phi$  es el caso” se cumple.*

La introducción de “ $\diamond$ ” se debe a que, en el ejemplo de Sistema Distribuido que se presentará durante la sección 2.6, es necesario mantener la consistencia en las creencias de los agentes. Sean  $\Phi$  y  $\Psi$  dos fórmulas que describen el ambiente de manera distinta. Un agente bajo cierto nivel de incertidumbre sobre el ambiente no sabría cuál de estas es la descripción correcta del mismo, y por ende, puede creer que ambas podrán tratarse de la correcta. Pero,  $B\Phi$  implicaría  $B\neg\Psi$ , y  $B\Psi$  implicaría  $B\neg\Phi$ , lo cual llevaría a una contradicción, y el agente podría terminar creyendo cualquier cosa por *principio de explosión*.

Sin embargo, no hay contradicción al afirmar que  $\Phi$  y  $\Psi$  son, *posiblemente*, el las descripciones correctas del ambiente:  $B\diamond\Phi \ \& \ B\diamond\Psi$ . Nótese que solo se requiere un estado donde se satisfaga una fórmula para que esta sea posible ( $\diamond$ ).

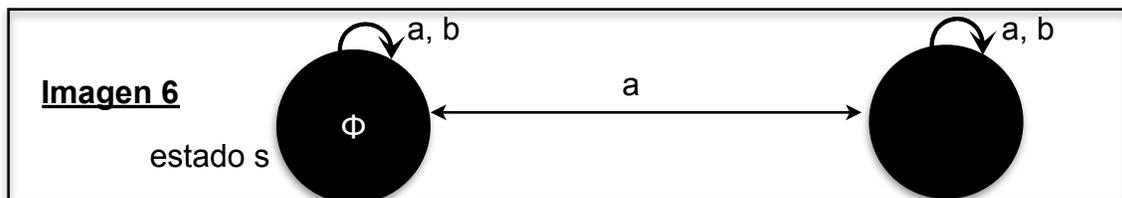
## 2.2 ¿Qué importancia tienen estos axiomas en agentes epistémicos?

Los axiomas de  $S_5$  son interesantes para reflexionar sobre el conocimiento de sistemas computacionales (p. 189, Meyer, 2001). Por ejemplo, de acuerdo con el axioma de veracidad (T) ( $K\Phi \rightarrow \Phi$ ) se expresa que todo lo que un agente puede conocer son hechos verídicos, i.e. en los estados del agente se cumple la propiedad de reflexividad. Otro axioma interesante es (5); con éste se expresa conciencia sobre el conocimiento

que se ignora, y un agente puede identificar que hay fórmulas que no figuran en su información (p. 183, Meyer)<sup>22</sup>.

De (5) y (T) se deriva  $(\sim\Phi \rightarrow K\sim K\Phi)$ <sup>23</sup>, y se le conoce como el teorema de la *sabiduría socrática*, ya que con éste se expresa la propia ignorancia (p. 146, Hendricks & Symons). La fórmula dice que *si  $\Phi$  no es el caso, entonces se sabe que no se conoce  $\Phi$* . El conocimiento de un agente está garantizado por el hecho de que el agente no puede adscribirse conocimiento que no puede poseer, y por lo tanto, no es blanco de ataques por parte del escepticismo epistémico.

Fagin es uno de los autores más importantes que motivan el uso de tales axiomas para agentes con inteligencia artificial (p. 882, Thomason). El autor defiende el uso de  $S_5$  porque permite formalizar el razonamiento de agentes que reflexionan sobre el conocimiento e ignorancia de otros (p. 189, Meyer). Por ejemplo, en un marco de  $S_5$  donde sean satisfechas ciertas condiciones de acceso para agentes (Imagen 6), pueden cumplirse condiciones como *si  $(M, s) \models K_b\Phi$  entonces  $(M, s) \models K_a(K_b\Phi \vee K_b\sim\Phi)$* , donde el agente a reflexiona sobre el conocimiento del agente b. O bien, el agente a, quien ignora el valor de verdad de  $\Phi$ ,  $(M, s) \models (\sim K_a\Phi \ \& \ \sim K_a\sim\Phi)$ , puede reflexionar sobre su propia ignorancia,  $(M, s) \models K_a(\sim K_a\Phi \ \& \ \sim K_a\sim\Phi)$ :



<sup>22</sup> Una versión más débil que  $S_5$  es  $S_4$ , donde  $S_4 = S_5 - (5)$ . Bajo  $S_5$  los agentes son epistémicamente más capaces que en  $S_4$ . Los primeros están conscientes de lo que conocen e ignoran, mientras que los segundos solamente sobre lo que conocen (p. 146, Hendricks & Symons).

<sup>23</sup> La fórmula es válida en marcos simétricos. La simetría es el resultado de combinar las propiedades euclidiana y reflexiva.

Algo similar sucede con los axiomas de  $KD_{45}$ , en particular, esta lógica es útil para expresar conocimiento falible. Este tipo de conocimiento (i.e. creencias) debe mantener algunas propiedades básicas, por ejemplo, la consistencia, para lo cual se emplea el axioma (D)  $B\Phi \rightarrow \sim B\sim\Phi$ . Con este axioma se evita que sucedan casos como  $(B\Phi \ \& \ B\sim\Phi) \rightarrow B\perp$ <sup>24</sup>, y por ende, terminar creyendo cualquier cosa.

En este trabajo se utilizará el operador de posibilidad ( $\diamond$ ) para formar expresiones del tipo  $B_i\diamond\Phi$  (el agente  $i$  cree que  $\Phi$  es posible). Ello permitirá abordar casos donde fórmulas del tipo  $(B_i\diamond\Phi \ \& \ B_i\diamond\sim\Phi)$  se satisfagan en un solo estado, sin contradicción alguna. La razón es que la satisfacción semántica de " $\diamond$ " solamente requiere acceso a un estado donde se cumpla la fórmula que procede al operador.

Con la combinación de los operadores  $K$  y  $B$  se abre la posibilidad de modelar otras nociones computacionales interesantes para agentes artificiales, como sucede con las versiones fuertes de los axiomas (4) y (5) para  $KD_{45}$ . Por ejemplo,  $BK_i\Phi$  es una actitud epistémica que refleja un nivel de certeza intermedio entre el conocimiento y la creencia, esta se lee como *i cree que sabe  $\Phi$*  (p. 148, van Benthem).

### 2.3 ¿Cuál es la conexión de los SI con relación a $S_5$ y $KD_{45}$ ?<sup>25</sup>

---

<sup>24</sup> Existe una técnica para hacer de  $KD_{45}$  una lógica doxástica más expresiva. Esta consiste en reemplazar las relaciones de acceso estándar a estados posibles por relaciones de plausibilidad. En términos generales, el agente ordena sus creencias con base en una jerarquía de plausibilidad para estados. Por ejemplo, para evitar contradicción en  $B_i(\Phi \ \& \ \sim\Phi)$ ,  $i$  podría considerar que  $\sim\Phi$  se satisface en estados menos plausibles que  $\Phi$ , y entonces  $\Phi$  es la creencia dominante.

Por razones de espacio no se abordará esta técnica. Más información sobre relaciones de plausibilidad en Halpern (pp. 288-291) y van Benthem (pp. 149-151).

<sup>25</sup> El contenido de esta sección pertenece principalmente a Halpern, et.al. (pp. 116-121). Como texto de apoyo se ha revisado también *Interpreted Systems and Epistemic Logic* (Sergot, 2008, pp. 2-3). Hay algunas modificaciones en el formalismo para mantener consistencia con la notación de otras secciones.

Como se ha visto en la sección 1.8, los SD, en el contexto de los SI, son modelos compuestos de corridas conformadas por puntos, también llamados estados globales. En cada estado global de cada corrida los agentes tienen acceso a estados alternos del mismo, con los cuales se representa su punto de vista de cada estado global. Además, para todo estado global y estados alternos, hay enunciados primitivos que representan acciones, entendidas como eventos susceptibles de satisfacerse en estos estados.

Con base en esta idea, un SD (bajo SI) se define como el conjunto  $(\mathbf{R}, V)$ , donde  $\mathbf{R}$  es un conjunto de corridas con estados (puntos), con relaciones de acceso para estados alternos, y  $V$  una función para evaluar enunciados primitivos en los estados de las corridas. Dicho de otro modo,  $V$  es una función que induce a interpretaciones sobre los puntos que conforman las corridas (p. 117, Halpern, et.al.). Dado que las relaciones de agentes a posibles estados del sistema son de equivalencia, es posible asociar un SI con un modelo  $(M)$   $S_5$  de Kripke,  $M = (W, \kappa \rightarrow, V)$ , para todo SI y todo  $M$  (p. 117, Halpern, et.al.).

De acuerdo con tal asociación, el conjunto  $W$  de mundos posibles en  $M$  corresponde al conjunto de estados globales de corridas en SI, y viceversa. Además, ahora es evidente que la razón por la cual se adoptan relaciones de equivalencia entre los estados globales se debe a que son de suma importancia para habilitar la reflexión sobre el conocimiento de agentes mediante la lógica  $S_5$ ; esto es, las relaciones  $\kappa \rightarrow$  corresponden a las relaciones del tipo  $R_i$ , y viceversa.

Así pues, la correspondencia entre la función de valuación de modelos SI y de  $M$  de sigue naturalmente, pues toda fórmula primitiva (enunciado) que la función  $V$  de SI

haga verdadera en un estado  $r(m)$ , también será verdadera bajo la función  $V$  de  $M$ , y viceversa:

- $\forall r(m) \forall \Phi. (SI, r(m)) \models \Phi \text{ sii } (M, r(m)) \models \Phi$

Véase que  $M$  y  $SI$  representan el mismo sistema de estados con relaciones de acceso.

*“La relación de acceso epistémico se emplea para asociar un modelo de Kripke con un SI, de manera que se exprese un modelo epistémico”* (p. 19, Wan, trad.)<sup>26</sup>.

Consecuentemente, las condiciones de satisfacción semántica en  $SI$  y  $M$  son las mismas:

- $(SI, r(m)) \models p \text{ sii } V$  hace verdadera a  $p$  en el estado  $r(m)$
- Para la verdad y validez de fórmulas el caso es el mismo. En particular, que la condición para la validez sea similar significa que todos los axiomas de  $S_5$  y tautologías proposicionales se cumplen en todo modelo  $SI$  (p. 119, Halpern, et.al.)
- $(SI, r_i(m)) \models K_i\Phi \text{ sii } \forall r'_i(m'). [r_i(m) = r'_i(m') \rightarrow (IS, r'_i(m')) \models \Phi]$ <sup>27</sup>

Hechos como *el agente  $i$  no sabe  $\Phi$*  significa que, para el agente  $i$ , el sistema podría encontrarse en un punto donde  $\Phi$  no se mantiene. Como se ha mencionado antes (sección 1.8), lo que saben los agentes, en un modelo  $SI$ , se determina con base en la información que contiene su estado local, por lo cual,  $i$  no puede distinguir entre dos puntos del sistema (estados globales) en los que el estado local es el mismo (pp.

---

<sup>26</sup> “The epistemic accessibility relation is defined to associate a Kripke model with a given interpreted system  $IS$  to express an epistemic model.” (cita original).

<sup>27</sup> Tanto el símbolo “=” como “ $K \rightarrow$ ” denotan relaciones de acceso de equivalencia entre estados posibles.

116-117, Halpern et.al.)<sup>28</sup>. Entonces que  $i$  no sepa  $\Phi$  significa que tiene acceso a un estado donde no se cumple  $\Phi$ , el cual no distingue de su estado local presente.

Formalmente,  $K_i\Phi$  es el caso en  $r(m)$  desde el punto de vista del agente  $i$  (esto es, en  $r_i(m)$ ), cuando  $\Phi$  es verdadero en todos los posibles estados globales alternos a  $r(m)$  a los que  $i$  tiene acceso mediante una relación de equivalencia, es decir, todos los estados posibles que contempla son equivalentes para éste en términos de información (p. 118, Halpern, et.al.). En otros términos,  $K_i\Phi$  se cumple  $r_i(m)$ , solo si  $r(m)$  es para  $i$  un estado donde el SD puede encontrarse solamente cuando  $\Phi$  es verdadera, y  $\Phi$  es verdadero en  $r_i(m)$  cuando a todos los estados a los que tiene acceso desde  $r(m)$  se cumple  $\Phi$  (pp. 257-258, Stalnaker, 1999).

A excepción de que en  $KD_{45}$  se suprime la reflexividad y se emplean relaciones del tipo  $B \rightarrow_i$ , la condición para el operador  $B_i$  en un SD se sigue de manera similar (p. 119, Halpern, et.al.).  $B_i\Phi$  es el caso en  $r_i(m)$  cuando  $\Phi$  es verdadero en todos los estados a los que  $i$  tiene acceso doxástico a partir de este:

- $(SI, r_i(m)) \models B_i\Phi$  sii  $\forall r'_i(m'). [r_i(m) B \rightarrow r'_i(m')] \rightarrow (IS, r'_i(m')) \models \Phi$

Puede observarse que la condiciones semánticas para creencia y conocimiento de agentes en SDs se comportan como las anteriormente definidas para  $S_5$  y  $KD_{45}$ .

---

<sup>28</sup> "Las afirmaciones como ' $R$  no sabe  $\Phi$ ' significa que, hasta donde  $R$  esta consciente, el sistema podría estar en un punto donde no  $\Phi$  se mantiene... .. Pensamos en el conocimiento de  $R$  estando por su estado local, tal que  $R$  no puede distinguir entre dos puntos del sistema en los cuales tiene el mismo estado local, y puede distinguir aquellos puntos donde sus estados locales difieren unos de los otros." (trad.)

"... statement such as ' $R$  does not know  $\Phi$ ' means that, as far as  $R$  is concerned, the system could be at a point where  $\Phi$  does not hold... .. We think of  $R$ 's knowledge as being determined by its local state, so that  $R$  cannot distinguish between two points of the system in which it has the same local state, and it can distinguish points in which its local state differs." (cita original).

A continuación se presenta una clase de SDs: *los Sistemas de Mensajería*. En ellos el conocimiento de agentes sufre transformaciones con el envío y recepción de información. Estos se abordarán como modelos de SI asociados con modelos de mundos posibles de  $S_5$  y  $KD_{45}$ .

## 2.4 ¿Qué son los Sistema de Mensajería?

En un *Sistema Distribuido de Mensajería* (SDM) los agentes ejecutan protocolos de comunicación, con los cuales se conforman secuencias de acciones sobre envío y recepción de mensajes entre agentes, admitidas por el SD, y cuyo fin es cumplir con un objetivo en particular, que está relacionado con la transferencia de información<sup>29</sup>. Estas acciones causan cambios en el sistema a lo largo del tiempo, los cuales a su vez, transforman el conocimiento de los agentes. Estos cambios son registrados en historiales<sup>30</sup>.

Para definir formalmente un SDM se contemplan los conjuntos  $(\Sigma_i, \text{MSG})$  para cada agente  $i = 1, \dots, n$ .<sup>31</sup> Donde  $\Sigma_i$  es el conjunto de estados iniciales para  $i$  y  $\text{MSG}$  es un conjunto de mensajes. Para indicar el envío y recepción de mensajes  $\partial \in \text{MSG}$ , se utilizan las siguientes expresiones:

- *envía*( $\partial, j, i$ ), que significa *i envía el mensaje  $\partial$  a  $j$* .
- *recibe*( $\partial, j, i$ ), que significa *i recibe el mensaje  $\partial$  de  $j$* .

---

<sup>29</sup> Thomason menciona que las lógicas modales son útiles en Sistemas Distribuidos con este tipo de protocolos (p. 883).

<sup>30</sup> Para sistemas más complejos se consideran protocolos e historiales para el ambiente donde subsiste el sistema, y un contexto en el que se describen las características de este ambiente. Los elementos del ambiente se omitirán para delimitar el análisis a aspectos epistémicos de los agentes.

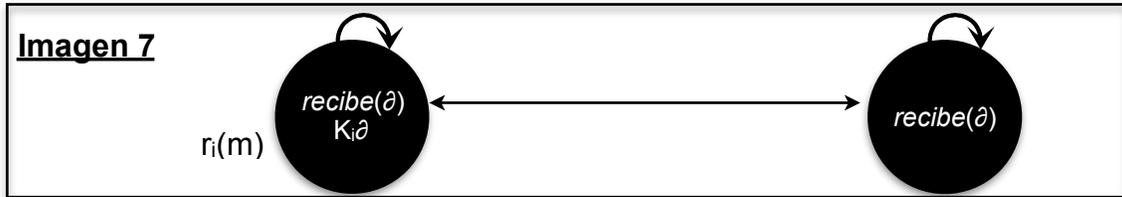
<sup>31</sup> La notación presentada corresponde a los capítulos 4 y 5 de *Reasoning About Knowledge* (Halpern et.al.) excepto donde se indica.

Una historia para  $i$  es una secuencia de elementos, donde el primero es un elemento de  $\Sigma_i$ , y los siguientes son conjuntos no vacíos con elementos del tipo  $envía(\partial, j, i)$  y  $recibe(\partial, j, i)$ . En el punto  $(r, 0)$  del sistema el historial de  $i$  es  $\{s\}$ , donde  $s \in \Sigma_i$ . Si durante el tiempo  $m$  de la corrida  $r$ , el agente  $i$  envía  $\partial$  al agente  $j$ , entonces, su historia en el punto  $(r, m)$  es el resultado de añadir  $envía(\partial, j, i)$  a su historia de  $(r, m-1)$ .

Las acciones se convierten en eventos registrados en historiales. Mediante los eventos se define el estado local de un agente. Por ejemplo,  $r_i(m)$  es la secuencia de todos los eventos con relación a  $i$  que han sido registrados hasta  $m$ .

La conexión entre estados locales y acciones se sigue de manera intuitiva. Un protocolo  $P_i$  es una función que determina, para cada  $r_i(m)$ , las acciones que  $i$  tiene disponibles en ese estado,  $P_i: r_i(m) \rightarrow A_i$ . Para reflejar la interacción entre acciones individuales se definen acciones conjuntas, que resultan de la unión de todas las acciones ejecutadas en un estado global del sistema. Las acciones conjuntas inducen a transiciones de estados globales a otros,  $\tau: r(m) \rightarrow r(m')$ . Por ejemplo,  $\tau(a_1, \dots, a_n)$   $(r_1(m), \dots, r_n(m)) = (r_1(m+1), \dots, r_n(m+1))$ , donde  $a_1, \dots, a_n \in A$ . Las transiciones locales se definen de manera similar.

Nótese que cuando  $recibe(\partial, j, i)$  es el caso, y en consecuencia el agente  $i$  transita a un estado  $r_i(m)$ , entonces se cumple  $K_i\partial$  en  $r_i(m)$ . Esto es así porque de acuerdo a la condición semántica de conocimiento en SI (sección 2.3), desde el punto de vista de  $i$  el sistema se encuentra en un estado global donde solo puede encontrarse cuando  $\partial$  es verdadero. Y  $\partial$  es verdadero para  $i$  desde  $r_i(m)$ , porque a todos los estados locales epistémicamente posibles a los que tiene acceso se cumple  $recibe(\partial, j, i)$ :



## 2.5 ¿Qué otras propiedades son necesarias para el análisis epistémico del SDM?

Aunque se dará mayor prioridad a propiedades epistémicas, se adoptarán dos propiedades temporales, dada su estrecha relación con las primeras:

- Memoria Perfecta: un agente con memoria perfecta nunca olvida la información que adquiere en los estados anteriores del sistema, pues su estado local registra el historial completo de su actividad durante la corrida actual.
- Sincronía y Asíncrona: de acuerdo con la primera, todos los agentes del sistema comparten un reloj global (p. 6, Sergot). Es decir, el tiempo sucede de la misma manera desde el punto de vista de todos los agentes. Un agente sincrónico satisface  $r_i(m) \neq r_i(m+1)$ ,  $\forall i$ . Caso contrario para agentes asíncronos.

La clase de agentes sincrónicos con memoria perfecta se caracteriza por satisfacer  $K_i XA \rightarrow XK_i A$ , donde  $X$  es el operador temporal *next* y  $A$  una fórmula arbitraria:  $XA$  es verdadero en tiempo  $n$  cuando  $A$  es verdadero en tiempo  $n+1$ . (p. 8, Sergot).

## 2.6 Listener-Teller. Un SDM con Protocolo de Comunicación

En *Reasoning About Uncertainty* Halpern expone algunos Sistemas Distribuidos con diferentes protocolos. Entre los objetivos de tal exposición está el demostrar que el comportamiento de un conjunto de agentes (i.e. procesos) depende del protocolo que ejecuten.

Uno de estos ejemplos es el Sistema Distribuido de Mensajería (SDM) *Listener-Teller* (pp. 210-213, Halpern; pp. 123-128, Halpern, et.al.). En él, dos agentes, Teller y Listener, ejecutan un protocolo de comunicación. El protocolo se compone de dos fases; en la primera Teller envía información a Listener durante cada punto del sistema, y durante la segunda, Teller consulta a Listener, y este le responde de acuerdo a la información que ha obtenido en la primera fase. El objetivo del protocolo, durante la primera fase, es almacenar información en la memoria de Listener, y el de la segunda fase, extraer consecuencias lógicas de este cuerpo de datos.

Esto es, de acuerdo con los tipos de inteligencia definidos (sección 1.2), y dado el protocolo establecido, Listener toma el papel de una agente con inteligencia reactiva, al ser la receptora de la información de otro agente (primera fase), y responder a las preguntas de Listener (segunda fase). En cambio, durante ambas fases, Teller es un agente con inteligencia pro-activa, pues toma la iniciativa por medio de sus actos para cumplir con los dos objetivos que estipula el protocolo.

En este trabajo se presenta un sistema como *Listener-Teller* para poner de manifiesto dos fenómenos epistémicos en una agente con inteligencia artificial como Listener, y que tienen como consecuencia la Omnisciencia Lógica (que se revisará a detalle como un problema en el 3er capítulo). El primero es la transformación de su conocimiento y creencias con la llegada de nueva información. El segundo es dar cuenta de su capacidad para deducir consecuencias lógicas. Se verá que esta capacidad en Listener es perfecta, a causa del empleo de las lógicas  $S_5$  y  $KD_{45}$  para formalizar su conocimiento y creencias.

En este SDM, Teller y Listener comparten un mismo ambiente. Teller tiene un conocimiento total de este ambiente, y además, tiene memoria perfecta. Esto es, conoce la descripción completa del ambiente en todo momento durante el protocolo. Por otro lado, Listener tiene nulo conocimiento del ambiente al inicio del protocolo, pero su ignorancia se reduce gradualmente gracias al canal de comunicación que establece con Teller.

Podría pensarse en una manera menos abstracta de esta situación, por ejemplo, suponiendo que Listener es una turista, y Teller un lugareño de la zona. Teller proporciona indicaciones del lugar a Listener, y posteriormente le hace preguntas para corroborar que las ha aprendido con éxito.

### **2.6.1 Primera fase**

La 1ª fase se divide en rondas, donde cada ronda sucede durante un estado global del sistema. En esta fase, Teller informa a Listener sobre el ambiente en cada ronda, pero esta información es muy limitada, pues solamente le informa sobre cuáles son los ambientes que no corresponden al verdadero para reducir su incertidumbre, pero nunca le dice cuál de hecho es el verdadero. A pesar de lo limitado de la información, se verá que Listener puede deducir una gran cantidad de consecuencias lógicas.

La descripción completa del ambiente, en el que ambos agentes habitan, se representa con una triada de variables binarias. Por ejemplo, (101) es una representación completa del ambiente. Con base en estas descripciones, Teller ejecuta acciones del tipo *envía*(~100, L, T), es decir, envía en cierta ronda un mensaje a Listener comunicándole que *el verdadero ambiente no es 100*.

Esta fase se presenta en dos versiones<sup>32</sup>. En ambas, Teller elige de manera determinista la descripción de un ambiente en cada ronda para hacerle saber a Listener que es falso. Primero ordena las descripciones del (000) al (111) y elige la descripción  $d_n$  en la ronda  $n$ , si  $d_n$  es una descripción falsa; pero si  $d_n$  es la descripción verdadera, entonces enviará el mensaje  $\sim d_{n+1}$  a Listener, para evitar revelar que  $d_n$  es la descripción verdadera del ambiente. En la primera versión Listener solo recuerda la información de su estado presente, denominada corrida  $R_1$ . En la segunda versión ( $R_2$ ), recuerda la información de los dos estados más recientes, y en una tercera ( $R_3$ ) tiene memoria perfecta, pero esta se reservará para la segunda fase.

Sea el SDM anteriormente definido el conjunto de corridas  $R_i$  (Tabla 1), donde  $i$  es la variable que representa cada una de estas tres corridas. Nótese que cuando el protocolo de un SD es determinista, hay una sola posible corrida por estado inicial, pues solo hay una acción que los agentes pueden ejecutar en cada estado, lo cual, induce a una sola secuencia de acciones. En *Listener-Teller* el protocolo es determinista, y tiene 3 posibles corridas, ya que son tres las posibles configuraciones iniciales, Listener con un bit de memoria, con dos bits, y con memoria perfecta:

<b>Tabla 1.</b> $R_i$	<i>Listener no tiene memoria</i>	<i>Listener recuerda la información de los dos estados más recientes</i>	<i>Listener tiene memoria perfecta</i>
<i>Protocolo determinista</i>	$R_1$	$R_2$	$R_3$

<sup>32</sup> Originalmente, el ejemplo *Listener-Teller* en Halpern (2003) se presenta en seis versiones. Las tres versiones que no se expondrán en este trabajo corresponden a casos donde el protocolo de Teller es probabilístico. La razón de no incluirlos se debe a que se prefiere aprovechar el espacio de este trabajo para exponer con detalle la manera en cómo se lleva a cabo un análisis epistémico de Sistemas Distribuidos, y las ideas en torno al conocimiento de agentes artificiales que hay de fondo. Serán suficientes estas tres versiones para mostrar que la Omnisciencia Lógica es un problema en *Listener-Teller*.

En las tres corridas se asumirá que Listener conoce el conjunto de todos las posibles descripciones del ambiente. Como consecuencia, la información que Teller le dará en cada ronda le permitirá reducir gradualmente la incertidumbre.

Como cada descripción corresponde a una terna de valores binarios, hay  $2^3$  posibles descripciones verdaderas:

- |          |          |          |          |
|----------|----------|----------|----------|
| 1. (000) | 3. (010) | 5. (110) | 7. (101) |
| 2. (100) | 4. (001) | 6. (011) | 8. (111) |

Teller comunica a Listener la descripción (1) durante el estado inicial  $r(s) \in \Sigma$  de cada corrida, y cuando la agente recibe esta información el sistema transita a la ronda 1, es decir, al estado  $r(1)$ , y posteriormente en esta ronda recibirá la descripción (2), etc. El protocolo prescribe que Teller envíe una descripción a Listener en cada ronda siguiendo este orden.

Sea  $L = \text{Listener}$ ,  $T = \text{Teller}$ , y  $A$  un conjunto de acciones:

- El conjunto de acciones disponibles para Teller ( $A_T$ ), en cada ronda de la primera fase, se compone de un elemento, que es,  $envía(\sim d_n, L, T)$ , donde  $d_n$  es cualquiera de las 7 descripciones del ambiente que no son las verdaderas.
- El conjunto de acciones disponibles para Listener es  $A_L$ , y contiene un solo elemento, que es,  $recibe(\sim d_n, T, L)$ . Para evitar redundancia, se omitirá en algunas casos la acción de envío de Teller, y se asumirá que en la acción receptiva de Listener hay un envío anterior de por medio. Es decir,  $recibe(\sim d_n, T, L)$  implica que la acción  $envía(\sim d_n, L, T)$  ya ha sido ejecutada anteriormente por Teller.

Si el historial de Teller es  $(d_7, \langle \sim d_1, \dots, \sim d_4 \rangle)$  en  $(r, m)$ , donde  $d_7$  es la descripción verdadera del ambiente, es decir,  $(IS, V) \models d_7$ . Y donde la información " $\sim d_1, \dots, \sim d_4$ "

corresponde a las descripciones que Teller ya ha informado a Listener de su falsedad, entonces los historiales de Listener son:

- $\langle \sim d_4 \rangle$ , si no tiene memoria, pues lo único que recuerda es la información de la última descripción del ambiente que Teller le ha proporcionado.
- $\langle \sim d_3, \sim d_4 \rangle$ , si tiene dos espacios de memoria. Es decir, recuerda la información de las dos últimas descripciones que Teller le ha dicho.
- $\langle \sim d_1, \dots, \sim d_4 \rangle$ , cuando tiene memoria perfecta. Recuerda, en todo momento, la información pasada y presente que Teller le ha comunicado durante el protocolo.

Listener transita de un estado local a otro siempre que recibe información de Teller. Y como los espacios de su memoria se definen con base en la información que puede retener durante la transición de estados, se dirá que, entre más información pueda almacenar Listener con el paso de las transiciones, mayor será su capacidad de memoria.

#### **2.6.1.1 $R_1$**

En  $R_1$  Listener solo recuerda la información del estado presente, por lo que cada estado es indistinguible para ella. Es decir, no puede saber que  $n-1$  es la ronda anterior, y  $n$  la ronda actual. Si bien es cierto que sabe, que recibir el mensaje ( $\sim 100$ ) pertenece a la primera ronda, ella no sabe si en realidad Teller le envía esta información en la ronda inicial porque (000) es la descripción verdadera, o si de hecho se trata de la primera ronda.

De acuerdo con las propiedades temporales establecidas (sección 2.6), Listener es una agente asíncrona para el sistema, pues no percibe transiciones, es decir, olvida

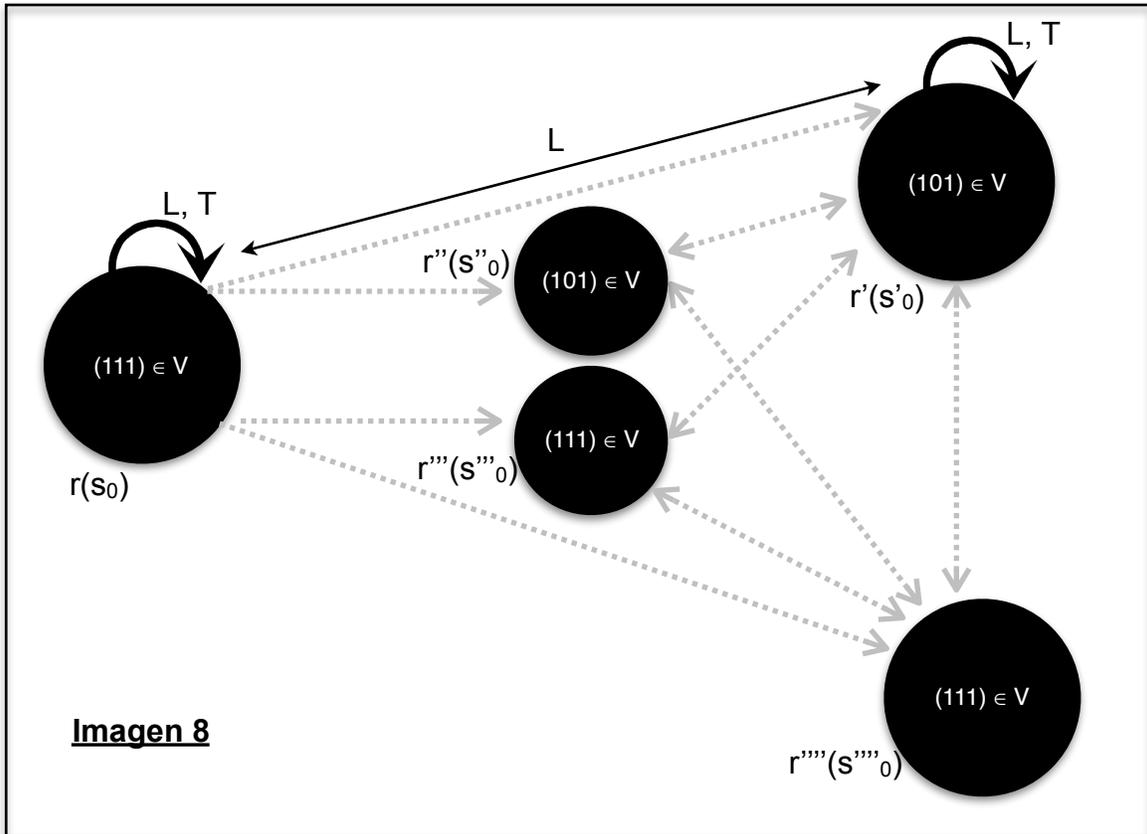
lo que sucede en el estado anterior cuando llega al siguiente. Véase que, a causa de la asíncrona en Listener, su conocimiento y creencias nunca aumentan:

- i) Supóngase que el siguiente historial representa el estado presente de Listener y Teller: (111,  $\langle \sim 000, \sim 100 \rangle, \langle \sim 100 \rangle$ ). De acuerdo con el historial, la descripción verdadera del ambiente es (111); la ronda presente es la segunda, dado que Teller ha informado a Listener que las dos primeras descripciones, (000) y (100), son falsas; y Listener recuerda la última pieza de información, que es ( $\sim 100$ ).
- ii) En términos epistémicos, la fórmula  $K_L(\sim 100)$  se satisface en  $r_L(2)$  esto es, Listener sabe en la segunda ronda que (100) es una descripción falsa del ambiente.
- iii) En la medida que Listener recibe información sobre cuáles son los ambientes que no corresponden al verdadero, sus creencias también se ven afectadas. Mientras que en su estado inicial ella cree que todas las descripciones que conoce son posiblemente verdaderas, no cree más en la posible verdad de (100) en (r, 2):  $(SI, r_L(2)) \models B_L(\diamond 000 \ \& \ \diamond 010 \ \& \ \diamond 001 \ \& \ \diamond 110 \ \& \ \diamond 011 \ \& \ \diamond 101 \ \& \ \diamond 111)$ .
- iv) Durante el mismo punto, (r, 2), Teller informa a Listener que (010) es otra de las descripciones falsas del ambiente; formalmente,  $(SI, r_L(2)) \models \text{recibe}(\sim 010, T, L)$ .
- v) La transición correspondiente se expresa como  $\tau(r_L(2))(\text{recibe}(\sim 010, T, L)) = r_L(3)$ . Listener transita de  $r_L(2)$  a  $r_L(3)$  cuando recibe la información ( $\sim 010$ ). Lo que sucede es que, al transitar al estado  $r(3)$ , Listener sabe ( $\sim 010$ ), y a causa de aprender esta información, olvida ( $\sim 100$ ), i.e.  $(SI, r_L(3)) \models K_L(\sim 010) \ \& \ \sim K_L(\sim 100)$ .
- vi) Por lo tanto, el historial resultante en el punto (r, 3) del sistema es (111,  $\langle \sim 000, \sim 100, \sim 010 \rangle, \langle \sim 010 \rangle$ ). Esta situación se repetirá en cada ronda, y por lo tanto, nunca aprenderá del ambiente, más allá de lo que le Teller le informa en un estado.

Sin embargo, a pesar de tener un solo espacio en su memoria y ser asíncrona, Listener puede, durante un solo estado, deducir todas las consecuencias lógicas de la información que posee en el presente, incluso si la información que Teller le proporciona es limitada. Algunas de las consecuencias lógicas que conoce (y cree) en su estado local  $r_L(3)$  son las siguientes:

- a)  $K_L K_L(\sim 010)$
- b)  $K_L[K_L(\sim 010) \rightarrow K_L(000 \vee 100 \vee 001 \vee 110 \vee 011 \vee 101 \vee 111)]$
- c)  $B_L(\sim \diamond 010 \rightarrow (\diamond 000 \ \& \ \diamond 100 \ \& \ \diamond 001 \ \& \ \diamond 110 \ \& \ \diamond 011 \ \& \ \diamond 101 \ \& \ \diamond 111))$
- d)  $B_L B_L \sim B_L(\diamond 010)$
- e) Listener, incluso, conoce todas las tautologías de lógica proposicional.

A manera de ejemplo, obsérvese que en el siguiente modelo de Kripke se cumplen fórmulas parecidas a las de los incisos (a), (c) y (e). Con el propósito de no hacer demasiado complejo el modelo, se asumirá que solo son dos las posibles descripciones del ambiente, (101) y (111). Las relaciones de acceso doxásticas de  $(B \rightarrow L)$  son grises, y todas pertenecen a Listener; las epistémicas  $(K \rightarrow)$  son negras y se indica con etiquetas a qué agentes corresponden. Los axiomas de  $KD_{45}$  se cumplen en virtud de las primeras, y los de  $S_5$ , en virtud de las segundas. Nótese que  $r_L(s_0)$  representa el estado inicial de Listener, donde  $r(s_0) \in \Sigma_L$ :



Este modelo representa el estado global durante la ronda inicial de la corrida  $R_1$ ,  $r(s_0)$ . Listener no sabe que la descripción verdadera es (111), pues tiene acceso epistémico a un estado  $r'(s'_0)$  donde la verdadera descripción es (101). Estos estados son indistinguibles para Listener porque en ambos se satisface su incertidumbre sobre la verdadera descripción, que se expresa con la fórmula  $\sim K_L(111) \ \& \ \sim K_L \sim(111)$ . Sin embargo, como es de esperarse, Teller conoce la descripción del ambiente, no solo en  $r(s_0)$  sino también en  $r'(s'_0)$ , pues en estos estados sus relaciones de acceso se limitan a las reflexivas.

Listener también mantiene creencias. De acuerdo con sus relaciones doxásticas, la agente cree que posiblemente (111) es el ambiente verdadero,  $B_L(\diamond 111)$ , así como

que (101) es posiblemente el ambiente verdadero  $B_L(\diamond 101)$ . A continuación se presentan algunas pruebas formales sobre las fórmulas que Listener conoce y cree.

\*a) Prueba.  $(SI, r_L(s)) \models K_L K_L(\sim 011)$

- i) Dado que las relaciones  $(r_L(s) \kappa \rightarrow_L r_L(s))$  y  $(r_L(s) \kappa \rightarrow_L r'_L(s'))$  se cumplen en este modelo, la fórmula  $K_L(\sim 011)$  debe satisfacerse en los estados  $r_L(s)$  y  $r'_L(s')$ .
- ii) Esto es el caso, pues a todos los estados que Listener tiene acceso (del tipo  $\kappa \rightarrow$ ) desde  $r_L(s)$  y  $r'_L(s')$  se satisface la ausencia de (011), esto es,  $(\sim 011)$ .  
Formalmente, se satisfacen las fórmulas  $(SI, r_L(s)) \models K_L(\sim 011)$  y  $(SI, r'_L(s')) \models K_L(\sim 011)$ . Ergo,  $(SI, r_L(s)) \models K_L K_L(\sim 011)$ .
- iii) Una manera más directa de llegar a esta conclusión es observando que si  $K_L(\sim 011)$  se cumple en  $r_L(s)$ , entonces  $K_L K_L(\sim 011)$  también se cumple por axioma (4).

\*c) Prueba.  $(SI, r_L(s)) \models B_L(\diamond 101 \ \& \ \diamond 111)$

- i) En el modelo se cumplen las relaciones  $(r_L(s) B \rightarrow_L r'_L(s'))$ ,  $(r_L(s) B \rightarrow_L r''_L(s''))$ ,  $(r_L(s) B \rightarrow_L r'''_L(s'''))$  y  $(r_L(s) B \rightarrow_L r''''_L(s''''))$ , y cada uno de estos estados, accesibles desde  $r_L(s)$ , tiene relación con al menos un estado donde  $(\diamond 101)$  y  $(\diamond 111)$  se satisface.
- ii) Formalmente,  $(SI, r'_L(s')) \models (\diamond 101)$  se cumple porque hay una relación  $(r'_L(s') B \rightarrow_L r''_L(s''))$ , y (101) se satisface en  $r''_L(s'')$ . Además,  $(SI, r'_L(s')) \models (\diamond 111)$  se cumple porque hay una relación del tipo  $(r'_L(s') B \rightarrow_L r'''_L(s'''))$ , donde  $(SI, r'''_L(s''')) \models (111)$ .  
Algo similar ocurre en el resto de los estados. Ergo,  $(SI, r_L(s)) \models B_L(\diamond 101 \ \& \ \diamond 111)$ .

\*e) Prueba.  $(SI, r_L(s)) \models K_L \Phi \rightarrow K_L(\Phi \vee \Psi)$

- i)  $(\Phi \rightarrow (\Phi \vee \Psi))$  es una tautología proposicional, y Listener la conoce. Si  $\Phi$  se satisficiera en  $r_L(s)$  y  $r'_L(s')$ , también sería satisfecha la fórmula  $(\Phi \vee \Psi)$  en estos

estados, pues sería suficiente que  $\Phi$  figurara en ambos para que  $(\Phi \vee \Psi)$  fuera el caso. Ergo,  $(S_I, r_L(s)) \models K_L\Phi \rightarrow K_L(\Phi \vee \Psi)$ .

Otras tipo de fórmulas que se satisfacen en el estado  $r(s_0)$  son las de múltiples agentes, por ejemplo, Listener está consciente de su propia ignorancia sobre el ambiente verdadero,  $K_L(\sim K_L(111) \ \& \ \sim K_L\sim(111))$ , pero sabe que Teller sí lo conoce,  $K_L(K_T(111) \vee K_T(101))$ .

Nótese que entre mayor sea el número de fórmulas que Listener conozca (y crea) en su estado actual, mayor será la cantidad de consecuencias lógicas que la agente sabrá, de manera automática, y sin importar lo complejas que puedan resultar y lo difícil que resultaría manipular la información que representan.

Si por medio de  $S_5$  y  $KD_{45}$  se le adscribe a Listener conocimiento y creencias de fórmulas de las que no se tiene una explicación sobre cómo las obtuvo *¿por qué razón se utilizan para el análisis epistémico de SDs?* Adscribirle tanta información parece contraintuitivo, al considerar que su memoria es muy limitada. La razón obedece a la noción de conocimiento y creencia que se utiliza en el análisis epistémico de SDs:

**Imagen 9**

<b>r(3)</b>	<b>r<sub>L</sub>(3)</b>	
<i>envía</i> (~010, L, T) <i>recibe</i> (~010, T, L)	◇ (000)	(~010)
	◇ (100)	
	◇ (001)	
	◇ (110)	
	◇ (011)	
	◇ (101)	
	◇ (111)	

La Imagen 9 representa una captura del sistema durante la tercera ronda del protocolo. Que el sistema se encuentre en el estado global  $r(3)$ , significa que en este se

satisfacen la acciones *envía*(~010, L, T) y *recibe*(~010, T, L). Por ende, desde el punto de vista Listener, la información (~010) se satisface en su estado local; en otros términos,  $(SI, rL(3)) \models K_L(\sim 010)$ . Como se ha asumido que Listener conoce el protocolo, fórmulas del tipo  $B_L \diamond (101)$  también se satisfacen, pues cree posibles todas las descripciones que, desde su punto de vista, Teller aún no le informa que son falsas.

Son dos las características del análisis epistémico en *Listener-Teller* que se ponen de manifiesto con en esta imagen. Una es que las nociones de conocimiento y creencia que se les atribuye a los agentes es externa (Halpern, et.al. pp. 8-10). Esto no es atípico, pues normalmente un análisis de este tipo se lleva a cabo de tal manera. La segunda característica, por el contrario, es inquietante, ya que estas nociones en realidad no dicen nada sobre cómo el agente obtiene la información. Por ejemplo, si el sistema se encuentra en el estado  $r(m)$  donde  $\Phi$  es verdadero, al agente se le atribuye el conocimiento de  $\Phi$  aunque no posea una representación de esta fórmula, y sin que sus capacidades deductivas sean tomadas en cuenta (pp. 258-260, Stalnaker).

Lo que sucede es que los operadores  $K$  y  $B$ , en el contexto de SDs, toman una significado particular, pues con estos no se tiene el objetivo de expresar el significado intuitivo de las palabras *conocimiento* ni *creencia*. Por ejemplo, cuando se dice que *i conoce*  $\Phi$  no significa que el agente ha computado la información requerida para deducir  $\Phi$ , sino que se trata de *conocimiento en potencia*.

En Sistemas Distribuidos, fórmulas como  $K_i \Phi$  “*significan que el agente  $i$  tiene información suficiente en sus estados locales para determinar que  $\Phi$  se mantiene globalmente en el sistema... [en sentido estricto] no significa que ha ejecutado el*

razonamiento requerido para determinar que  $\Phi$  es el caso” (p. Sergot, p. 3, trad.)<sup>33</sup>. De acuerdo con esta idea, las tautologías y consecuencias lógicas atribuidas a Listener son fórmulas que podría conocer en principio, es decir, su estado local tiene información suficiente para derivarlas<sup>34</sup>.

Véase ahora las consecuencias lógicas que Listener puede deducir con dos espacios de memoria.

### 2.6.1.2 $R_2$

En esta corrida Listener satisface sincronía, por lo que puede detectar transiciones de unos estados a otros. Para todo estado local  $r_L(m)$  y  $r_L(m+1)$ , ella identifica que conoce más en  $m+1$  que en  $m$ , pues recuerda las dos últimas cosas que Teller le ha dicho. Aún más, Listener sabe que  $n$  es la ronda actual, pues conoce la manera en cómo se lleva a cabo el protocolo, además de las ocho posibles descripciones del ambiente.

Si la acción *recibe*( $\sim 000$ , T, L) se satisface en el estado  $r(s)$ , entonces Listener transita al estado local  $r_L(1)$  donde se cumple  $K_L(\sim 000)$ . Si posteriormente recibe la información ( $\sim 100$ ), Listener transitará a  $r_L(2)$  y ella lo sabrá. Esto se debe a que reconoce la acción de haber recibido la falsa descripción (100) como perteneciente a la segunda ronda, pues sabe que en la ronda anterior recibió el dato ( $\sim 000$ ).

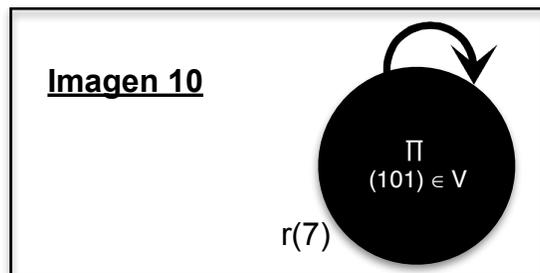
Véase ahora que es suficiente que Listener sea una agente sincrónica para conocer la verdadera descripción del ambiente durante un solo estado, sin importar que su memoria continúe siendo muy limitada:

---

<sup>33</sup> “ $K_i\Phi$  means that  $i$  has enough information in its own local states to determine that  $\Phi$  holds globally in the system as a whole. It does not mean that  $i$  has actually performed whatever reasoning is required to determine that  $\Phi$  holds.” (cita original).

<sup>34</sup> Durante el siguiente capítulo se defenderá la idea de que esta visión del conocimiento y la creencia deriva en un fenómeno denominado Omnisciencia Lógica, el cual resulta problemático, en diversos sentidos, para el sistema *Listener-Teller*.

- i) Supóngase que las siguientes condiciones se satisfacen:  $(101) \in V$ , y la ronda actual es la sexta. Esta es una situación similar a la de la Imagen 8, pero con una descripción verdadera y ronda presente distintas.
- ii) De acuerdo con el protocolo, Teller no puede revelar directamente a Listener que  $(101)$  es la verdadera descripción, por lo cual ejecuta la acción  $envía(\sim 111, L, T)$
- iii) durante el estado  $r(6)$ , y en consecuencia el sistema transita a  $r(7)$ , (Imagen 10):



- iv) Listener recibe esta información, y por lo tanto,  $recibe(\sim 111, T, L)$  (acción representada en el modelo por la fórmula primitiva  $\square$ ) se satisface en  $r(7)$ .
- v) Ya que Listener conoce el protocolo y sabía que se encontraba en la sexta ronda, y como también sabe que la ronda donde la acción  $envía(\sim 111, L, T)$  debería ejecutarse en la séptima, entonces las fórmulas  $K_L(101)$  y  $K(\sim 111)$  se satisfacen automáticamente en  $r_L(7)$ . Con la recepción del mensaje, Listener ya no concibe estados con acceso epistémico donde  $(111)$  sea la descripción verdadera, o estados con acceso doxástico donde se cumpla la posibilidad de la verdad sobre  $(111)$ .
- vi) El único estado al que Listener tiene ahora acceso es a  $r(7)$  por reflexividad, y dado que, evidentemente, este estado es para la agente indistinguible consigo mismo, entonces no hay más incertidumbre para ella.

Una vez más, obsérvese que el operador  $K$  se utiliza para expresar que el estado  $r(7)$  contiene los datos suficientes para que Listener derive la información  $(101)$ .

### 2.6.2 Segunda fase, $R_3$

En  $R_3$  Listener satisface memoria perfecta, y por lo tanto, se trata de una *Base de Datos*. Así pues, posee la suficiente capacidad de almacenamiento para deducir cuerpos de información aún más grandes que en  $R_1$  y  $R_2$  en cada estado.

Dado que Listener (ahora BD) es sincrónica y además tiene memoria perfecta, en esta corrida satisface, durante todos sus estados locales, la fórmula  $K_{BD}XA \rightarrow XK_{BD}A$ . Véase la demostración por *Modus Tollens*:

- i) Sea  $X$  el operador temporal *next*,  $A$  un enunciado que dice: *conozco una falsa descripción más, que es  $\sim d_{m+1}$* . Y supóngase que Listener sabe que la presente ronda es  $m$ , y además,  $m+1 = X$ .
- ii) Asúmase que  $(SI, r_{BD}(m+1)) \models K_{BD}A$  no se cumple. Esto es, Teller no informó a BD sobre la falsedad de otra descripción durante la ronda  $m+1$ , por lo que  $\exists(r, m+1)$  que es el punto del sistema que le precede a  $(r, m)$ , donde  $(SI, r_{BD}(m+1)) \not\models A$ .
- iii) Entonces  $(SI, r_{BD}(m)) \not\models K_{BD}XA$ .

En un SDM el rol de BD consiste en ser consultada por otro(s) agente(s) sobre hechos del ambiente, y responder de acuerdo con lo que conoce<sup>35</sup>. Para responder, BD deriva consecuencias lógicas de la información que posee. Sea  $\Phi$  una pregunta en  $(r, m)$  para BD:

- BD responde *Sí* cuando  $(SI, r_{BD}(m)) \models K_{BD} \Phi$
- BD responde *No* cuando  $(SI, r_{BD}(m)) \models K_{BD} \sim \Phi$
- BD responde *No sé* si ninguna de las de arriba es el caso

---

<sup>35</sup> Otro uso interesante para  $KD_{45}$  sería el caso donde Teller mantuviera un nivel de incertidumbre sobre el ambiente actual. Este tendría acceso a los estados que, de acuerdo con sus creencias, posiblemente contienen la información para saber cuál es este ambiente. Pero para evitar comunicar conocimiento erróneo a BD, con creencias falsas, tendría que imponerse una regla del tipo  $K_T \partial \rightarrow envía(\partial, BD, T)$ .

Para derivar consecuencias lógicas proposicionales, BD computa fórmulas sintácticamente equivalentes a las que posee; o bien puede derivar tautologías. Inclusive, no es posible que en semántica de Kripke existan tautologías o consecuencias lógicas no cognoscibles por BD, pues esta posibilidad deriva en contradicción. Esto se demuestra en la siguiente prueba.

Sea  $\mathbf{M}$  una clase de todos los modelos de Kripke cuyas relaciones satisfacen la propiedad de equivalencia, y sea  $SI^*$  un sistema en el que se cumplen las siguientes condiciones:

- Los valores de verdad de las fórmulas que describen el ambiente son estables, por lo que no cambian con el tiempo.
- Todo lo que sabe BD son hechos del ambiente, los cuales se expresan con fórmulas de lógica proposicional (i.e. no epistémicas) del tipo  $\phi_1, \dots, \phi_k$ .
- No hay conocimiento a priori por parte de BD.
- Finalmente, el historial  $(\phi_1, \dots, \phi_k)$  se cumple en  $r_{BD}(m)$ .

Si  $\kappa$  equivale al conjunto de fórmulas  $\phi_1 \& \dots \& \phi_k$ , y  $\Psi$  es fórmula de lógica proposicional, entonces las siguientes afirmaciones son equivalentes:

**(a)**  $(SI^*, r_{BD}(m)) \models K_{BD}\Psi$ ; **(b)**  $\kappa \rightarrow \Psi$  es TAUT; **(c)**  $\mathbf{M} \models K_{BD}\kappa \rightarrow K_{BD}\Psi$

Demostración:

**[(b)  $\rightarrow$  (c)]:** Si  $\kappa \rightarrow \Psi$  es TAUT entonces  $\mathbf{M} \models K_{BD}\kappa \rightarrow K_{BD}\Psi$ , pues  $(M, s) \models TAUT \forall M \in \mathbf{M}, \forall s \in M$  y  $\forall TAUT$ , y por definición semántica del operador  $K$ , se sigue  $(M, t) \models K_{BD}(\kappa \rightarrow \Psi) \forall t \in M. (t \rightarrow_{BD} s)$ . Luego,  $(M, t) \models K_{BD}\kappa \rightarrow K_{BD}\Psi$  por axioma (K), y por lo tanto se sigue **(c)**  $\mathbf{M} \models K_{BD}\kappa \rightarrow K_{BD}\Psi$ . Ergo, **[(b)  $\rightarrow$  (c)]**.

**[(c) → (b)]:** Supóngase que **(b)** no es el caso, formalmente,  $\kappa \rightarrow \Psi$  no es TAUT. Esto significa que  $\exists M \in \mathbf{M}$  tal que  $(M) \not\models K_{BD}\kappa \rightarrow K_{BD}\Psi$ . Por ejemplo, supóngase un modelo  $M$  de un estado  $s$  donde  $(s \rightarrow_{BD} s)$  y además  $(M, s) \models (\kappa \ \& \ \sim\Psi)$ . Como consecuencia,  $\mathbf{M} \not\models K_{BD}\kappa \rightarrow K_{BD}\Psi$ . Dado que **[(b) → (c)]** y **[(c) → (b)]** son el caso, **[(c) = (b)]** también lo es.

**[(b) → (a)]:** Supóngase que **(b)** es el caso, y supóngase también un  $(r', m')$  en  $SI^*$  donde, como en toda estructura  $SI$ , incluyendo  $SI^*$ , se cumple una relación de equivalencia entre sus puntos, en este caso particular:  $(r, m) =_{BD} (r', m')$ . Como es un hecho que  $(SI^*, r_{BD}(m)) \models \kappa$ , entonces  $(SI^*, r'_{BD}(m')) \models \kappa$  también lo es, al ser dos estados indistinguibles desde el punto de vista de  $BD$ . Dado que ya se ha demostrado **[(b) → (c)]** y se ha asumido que **(b)** es el caso, entonces **(c)** también lo es:  $\mathbf{M} \models K_{BD}\kappa \rightarrow K_{BD}\Psi$ .

Ahora bien, recuérdese que al ser  $(r, m) =_{BD} (r', m')$  una relación de equivalencia entre dos puntos de  $SI^*$ , entonces esta estructura  $SI^*$  puede interpretarse como un modelo de  $\mathbf{M}$ . Como  $(M, s) \models K_{BD}\kappa \rightarrow K_{BD}\Psi \ \forall M \in \mathbf{M}$  y  $\forall s \in M$ ,  $(SI^*, r_{BD}(m)) \models K_{BD}\kappa \rightarrow K_{BD}\Psi$  es el caso, y ya que evidentemente  $(SI^*, r_{BD}(m)) \models K_{BD}\kappa$  se cumple, dado que  $\kappa$  se cumple en  $r_{BD}(m)$  y  $r'_{BD}(m')$ , se sigue  $(SI^*, r_{BD}(m)) \models K_{BD}\Psi$ .

**[(a) → (b)]:** Supóngase que **(b)** no es el caso, entonces debe haber un  $(r', m')$  de  $SI^*$  donde  $(\kappa \ \& \ \sim\Psi)$  se cumpla. Por definición de las relaciones entre puntos de  $SI^*$ , es el caso que  $(r, m) =_{BD} (r', m')$ . Finalmente, ya que se ha asumido  $(SI^*, r'_{BD}(m')) \models \sim\Psi$ , y dado que hay una relación de  $r_{BD}(m)$  a  $r'_{BD}(m')$ , entonces  $(SI^*, r_{BD}(m)) \models \sim K_{BD}\Psi$ , ergo **(a)** no es el caso. Dado que **[(b) → (a)]** y **[(a) → (b)]** se cumplen, **[(a) = (b)]** también. Y por transitividad **[(a) = (c)]** también.

Fin de la demostración.

Que **(a)**, **(b)** y **(c)** sean equivalentes pone en evidencia dos hechos. Por un lado, agentes como BD, en la semántica de Kripke de  $S_5$ , conocen todas las fórmulas tautológicas proposicionales, y además, saben todas las consecuencias lógicas proposicionales de su conocimiento. Por otro lado, todas las fórmulas proposicionales que a BD se les pueden adscribir como conocimiento son de dos tipos, tautologías o consecuencias lógicas.

*¿Qué sucede cuando las fórmulas sobre las que a BD se le pide inferir su valor de verdad no son de lógica proposicional sino epistémicas y doxásticas? Para dar a conocer el valor de verdad de este tipo de fórmulas, BD debe hacer uso de los axiomas de  $S_5$  y  $KD_{45}$  para obtener como consecuencia lógica su verdad o falsedad. Como todos los agentes de estas lógicas conocen estos axiomas, entonces, no es posible que no conozcan (o crean) todas las consecuencias de la información que poseen.*

Pero para que esto sea posible, BD no puede responder como si se tratara de fórmulas de lógica proposicional. Por ejemplo, supóngase que BD interpreta la fórmula  $(p \rightarrow K_{BD}p)$  como una fórmula proposicional  $\Phi$ . En el caso que  $\Phi$  fuera falsa, su respuesta sería “No”, es decir “ $K_{BD}\sim\Phi$ ”. Pero ya que  $\Phi = (p \rightarrow K_{BD}p)$ ,  $K_{BD}\sim\Phi$  equivaldría a  $K_{BD}(p \ \& \ \sim K_{BD}p)$ , la cual es una fórmula inconsistente con  $S_5$ .

Para evitar respuestas como la del ejemplo anterior, BD debe usar sus habilidades introspectivas, modeladas con los axiomas (4) y (5), para responder si  $K_{BD}p$  es una consecuencia lógica de la información que posee, o si bien  $\sim K_{BD}p$  lo es. Modelar de esta manera las respuestas de Listener para fórmulas epistémicas, permite ver con

claridad que, como se observa en este caso, la agente siempre sabe lo que conoce así como lo que ignora.

Supóngase que el sistema se encontrara en un estado donde la expresión  $K_{BD}p$  se satisficiera. Entonces BD podría responder *Sí* a preguntas como  $\neg K_{BD}K_{BD}p$ ? Para hacerlo, revisa que en su estado presente se satisfaga  $p$ ; si  $p$  es el caso entonces  $K_{BD}p$  también se satisface, finalmente, BD obtendría  $K_{BD}K_{BD}p$  automáticamente, por derivación de una consecuencia lógica del axioma (4).

En este capítulo se introdujo el resto de la maquinaria formal para estudiar *Listener-Teller* bajo la perspectiva de las lógicas epistémica y doxástica, lenguajes para modelar en términos axiomáticos el conocimiento y las creencias de agentes, respectivamente. Al analizar *Listener-Teller* se mostró que, en las tres corridas de este sistema, Listener conoce todas las consecuencias lógicas de su información, además de todas las tautologías proposicionales, epistémicas y doxásticas, sin una justificación aparente, y sin importar la capacidad de almacenamiento en su memoria.

Que Listener conozca todas las consecuencias lógicas de su información actual, además de todas las tautologías, significa que se trata de una agente con la propiedad de *Omnisciencia Lógica*. La naturaleza de esta propiedad, así como sus motivaciones filosóficas, se revisarán en el siguiente capítulo. Además, se explicaran las razones formales por las que los agentes de lógica epistémica y doxástica son omniscientes, y las razones específicas por las cuales Listener lo es. Finalmente se revisarán dos desventajas que surgen al analizar *Listener-Teller* con agentes omniscientes.

### **3. Omnisciencia Lógica. Un problema de representación epistémica y sus consecuencias en Listener**

La Omnisciencia Lógica es una propiedad epistémica que satisfacen los agentes cuyo conocimiento y creencias se modelan con los axiomas<sup>36</sup> de las lógicas  $S_5$  y  $KD_{45}$ . Un agente omnisciente sabe todas las consecuencias lógicas de lo que conoce y cree. A causa de esta propiedad, no es posible distinguir las consecuencias lógicas que los agentes pueden derivar mediante su capacidad deductiva, de las que no pueden.

En este capítulo, después de presentar de manera general la propiedad de Omnisciencia Lógica, se explicará por qué, en particular, Listener es lógicamente omnisciente. Finalmente, se defenderá que la omnisciencia en la agente tiene consecuencias indeseables para el análisis epistémico.

#### **3.1 ¿Qué es la Omnisciencia Lógica?**

Para explicar qué es la Omnisciencia Lógica, Rohit Parikh (2007, Parikh2) parte de *El Menón* de Platón (p. 2). En este diálogo, Sócrates plantea un problema geométrico al esclavo de Menón. El esclavo logra inferir la respuesta después de que Sócrates le hace preguntas específicas sobre el mismo. La moraleja de Sócrates, dice Parikh, es que en realidad no aprendemos nada de nuestras inferencias (p. 1). Sócrates cree que el esclavo ya conocía la respuesta al problema, pues la dedujo a partir del conocimiento que ya tenía anteriormente a las preguntas que le hizo.

*El Menón* es conocido como uno de los primeros textos donde se aborda el tema de la Omnisciencia Lógica. La Omnisciencia Lógica es una capacidad epistémica. Se

---

<sup>36</sup> (G) figura en este conjunto, pero no se trata de un axioma. Sin embargo, para hacer la redacción más fluida, se asumirá la mención de (G) cuando se hable de axiomas omniscientes en general.

es lógicamente omnisciente cuando se conocen todas las consecuencias lógicas que se encuentran implícitas en el conocimiento que se posee de forma explícita.

Pero el problema de Omnisciencia se origina a partir de otro: el del *argumento deductivo correcto*. Si en un argumento las premisas contienen la conclusión quizá no hay nada nuevo al deducirla, pues parece que la conclusión ya se conocía (p. 1, Parikh1, 1995). Pareciera también que una entidad omnisciente no necesitaría razonar para derivar nuevo conocimiento del que posee, puesto que ya sabe todo lo que puede derivar via inferencias deductivas.

Otra manera de explicar la Omnisciencia Lógica es la de Fred Dretske (2005). El autor la presenta como un problema en la epistemología, y evalúa casos de omnisciencia relativos a la evidencia sensorial, por ejemplo, donde el hecho de oler, sentir o percibir de manera sensible P implica que también se percibe sensorialmente Q. Este tipo de casos parece extraño en el caso de agentes omniscientes con inteligencia artificial, pero el problema que Dretske detecta a través de diversos ejemplos en su texto es, en esencia, el mismo que se presenta en este trabajo.

Dretske identifica *implicaciones ligeras e implicaciones pesadas* entre las consecuencias lógicas del conocimiento explícito (pp. 17 y 19-20). Supóngase que  $(\Phi \& \Psi)$  es el conocimiento actual de Jimmy. Hay dos posibilidades. El enunciado  $\Phi$  es el consecuente de una implicación ligera cuando la evidencia que se posee de  $(\Phi \& \Psi)$  tiene el alcance suficiente para ser evidencia de  $\Phi$ . O bien,  $\Phi$  es el consecuente de una implicación pesada cuando la evidencia de  $(\Phi \& \Psi)$  no es suficiente para justificar  $\Phi$ .

Por ejemplo, bien Jimmy puede saber que las galletas son objetos que existen independientemente de su propia existencia, pero esto no significa que percibir

sensorialmente que hay galletas en el tarro sobre la mesa, implique que se tenga la percepción que cada uno de los objetos que existen en el mundo son independientes de su existencia (de Jimmy) (pp. 14-15). Percibir las galletas no implica que se perciba un mundo donde todos los objetos son independientes de la existencia de los sujetos.

En términos epistémicos, si bien, para que existan las galletas se requiere que existan en el mundo objetos independientes del sujeto, no se sigue que conocer la existencia de las galletas implique saber que el mundo esta formado por objetos independientes. Lo que sucede es que percibir o conocer la existencia de todos los objetos independientes es consecuente de una implicación pesada, pues la existencia de las galletas no es evidencia suficiente para justificar un hecho de tal magnitud.

Del ejemplo anterior el autor concluye dos cosas. En primer lugar, Jimmy conoce esta consecuencia solo cuando se acepta el compromiso de Omnisciencia Lógica. En segundo lugar, la Omnisciencia Lógica es un problema y debe evitarse a toda costa. Para sustentar lo segundo el autor introduce la noción epistémica de *transmisión de la evidencia* (p. 14), pues no esta de acuerdo en que la evidencia del conocimiento explícito se transmita a todos los consecuentes de consecuencias lógicas.

Se observa que a causa de la Omnisciencia Lógica emergen situaciones poco deseables para analizar agentes epistémicos; pero entonces *¿qué es lo que ha motivado su existencia?*

Stalnaker identifica cuatro posibles motivos por los que se introduce la Omnisciencia Lógica en teorías formales del conocimiento y creencia (pp. 242-245). La primera responde al objetivo de estudiar *el mecanismo natural de derivación de consecuencias lógicas*, sin la intervención de otros mecanismos que provocan una

*fricción cognitiva* que impide este proceso, por ejemplo, la ausencia de una memoria perfecta. El autor rechaza esta motivación, puesto que descansa en una idea idealizada y poco plausible de lo que es conocer y mantener creencias.

La segunda motivación es la simplificación de un fenómeno tan complejo como lo es el epistémico. Sin embargo, para Stalnaker la simplificación tiene un alto costo en la caracterización de los agentes, pues al concebirllos omniscientes deja de tener sentido su racionalidad (i.e. un agente omnisciente no necesita razonar, pues ya sabe todas las consecuencias de lo que conoce). La tercera motivación responde a perseguir un ideal normativo: *la Omnisciencia Lógica es el modelo al que todo agente debe aspirar*. Aunque por otro lado, dice el autor, esta no es razón suficiente para construir teorías formales bajo tal norma.

La cuarta motivación es para Stalnaker la razón que más se acerca a explicar la existencia de la Omnisciencia Lógica. La lógica epistémica es la mejor teoría formal del conocimiento que ha existido por un largo tiempo, y es cuestión de diseñar nuevas teorías para lidiar con el problema bajo una mejor estrategia (p. 245).

Nótese que la tercera motivación no es necesariamente equivocada, pues dependiendo de los objetivos a lograr en un modelo epistémico, bien podría requerirse modelar agentes con características omniscientes para cumplirlos. Sin embargo, como en este trabajo se defenderá que una perspectiva de agentes no omniscientes cuenta con ventajas frente a una de agentes omniscientes, en el caso particular de *Listener-Teller*, se adoptará la cuarta motivación que propone Stalnaker como la mejor explicación sobre la existencia de la Omnisciencia Lógica.

### 3.2 ¿Porqué los agentes de $S_5$ y $KD_{45}$ son omniscientes?

La Omnisciencia Lógica es una propiedad intrínseca a la semántica de Mundos Posibles para  $S_5$  y  $KD_{45}$ , dicho de otra forma, no es posible la ausencia de Omnisciencia Lógica si se pretende preservar la definición semántica estándar de los operadores K y B. En particular, son los axiomas de ambas lógicas por medio de los cuales se modela el comportamiento de los operadores, y es por esta razón que también se les suele llamar axiomas omniscientes.

En  $S_5$  el axioma más controversial es (K). De acuerdo con éste, si se conoce que *de los axiomas de la aritmética de Peano se derivan todos sus teoremas*, y si es un hecho se conocen estos axiomas, se sigue que se conocen todos sus teoremas (p. 1, Parikh<sup>2</sup>). (K) resulta problemático para agentes como Listener porque su aceptación lleva a asumir que posee una perfecta capacidad deductiva, y por tanto, que conoce todas las consecuencias lógicas de lo que sabe en el presente.

Otros axiomas controversiales respecto a la Omnisciencia Lógica son (4) y (5), además de la regla de Generalización (G) (p. 3, Halpern & Puccella, 2011). De acuerdo con (G), toda fórmula que es válida forma parte del conocimiento del agente, por ejemplo un agente conoce la tautología  $\Phi \rightarrow (\Phi \vee \Psi)$  en virtud de (G). Acerca de (4) y (5), no es realista pensar que se es consciente, al menos en todo momento, de lo que se conoce y lo que se ignora. Ello es más evidente en niveles profundos de consciencia lógica:  $KKKK\Phi$ . En el mismo sentido, axiomas derivados de estos como el de *sabiduría socrática* (sección 2.2) son también problemáticos.

Además de (K), (4) y (5), la situación es similar para axiomas omniscientes en  $KD_{45}$ . De los nueve axiomas doxásticos omniscientes que Meyer y Hoek listan (p. 72),

siendo todos ellos válidos en  $KD_{45}$ , tres de ellos son derivados de (G) (por lo que también se encuentran en  $S_5$ ):

- (AD1)  $\models \Phi$  implica  $\models B\Phi$  (*creencia de fórmulas válidas*)
- (AD2)  $\models \Phi \rightarrow \Psi$  implica  $\models B\Phi \rightarrow B\Psi$  (*cerradura bajo implicación válida*)
- (AD3)  $\models \Phi \leftrightarrow \Psi$  implica  $\models B\Phi \leftrightarrow B\Psi$  (*creencia de fórmulas equivalentes*)

En el caso del primero (AD1), si  $\Phi$  es una fórmula válida, entonces  $B\Phi$  también lo es. El caso es parecido para el segundo (AD2) y tercer (AD3) axioma.

Dos de estos axiomas son similares a (K), al expresar que las creencias están cerradas, en este caso bajo conjunción y disyunción:

- (AD4)  $(B\Phi \ \& \ B\Psi) \rightarrow B(\Phi \ \& \ \Psi)$  (*cerradura bajo conjunción*)
- (AD5)  $B\Phi \rightarrow B(\Phi \vee \Psi)$  (*debilitamiento de creencia*)

(AD4) implica que un agente conoce la conjunción de todas sus creencias. (AD5) implica que se consideran alternativas para toda creencia que se mantiene. Al igual que los anteriores, se obtienen los axiomas análogos a  $S_5$  al sustituir el operador B por K.

Por último, las siguientes expresiones llevan al compromiso de mantener un tipo particular de creencias, aquellas que son consistentes y correctas:

- (AD6)  $B\Phi \rightarrow \sim B\sim\Phi$  (*consistencia de creencias*), que es el axioma (D).
- (AD7)  $B(B\Phi \rightarrow \Phi)$  (*creencia de no tener creencias falsas*)
- (AD8)  $B \text{ true}$  (*creer en la verdad*)<sup>37</sup>

(AD6) se trata de (D). Finalmente, la validez de axiomas (AD7) y (AD8) hacen aún más evidente que el operador B (así como K) tiene el mismo comportamiento que el operador modal de necesidad ( $\square$ ). Una fórmula del tipo  $\square\Phi$  significa que  $\Phi$  se satisface

---

<sup>37</sup> Así mismo, estos tres axiomas (AD6, AD7 y AD8) forman parte de  $S_5$ , ya que el conocimiento también satisface estas características.

en todos los estados a los que se tiene acceso desde el estado del agente, siendo el mismo caso para la creencias.

El uso de estos axiomas, en sistemas como los SDM, pone de manifiesto que la consecuencias lógicas derivadas por un agente artificial deberían ser completas y correctas respecto a su información (p. 242, Russell & Norving; p. 49, KRR). Deberían ser correctas porque el agente no podría deducir algo diferente a las fórmulas que se encuentran implícitas en su conocimiento actual. Y deberían ser completas porque tendría que ser capaz de derivar todas las fórmulas implícitas. Pero esta clase de agentes son idealizaciones, pues ninguno posee en realidad tales capacidades<sup>38</sup>.

### **3.3 ¿Porqué es Listener una agente omnisciente?**

Hay dos razones por las que Listener (y BD) es lógicamente omnisciente. La primera es el uso (implícito) de axiomas epistémicos y doxásticos para modelar sus procesos inferenciales. La segunda es el uso de una noción de memoria que refleja una capacidad de almacenamiento de información idealizada.

#### ***3.3.1 Omnisciencia lógica con relación al uso de axiomas***

El problema de Omnisciencia Lógica en agentes de SDs, respecto al uso axiomas, se origina a partir de una concepción epistémica adoptada por muchos filósofos y científicos cognitivos, conocida como *el modelo de almacén* (p. 246, Stalnaker). Lo

---

<sup>38</sup> Es interesante mencionar la situación cuando se emplea lógica de primer orden como lenguaje epistémico (i.e. John McCarthy). En este caso, el compromiso de asumir la omnisciencia lógica sería aún más evidente, pues aunque bien la lógica de primer orden es *completa*, y por lo tanto es posible computar todas las consecuencias lógicas de un conjunto de fórmulas de predicados, esta es *indecidable*. La indecidibilidad en un lenguaje formal implica que no hay un método efectivo para derivar todas las consecuencias lógicas de un conjunto de fórmulas del mismo, y en términos epistémicos supondría que los agentes requerirían una inmensa cantidad de recursos temporales y espaciales para lograrlo.

esencial de este modelo es que el conocimiento y la creencia se determina de acuerdo a los enunciados almacenados en la memoria del agente.

Hay dos maneras en las que, a través de los axiomas de  $S_5$  y  $KD_{45}$ , este modelo condiciona lo que Listener conoce y cree (pp. 247-249). La primera responde a una manera imprecisa y abstracta sobre cómo se caracterizan los enunciados que expresan su conocimiento. En *Listener-Teller*, y en un análisis epistémico en general, los enunciados son composiciones de sub-formulas o literales atómicos con valores de verdad, falsos o verdaderos. Por lo tanto, un agente puede conocer, por ejemplo, diversas equivalencias semánticas entre enunciados cuando tienen el mismo valor, como es el caso de  $K_L(\Phi \ \& \ \Psi) \leftrightarrow K_L(\Psi \ \& \ \Phi)$ , una expresión válida en  $S_5$ .

La segunda manera con la que el modelo de almacén establece condiciones en los enunciados de Listener, es la imposición de propiedades que responden a una visión sobre cómo debe comportarse un conjunto de enunciados en su memoria, por ejemplo, que no mantenga creencias (y conocimiento) contradictorias, como sucede con los axiomas (D), y (AD8) (sección 3.2). A continuación se revisan algunos de estos axiomas con relación a la omnisciencia de Listener.

A lo largo de *Listener-Teller*, los axiomas anteriormente presentados determinan lo que Listener sabe en cada momento. Por ejemplo, su conocimiento de axiomas y tautologías le permiten acceder a una gran cantidad de información.

Un caso de conocimiento de axiomas por Listener surge con la fórmula  $B_L(\diamond 101 \ \& \ \diamond 111)$  (Imagen 8). Para llegar a esta información se asume que, en un mismo estado y de forma automática, Listener sabe, en virtud del axioma (AD8), que si  $[(\diamond 101) \ \& \ (\diamond 111)]$  es verdadero en este modelo, entonces  $[B(\diamond 101) \ \& \ B(\diamond 111)]$  también lo es.

Además, Listener conoce la siguiente instancia de (AD4),  $B(\diamond 101) \& B(\diamond 111) \rightarrow B_L(\diamond 101 \& \diamond 111)$ . Como  $B_L(\diamond 101) \& B_L(\diamond 111)$  es el caso, se sigue  $B_L(\diamond 101 \& \diamond 111)$ , es decir, sus creencias están cerradas bajo conjunción. También, el hecho que Listener mantenga en su cuerpo de creencias todas las fórmulas tautológicas, a causa de (AD1), implica además que puede demostrarse de una manera similar a la prueba \*e (p. 36 e Imagen 8) que  $B_L[\Phi \rightarrow (\Phi \vee \Psi)]$  es el caso en  $r(s)$ .

Otros casos de omnisciencia, con relación al uso de axiomas, es evidente atendiendo al tipo de creencias que Listener mantiene. Sus creencias son siempre consistentes en virtud de (D), y además, son siempre verdaderas, por lo cual también satisfacen (AD8)<sup>39</sup>.

La omnisciencia respecto al conocimiento sucede principalmente con (G) y (K), una vez más, el conocimiento de la tautología  $(\Phi \rightarrow (\Phi \vee \Psi))$  (prueba \*e) puede demostrarse fácilmente con ambos, así como la derivación de su consecuencia lógica correspondiente, a saber,  $(\Phi \vee \Psi)$ . Primero,  $K_L(\Phi \rightarrow (\Phi \vee \Psi))$  por (G). Y como en este caso conoce  $\Phi$ , entonces se sigue  $(\Phi \vee \Psi)$  por (K):  $K_L(\Phi \rightarrow (\Phi \vee \Psi)) \rightarrow (K_L\Phi \rightarrow K_L(\Phi \vee \Psi))$ .

Los axiomas (G) y (K) son válidos en cualquier marco y bajo cualquier evaluación. Estos modelan el conocimiento en Semántica de Kripke. Suprimirlos no es

---

<sup>39</sup> A manera de ejemplo, véase el modelo correspondiente a la Imagen 8 del segundo capítulo. En el modelo Listener cree posible en el estado  $s$  tanto 111 así como 010, es decir  $B_L(\diamond 111)$  y  $B_L(\diamond 101)$  se satisfacen en  $s$ . Ahora bien, tanto  $(\diamond 111)$  como  $(\diamond 101)$  se satisfacen en todos los estados del modelo, lo cual significa que ambas expresiones cumplen con la condición de verdad en el modelo; *algo es verdadero en un modelo cuando se satisface en todos los estados del mismo*. El caso es el mismo para el resto de las creencias de Listener, y para todos los modelos.

una opción, pues no solo  $S_5$  y  $KD_{45}$  colapsarían, sino todo sistema de lógica modal estándar<sup>40</sup>. Véase que el intento de eliminarlos deriva en contradicción:

- i) Supóngase que  $(G) \notin S_5$ , entonces entonces existe una fórmula válida  $\Phi$ , un modelo  $SI$  y un estado  $r(m) \in SI$ , tal que  $(SI, r(m)) \models \sim K\Phi$ . Por definición semántica del operador  $K$ ,  $\exists r'(m'). r(m) \vDash_{K \rightarrow} r'(m')$  donde  $(SI, r'(m')) \models (\Phi \ \& \ \sim \Psi)$ , lo cual significa que  $\Phi$  no es una fórmula válida. **Contradicción.**
- ii) Supóngase que  $(K) \notin S_5$ , entonces existe un  $SI$  y un estado  $r(m) \in SI$ , tal que  $(SI, r(m)) \models K(\Phi \rightarrow \Psi) \ \& \ \sim(K\Phi \rightarrow K\Psi)$ . Si  $(SI, r(m)) \models (K\Phi \ \& \ \sim K\Psi)$  entonces  $\exists r'(m'). r(m) \vDash_{K \rightarrow} r'(m')$  donde  $(SI, r'(m')) \models \sim \Psi$ . Pero para que se cumpla  $(SI, r(m)) \models K(\Phi \rightarrow \Psi)$  debe ser el caso que,  $\forall r'(m'). r(m) \vDash_{K \rightarrow} r'(m'), (SI, r'(m')) \models \sim(K\Phi \ \& \ \sim K\Psi)$ .

**Contradicción.**

Además, negar otros axiomas como (T) llevaría a eliminar la única característica que define al conocimiento frente a la creencia en este par de lógicas: *la verdad*. Al negar (4) y (5) no solo se perderían las capacidades introspectivas de los agentes, sino también la posibilidad de expresar conocimiento común como lo es en el caso del protocolo de Teller.

Al igual que (K) y (G), los axiomas omniscientes de  $KD_{45}$  son tan fundamentales en la semántica de Kripke, que su negación también derivaría en un colapso. Por ejemplo, eliminar (D) tendría como consecuencia la aceptación de creencias contradictorias, y por ende, el agente estaría autorizado en creer cualquier cosa a partir de ella.

---

<sup>40</sup> En la sección 4.1.2 se verán algunas propuestas para enfrentar la omnisciencia lógica en los sistemas modales del conocimiento y la creencia, en ellos es posible suprimir la validez de (K) y (G). Se verá también que hay consecuencias con las que debe lidiarse, las cuales no resultan tan convenientes.

### 3.3.2 Omnisciencia Lógica con relación a la noción de memoria

Hay evidencia para pensar que Listener es también omnisciente dado que su memoria responde a constreñimientos temporales del reloj del sistema, pero no al tiempo que le llevaría computar las consecuencias lógicas cuyo conocimiento se le adscriben:

- i) Sea  $(SI, r_L(m)) \models \text{recibe}(\partial, T, L)$  en  $R_1, R_2$  y  $R_3$ , donde  $\partial: (\sim 1XX \ \& \ \sim XX0)$
- ii)  $\tau(r_L(m)) (\text{recibe}(\partial, T, L)) = (SI, r_L(m+1)) \models B_L(\diamond 001 \ \& \ \diamond 011)$

Listener deriva  $B_L(\diamond 001 \ \& \ \diamond 011)$  en un mismo estado, sin importar su capacidad de almacenamiento, aunque datos como  $\sim 1XX$  podrían tratarse de grandes cuerpos de información que requirieran una gran memoria. Que ella pueda derivar  $B(\diamond 001 \ \& \ \diamond 011)$ , incluso con carencias de almacenamiento, se debe a que cada espacio de su memoria corresponde a un estado local. Mientras no haya transición al siguiente estado, que sucede cada vez que Teller le proporciona información, ella podrá continuar derivando inferencias aún si solo recuerda lo que Teller le dice en el presente<sup>41</sup>.

### 3.4 ¿Qué consecuencias tiene la Omnisciencia Lógica en Listener?

En el capítulo 2 (sección 2.6.1) se explicó que, para enfrentar la Omnisciencia Lógica, los defensores del uso de las lógicas modales del conocimiento y la creencia para el análisis epistémico de Sistemas Distribuidos (SDs) adscriben un significado particular a los operadores  $K$  y  $B$ .  $K_i\Phi$  significa que el estado local del agente  $i$  tiene la información suficiente para derivar  $\Phi$ , siendo un caso similar para la creencia. Mediante tal estrategia se ha intentado evitar dar una visión idealizada de las capacidades

---

<sup>41</sup> Durante el quinto capítulo se reflexionará sobre el problema de cómo modelar la memoria de Listener.

inferenciales de los agentes; sin embargo, hay una importante razón para pensar que el uso de esta noción es problemática en Sistemas Distribuidos.

A lo largo de este capítulo y el anterior se ha puesto en evidencia que una consecuencia directa de la Omnisciencia Lógica es que no es posible identificar las inferencias que están al alcance del agente, de las que no lo están. Esta consecuencia afecta directamente a aquellos agentes cuyas creencias y conocimiento se modelan bajo las nociones de *creencia* y *conocimiento potencial*. En el caso particular de *Listener-Teller*, que no sea posible identificar las inferencias que están al alcance de una agente como Listener tiene dos consecuencias problemáticas (i, ii):

- i) Hacer dependientes las acciones de la información requiere saber el momento en el que los agentes tienen acceso a la información que se les adscribe para utilizarla como una herramienta para tomar decisiones respecto a sus actos (pp. 260-261, Stalnaker). El problema, pues, es determinar la disponibilidad del conocimiento, lo cual no es posible cuando a los operadores K y B cobran un significado de conocimiento y creencia *en potencia*.

Supóngase que se ejecuta un protocolo de comunicación entre Listener y Teller. Con el objetivo que Listener llegue a la meta, el papel de Teller consiste en proporcionarle información para que ella pueda navegar en un mapa, donde hay espacios con y sin camino. En los primeros se satisface el dígito 1, y en los vacíos, 0.

Sea una descripción (d) una triada de valores binarios del tipo (011). Teller ejecuta acciones como *envía*(011, L, T) para informarle a Listener que (011) es la descripción de acuerdo con la cual debe desplazarse un espacio, desde la coordenada donde se encuentra en el presente (Imagen 11):

**Imagen 11**

	X	Y
A	META	0
B	1	1
C	0	Listener

Este es un caso donde el desplazamiento de Listener depende de su conocimiento,

- Si  $(SI, (C, Y)) \models K_L(d)$  entonces avanza un espacio donde  $\sim 0$ . Esto es, *cuando conozcas la descripción (d) del mapa estando en las coordenadas (C, Y), muévete un espacio evitando caer al vacío.*

De acuerdo con la imagen (arriba), Listener debe conocer la descripción (011) de manera que se garantice su desplazamiento a las coordenadas (B, X) o (B, Y). Pero si la expresión  $K_L(011)$  significa que, *dado que la acción recibe(011, T, L) se satisface en la coordenada (C, Y), el estado donde Listener se encuentra tiene la información suficiente para derivar (011)*, entonces no hay garantía que Listener haya ejecutado las inferencias necesarias para computar este dato.

Esta situación es problemática en diversos sentidos. Por ejemplo, supóngase que el protocolo obliga a que Listener se desplace un espacio durante cada ronda, entonces la agente corre el peligro de desplazarse al vacío en cualquier coordenada.

Un problema muy similar sucede cuando Listener es una base de datos y ésta es cuestionada por Teller. Ella debe identificar si las respuestas que necesita inferir están al alcance de sus capacidades deductivas (pp. 391-392, Halpern, et.al.). Una vez más, hay acciones que dependen de su conocimiento, pues así como en el ejemplo anterior, el empleo de una noción potencial de conocimiento no permite a la base de datos responder *Sí-No* a las preguntas de Teller.

ii) Bajo las nociones potenciales de conocimiento y creencia,  $K_L\Phi$  cobra el siguiente significado:  $\Phi$  es cognoscible para Listener. Fitting sugiere que, en consecuencia, la expresión  $\sim K_L\Phi$  toma un significado confuso, a saber,  *$\Phi$  no es cognoscible para Listener* (pp. 3-4, Fitting1). Supóngase que expresiones como  $\sim K\Phi$  cobraran tal significado *¿Qué causaría la adopción de esta interpretación?* Una posible consecuencia sería entender la no cognoscibilidad de  $\Phi$  como *la imposibilidad sobre el conocimiento de  $\Phi$* . Pero esta visión de imposibilidad epistémica no es clara, por lo cual, no está exenta de confusiones en caso de adoptarla.

En primer lugar *¿cómo entender la imposibilidad de conocer una fórmula?* Una posible interpretación es que es imposible conocer una fórmula cuando *no hay forma alguna por medio de la cual el agente pueda añadir esta fórmula a su cuerpo de conocimiento y creencias*. Pero no parece que esto sea lo que quiera decirse en situaciones del tipo  $\sim K\Phi$ . Véase que hay dos posibles escenarios con distintos resultados.

Supóngase un ambiente que contiene un solo agente. Es razonable pensar que su incapacidad de conocer fórmulas afectaría totalmente la posibilidad de que éste obtenga nueva información, pues su única manera de conocer, al no haber más agentes, sería la inferencia deductiva. Hasta ahora parece no haber problemas.

Pero ahora véase la misma situación para un ambiente de múltiples agentes. En este caso, existen medios alternativos para conocer aún si esta noción de imposibilidad epistémica fuera aceptada, pues cualquier agente podría comunicar ésta información sobre nuevas fórmulas *¿Qué razones habría para asegurar que es imposible el hecho que un agente comunica información a otro?* Parece que mantener esta interpretación

de imposibilidad epistémica resulta en una una noción de no cognoscibiidad demasiado restrictiva para acciones comunicativas de agentes, y por lo tanto poco convincente.

Este último escenario es muestra de lo que sucede en *Listener-Teller*. En un sentido formal, en el caso de adoptar esta noción de imposibilidad, la satisfacción de la fórmula  $(IS, r_L(m)) \sim K_L \Phi$  tendría como consecuencia la satisfacción de la fórmula  $\forall n. (IS, r_L(m+n)) \models \sim K_L \Phi$ . Pero esto significaría la imposibilidad de que Teller comunicara a Listener esta información en cualquier momento, imposibilidad que contradice el propósito de un protocolo de comunicación como el que se ejecuta en este Sistema Distribuido de Mensajería.

Alternativamente, con el objetivo de solventar estas restricciones para ambientes multi-agente, podría sugerirse interpretar la no cognoscibilidad como *imposibilidad de conocer por medio de inferencia deductiva*. Sin embargo, esto tampoco es correcto, pues cabe la posibilidad de que, si bien  $\sim K_L \Phi$  se satisface, Teller podría comunicar a Listener un conjunto de fórmulas, mediante las cuales, la agente podría en seguida derivar por inferencia deductiva la fórmula  $\Phi$ .

Finalmente, y como respuesta a este problema, podría sugerirse interpretar la no cognoscibilidad como la imposibilidad de conocer una fórmula por medio de inferencia dedutctiva, *en este momento particular*. Pero la adopción de esta última interpretación tendría como consecuencia abandonar la noción de conocimiento potencial del operador K, pues para mantener consistencia con la interpretación negativa del mismo,  $K\Phi$  tendría que entenderse como *el agente sabe  $\Phi$  en este momento particular*, interpretación que, como se ha visto anteriormente, no puede tener (véase el inciso i en esta misma sección).

Nótese que es muy razonable interpretar el operador K bajo la noción potencial de conocimiento, pues observando su comportamiento semántico es evidente que el término de *conocimiento* que representa no corresponde al significado cotidiano de lo que suele entenderse por *conocer*. El problema que aquí se pone de manifiesto es la falta de claridad que hay en la interpretación del operador K, lo cual es notorio en casos de ausencia de conocimiento ( $\sim K$ ).

Después de explicar la propiedad de Omnisciencia Lógica, y la razones por las que es satisfecha por agentes de  $S_5$  y  $KD_{45}$ , se revisó el caso particular de Listener como una agente omnisciente. Los problemas originados a causa de la omnisciencia de Listener sugieren la necesidad de un lenguaje formal para llevar a cabo un análisis epistémico de SDs, con el que sea posible determinar si la agente puede acceder a la información que requiere, o si por otro lado, no tiene la capacidad suficiente para hacerlo.

Con el objetivo de proponer un lenguaje con estas características, en el siguiente capítulo, se introduce una lógica que ofrece una perspectiva distinta de las capacidades inferenciales de agentes, *la lógica de justificaciones*. Toda adscripción de información que se le atribuye a los agentes de la lógica de justificaciones está antecedida por términos, los cuales representan evidencia que respalda tal información. Con ayuda de este mecanismo de evidencia es posible distinguir la información que los agentes mantienen explícitamente, y la que mantienen solo de forma implícita. Esta distinción supone un paso importante para lograr el objetivo de modelar agentes no omniscientes.

## **4. La Lógica de Justificaciones. Una propuesta alterna para enfrentar la Omnisciencia Lógica**

La mayoría de las propuestas que buscan resolver el problema de Omnisciencia Lógica no son satisfactorias, pues lo abordan solo parcialmente, o por el contrario, limitan en exceso la capacidad deductiva de los agentes. Dada la insatisfacción para tratar este problema se introduce la lógica de justificaciones, un lenguaje para modelar conocimiento y creencias justificadas, donde los agentes no son omniscientes, pero además preservan sus habilidades inferenciales. Al final de este capítulo se harán explícitas las ventajas de la lógica de justificaciones frente a las lógicas epistémica y doxástica.

### **4.1 ¿Qué tipo de propuestas existen para enfrentar la Omnisciencia Lógica?**

Hay tres tipos de posibles soluciones a la Omnisciencia Lógica, con las cuales se aborda el problema bajo distintas estrategias. Las primeras son *interpretaciones conceptuales alternativas de la noción de conocimiento* (a); mediante estas se tiene el objetivo de establecer un concepto de conocimiento con el cual logren evadirse algunos tipos de Omnisciencia Lógica. Las segundas son *condiciones alternas para la satisfacción del conocimiento* (b), las cuales se tratan de definiciones de conocimiento para la Semántica de Kripke bajo ciertas restricciones.

La tercera propuesta es *el enfoque de conocimiento algorítmico* (c). Este es un formalismo en el que se abandonan algunos elementos de la Semántica de Kripke, con el objetivo de establecer una noción distinta de conocimiento basada en algoritmos<sup>42</sup>.

#### **4.1.1 (a) Propuestas conceptuales**

La primera de las propuestas conceptuales consiste en reconocer la diferencia entre el conocimiento de enunciados y el de proposiciones<sup>43</sup> (pp. 23-24, Parikh1). Sean los enunciados *objetos sintácticos*, y las proposiciones, *el sentido de los enunciados*; el sentido se trata del contenido semántico, o en términos más formales, su valor de verdad. Sean  $\Phi$ ,  $\Psi$  enunciados, y además sea  $(\Phi \leftrightarrow \Psi)$  una expresión verdadera, así pues, las dos denotan la misma proposición, dado que tienen el mismo valor de verdad<sup>44</sup>.

Bajo tal distinción del conocimiento, entre conocimiento de proposiciones y conocimiento de enunciados, se ha defendido la idea de que se está justificado en conocer las equivalencias semánticas, y que por ende, no hay en estos casos una adscripción irracional de Omnisciencia Lógica de por medio<sup>45</sup>. La idea que hay de fondo es que los agentes reaccionan de manera similar ante el conocimiento de enunciados cuyo contenido es el mismo.

---

<sup>42</sup> Los agentes artificiales trabajan por medio de algoritmos para inferir su conocimiento. Un algoritmo es el conjunto de instrucciones con el que un agente es programado para efectuar tareas, por ejemplo la de inferir enunciados implícitos a partir de lo que sabe explícitamente.

<sup>43</sup> Solamente en esta sección se le otorga un significado particular a los términos “enunciado” y “proposición” respecto a este ejemplo. Al finalizar esta sección se continuarán empleando los términos “enunciado” y “proposición”, pero no bajo el significado específico que se le otorga en la propuesta de Parikh.

<sup>44</sup> Para que una equivalencia del tipo  $(\Phi \leftrightarrow \Psi)$  sea verdadera, el valor de verdad de cada fórmula debe ser el mismo. Tanto  $\Phi$  y  $\Psi$  deben ser falsas, o bien, verdaderas.

<sup>45</sup> Como se menciona en la sección 3.2,  $[ \models \Phi \leftrightarrow \Psi \text{ implica } \models K\Phi \leftrightarrow K\Psi ]$  es un axioma epistémico omnisciente derivado de la regla inferencial G. En virtud de esta regla, toda fórmula válida es conocida por los agentes, y en consecuencia, toda equivalencia semántica válida también lo es.

Para esta defensa se introduce el término de *b-conocimiento* (b de *behaviour*). Sea b-conocimiento el comportamiento de un agente ante la verdad o falsedad de un enunciado. Un termostato que activa el calor cuando la temperatura baja a 68°F también lo hará cuando baje a 20°C; esto se debe a que  $b-K(68^{\circ}\text{F}) \leftrightarrow b-K(20^{\circ}\text{C})$  (p. 24), suponiendo, entre otras posibles interpretaciones, que tanto la información “68°F” como “20°C” significan que *es verdadero que hace frío*.

Como 68°F y 20°C son enunciados que tienen el mismo significado, dado que poseen el mismo contenido semántico, y como el comportamiento reactivo del termostato es el mismo al conocer cualquiera de los dos contenidos, entonces parece razonable justificar la cerradura del b-conocimiento bajo equivalencia semántica. Pero en términos de conocimiento sintáctico, continúa Parikh, es posible que suceda el caso donde  $K\Phi$  pero  $\sim K\Psi$ : *cuando alguien dice que su hijo tiene dos años quizá no quiera decir que su edad es el único número  $n > 1$  para el que existen tres enteros  $x, y, z > 0$  tal que  $x^n + y^n = z^n$* . (p. 23).

Los defensores de ejemplos como este concluyen lo siguiente: es razonable que un agente no posea la capacidad de identificar la similitud entre construcciones sintácticas de la información, pues la intención de expresar, por ejemplo,  $\sim(\Phi \& \sim\Psi)$ , puede no ser la misma al expresar  $(\Phi \rightarrow \Psi)$ . Sin embargo, la similitud que se refleja en las reacciones de los agentes ante distintas construcciones sintácticas que poseen el mismo contenido semántico, es evidencia de su capacidad para saber que son contenidos equivalentes.

La segunda de este tipo de propuestas se trata de concebir una conceptualización distinta del conocimiento en agentes algorítmicos (pp. 26-27,

Parikh1). Para ello, se ha distinguido entre *conocimiento como hecho* y *conocimiento como procedimiento*. Un agente tiene conocimiento de hecho sobre lo que hay en su base de datos o lo que infiere a partir de ella. La segunda noción, de conocimiento como procedimiento, refiere a sus habilidades. Así pues, un agente no es lógicamente omnisciente si se adopta la segunda noción, es decir, si por *conocimiento* se entiende su habilidad para deducir consecuencias lógicas por medio de un algoritmo al que se le añade información de manera gradual.

Ninguna de las dos propuestas presentadas aborda la Omnisciencia Lógica satisfactoriamente, y a continuación se verá por qué. La primera propuesta, que es una defensa al conocimiento de equivalencias semánticas, es interesante porque hace notar explícitamente la diferencia que hay entre el valor de verdad del conocimiento y su composición sintáctica. Véase que, si bien  $(\Phi \leftrightarrow \Psi)$  y  $\sim\sim[\sim(\Phi \& \sim\Psi) \& \sim(\Psi \& \sim\Phi)]$  tienen el mismo valor de verdad, la formulación del primer enunciado difiere respecto al segundo, por lo cual, podría argumentarse que tal diferencia sintáctica es causada por distintos procesos inferenciales para derivarlas.

Sin embargo, la propuesta no está exenta de problemas. En general, no es totalmente certero que las formulaciones semánticamente equivalentes causen el mismo efecto en el comportamiento de agentes. Por mencionar brevemente algunos casos, considérese que en áreas de estudio como la *Lógica e Información* hay enfoques semánticos, *la semántica dinámica* por mencionar uno, donde el orden en el que la información llega a un agente determina cómo afectará su reacción ante ella, y de esta manera se enfatiza la diferencia entre fórmulas semánticamente equivalentes como  $(\Phi \& \Psi)$  y  $(\Psi \& \Phi)$ .

Con base en la misma idea, se pueden reproducir situaciones como las anteriores con *lógica dinámica epistémica de anuncios públicos*. Por ejemplo, es posible que dos secuencias de anuncios, que representan una equivalencia semántica, tengan efectos divergentes en un modelo de Kripke de estados posibles. Formalmente, sean  $[\Phi] \& [\Psi]$  y  $[\Psi] \& [\Phi]$  dos secuencias de anuncios, la primera tiene un efecto  $X$  en el modelo, mientras que la segunda un efecto  $Y$ , donde  $X \neq Y$ . En otros términos, el orden de los anuncios es relevante para la actualización de un modelo semántico<sup>46</sup>.

Además, para el caso particular del termostato, también podría argumentarse que el contenido semántico de  $(\Phi \& \Psi)$  y  $(\Psi \& \Phi)$  es producto de interpretaciones que se le adscriben a un hecho, i.e. el hecho de que el termostato detecta cierta temperatura; una interpretación corresponde a grados Fahrenheit, y otra, a Centígrados. Así pues, la equivalencia semántica recae en estas interpretaciones externas, lo cual no implica necesariamente que el termostato tenga capacidad de identificarlas como equivalentes.

Respecto a la segunda propuesta conceptual que expone Parikh, esta está basada en una concepción de agentes artificiales como entidades que emplean algoritmos para computar inferencias. Sin embargo, para re-definir el conocimiento bajo esta concepción se requiere un aparato formal que se aleja de la semántica estándar de Kripke (que es donde yace el problema de Omnisciencia Lógica). Este aparato, donde se formaliza el conocimiento algorítmico, se tratará en la sección 4.1.3 y se presentarán algunos de sus problemas al abordar la Omnisciencia Lógica.

---

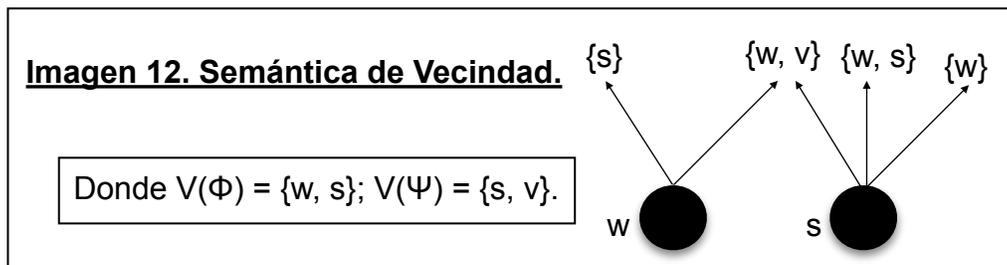
<sup>46</sup> Por razones de espacio no será posible abordar en este trabajo la semántica dinámica ni la lógica dinámica epistémica de anuncios públicos. Para una introducción a ambos temas puede consultarse *Dynamic Semantics* (2010) de Jan van Eijck y Albert Visser en *Stanford Encyclopedia of Philosophy* (<http://plato.stanford.edu/entries/dynamic-semantics/>) y *Dynamic Epistemic Logic* (2008) de Hans van Ditmarsch, Wiebe van der Hoek y Barteld Kooi (<http://www.iep.utm.edu/wp-content/media/DynamicEpistemicLogic.pdf>), respectivamente.

### 4.1.2 (b) Condiciones semánticas alternas del conocimiento

Existen diversas opciones, semánticas y sintácticas, para modificar la condición semántica del operador K. Entre las más relevantes están *la semántica de vecindad*, *el enfoque sintáctico*, *los estados imposibles*, *la semántica de Kripke con lógica no estándar*, y *el enfoque de conciencia lógica (awareness)*. Todos tienen sus problemas.

#### 4.1.2.1 Semántica de Vecindad

En semántica de vecindad ( $M_N$ ) (pp. 1-7, Pacuit, 2014), el agente no conoce hechos expresados por fórmulas, sino proposiciones. Las proposiciones están formadas por conjuntos de estados,  $(M_N, s) \models K\Phi$  sii en la lista de conjuntos de estados asignada a  $s$ , se encuentran todos aquellos donde  $\Phi$  se cumple. Intuitivamente, el agente debe tener acceso a ciertos conjuntos de estados epistémicamente posibles para conocer una proposición. Hay una función de vecindad ( $N$ ) que asigna a cada estado las proposiciones que se consideran necesarias. Sea el siguiente un modelo  $M_N$ :



En este modelo, la omnisciencia lógica no se mantiene en su totalidad, pues por ejemplo, el conocimiento no está cerrado respecto a la eliminación de la conjunción. En particular, hay un contra-ejemplo a una instancia de la misma, a saber  $K(\Phi \& \Psi) \rightarrow K(\Phi)$ .

Antes de explicar el ejemplo debe mencionarse que una conjunción, en Semántica de Vecindad, se satisface en el conjunto (o proposición) que surge de la

intersección ( $\cap$ ) de los conjuntos donde se satisface cada conyunto. Así pues, para el conocimiento de una conjunción en un estado  $t$ , se requiere que la función  $N$  asigne a  $t$  como conjunto vecino (o como proposición necesaria), el conjunto donde se satisface tal conjunción.

Ahora bien, véase que mientras  $K(\Phi \ \& \ \Psi)$  se satisface en  $w$ ,  $K\Phi$  no se satisface en el mismo estado. Primero,  $K(\Phi \ \& \ \Psi)$  se satisface en  $w$  porque a  $w$  se le ha asignado el conjunto  $\{s\}$ , en el cual se satisface  $(\Phi \ \& \ \Psi)$ , pues de  $\{w, s\} \cap \{s, v\}$ , conjuntos donde se satisface  $\Phi$  y  $\Psi$  respectivamente, da como resultado el conjunto  $\{s\}$ . Por otro lado, sucede que  $(M_N, w) \not\models K\Phi$ , pues para que  $K\Phi$  se satisficiera en  $w$ ,  $N$  tendría que asignar el conjunto de estados  $\{w, s\}$  como vecino de  $w$  pero no es así. Así pues,  $(M_N, w) \not\models K(\Phi \ \& \ \Psi) \rightarrow K(\Phi)$ <sup>47</sup>.

Lo que sucede con el conocimiento de agentes en Semántica de Vecindad es que éste es más difícil de obtener. A diferencia de la Semántica de Kripke estándar, donde las fórmulas conocidas deben cumplirse en todo estado que el agente considera posible, los agentes requieren un mayor nivel de introspección, es decir, un conocimiento total de los estados epistémicos sobre una proposición, y no solo los que consideran posibles respecto a la misma (p. 3, Artemov & Kuznets, 2014).

Pero la solución a la Omnisciencia Lógica es parcial, ya que el conocimiento de equivalencias sintácticas continúa siendo válido. Véase que  $(M_N, w) \models K(\Phi \ \& \ \Psi)$  y  $(M_N, w) \models K(\Psi \ \& \ \Phi)$ , pues  $\{s\} \cap \{w, v\} = \{w, s, v\}$ ; sucede que  $\{w, s\} \subseteq \{w, s, v\}$  donde  $V(\Phi) =$

---

<sup>47</sup> Nótese, sin embargo, que aunque  $K\Phi$  falla en  $w$ , es el caso que  $(M_N, w) \models KK\Phi$ , pues  $(M_N, s) \models K\Phi$  al tener el estado  $s$  como vecino al conjunto  $\{w, s\}$ , y  $N$  asigna  $\{s\}$  a  $w$  como conjunto vecino, siendo  $\{s\}$  el único conjunto donde  $K\Phi$  se satisface. Esta información no es crucial para el ejemplo, pero permite dar una imagen más completa del comportamiento semántico de fórmulas epistémicas en este enfoque, en este caso particular de la introspección positiva.

$\{w, s\}$ , y  $\{s, v\} \subseteq \{w, s, v\}$  donde  $V(\Psi) = \{s, v\}$ , y viceversa. Ergo, axiomas como  $K(\Phi \& \Psi) \leftrightarrow K(\Psi \& \Phi)$  continúan siendo válidos.

#### 4.1.2.2 Enfoque Sintáctico

El enfoque sintáctico (pp. 338-339, Halpern, et.al.; pp. 3-4, Halpern & Puccella) consiste en *listar explícitamente* las fórmulas que se conocen en cada estado posible. Se conoce  $\Phi$  explícitamente en  $w$  cuando  $\Phi$  pertenece a la lista (C) de formulas verdaderas en  $w$ :  $w \models K\Phi$  sii  $\Phi \in C(w)$ . En este enfoque se pierden las relaciones de acceso a estados posibles, pues lo que determina el valor de verdad de las fórmulas es una función sintáctica ( $\sigma$ ) que, reemplazando la función de valuación de lógica epistémica y doxástica ( $V$ ), su papel es asignar valores de verdad a las fórmulas.

Mediante esta función se determina el conocimiento (y creencias) que los agentes tienen en cada estado. Con base en esta estrategia, la expresión  $\sigma(s)(K_i\Phi) = \textit{verdadero}$  significa que el agente  $i$  conoce  $\Phi$  en el estado  $s$ .

La consecuencia del enfoque sintáctico es la ausencia total de agentes omniscientes; el conocimiento no está cerrado bajo implicación porque puede ser el caso que  $\sigma(s)(K_i(\Phi \rightarrow \Psi), K_i\Phi) = \textit{verdadero}$  pero  $\sigma(s)(K_i\Psi) = \textit{falso}$ . Lo mismo sucede con el conocimiento del resto de axiomas y equivalencias lógicas, ya que, por ejemplo, aunque  $(\Phi \leftrightarrow \Psi)$  sea el caso, no existe requisito alguno bajo el que la función  $\sigma$  deba hacer verdadero a  $K\Phi$  cuando  $K\Psi$  es el caso, y viceversa.

Pero al haber una ausencia total de omnisciencia se pierden todas las propiedades de los axiomas de  $S_5$  y  $KD_{45}$ . Como “ $\sigma$ ” no garantiza la validez de estas expresiones, propiedades como la veracidad (T) dejan de cumplirse, pues de  $(\sigma)(K_i\Phi) =$

*verdadero* no se sigue necesariamente  $(\sigma)(\Phi) = \textit{verdadero}$ . Ergo, las capacidades deductivas del agente son excesivamente limitadas en este enfoque.

#### **4.1.2.3 Estados Imposibles**

Esta propuesta consiste en que el agente considere relevantes para su conocimiento los estados imposibles (pp. 4-5, Halpern & Puccella; pp. 357-359, 361-362, Halpern, et.al.), aquellos donde los teoremas y axiomas no son válidos. Para ello, se combina la Semántica de Kripke con la función “ $\sigma$ ” del enfoque sintáctico. La parte semántica consiste en dar acceso a los agentes a estados imposibles, en los cuales, se asignan arbitrariamente valores de verdad a fórmulas epistémicas con “ $\sigma$ ”. La Omnisciencia Lógica es eliminada, por las mismas razones que en el enfoque sintáctico, pero, al igual que en este, las capacidades deductivas de los agentes son limitadas exceso, pues habría que forzar la función para recuperar propiedades epistémicas<sup>48</sup>.

#### **4.1.2.4 Semántica de Kripke con Lógica no Estándar**

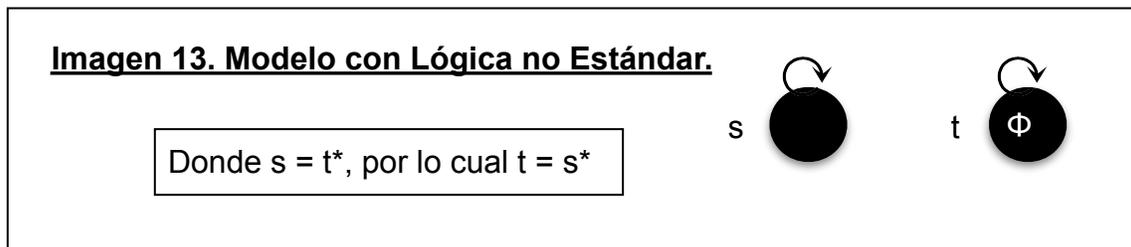
En lógica no estándar (pp. 346-351, Halpern, et.al.) hay una definición semántica distinta para la negación. Para ello se introducen estados adyacentes en modelos de Kripke. Sea  $s$  un estado en Semántica de Kripke con lógica no estándar, en éste existen dos conjuntos de formulas ( $F_F$  y  $F_V$ ). En  $F_F$  se encuentran las fórmulas falsas de  $s$ , y en  $F_V$ , las verdaderas. Con base esta idea, “ $s^*$ ” es un estado adyacente de  $s$  dado que posee los conjuntos  $F_F^*$  y  $F_V^*$ , donde en  $F_F^*$  figuran las fórmulas que no aparecen en  $F_F$  tal que  $F_V = F_F^*$ . Intuitivamente,  $s$  y  $s^*$  son dos BDs para un mismo estado<sup>49</sup>.

---

<sup>48</sup> Cabe mencionar que las tautologías proposicionales mantienen su validez en este contexto, pues no tiene sentido considerar estados lógicamente imposibles para definirlos semánticamente.

<sup>49</sup> Por esta razón, este enfoque es útil para representar el conocimiento de una BD con distintas fuentes de información, susceptibles de ser contradictorias una para la otra.

Bajo estas condiciones,  $(M, s) \models \sim\Phi$  sii  $(M, s^*) \models \Phi$ , i.e.  $\Phi \in F_V^*$ . Ello permite casos donde  $\Phi$  y  $\sim\Phi$  sean ambos verdaderos en un estado, pues es posible que  $\Phi \in F_V$  y  $\Phi \in F_V^*$ , por ejemplo, cuando las bases de datos  $s$  y  $s^*$  se contradicen una a la otra dado que su información proviene de distintas fuentes. Sea el siguiente un modelo de Kripke con lógica no estándar:



En el marco se demuestra que en este enfoque no hay formulas proposicionales válidas, ya que todo lo que se satisface en  $t$  no podrá satisfacerse en  $s$ , y viceversa. Además, los agentes no son omniscientes respecto al axioma (K), pues  $(M, t) \models \Phi$  se cumple, y  $(M, t) \models K(\Phi \rightarrow \Psi)$  también, ya que  $(\Phi \rightarrow \Psi) = (\sim\Phi \vee \Psi)$  y  $(M, t) \models \sim\Phi$  dado que  $\sim\Phi$  se satisface en  $s$  (i.e.  $t^*$ ), pero aún así,  $\Psi$  no se cumple en  $t$ ,  $(M, t) \models \sim K\Psi$ .

Aún así, el enfoque tiene problemas. Primero, los agentes continúan siendo omniscientes, por ejemplo, respecto a la cerradura bajo disyunción:  $(M, t) \models K\Phi \rightarrow K(\Phi \vee \Psi)$ . También, como la implicación material ya no representa la noción de *consecuencia lógica*, a causa de las modificaciones para la semántica de la negación, en estos modelos se introduce la *implicación estricta*  $[\Box(\Phi \rightarrow \Psi)]$ <sup>50</sup> para recuperar tal

<sup>50</sup> Donde  $\Box$  es un operador modal que denota necesidad, tal que  $\Psi$  es el caso siempre que  $\Phi$  lo es.

noción. Sucede que el conocimiento de agentes está cerrado bajo implicación estricta<sup>51</sup>.

#### **4.1.2.5 Función de Consciencia Lógica**

En esta propuesta (pp. 362-365, Halpern, et.al.) se introduce una función sintáctica de consciencia "A" (*awareness*) tal que  $(M, s) \models A_i\Phi$  sii  $\Phi \in A_i(s)$ , *i está consciente de  $\Phi$  en s si  $\Phi$  pertenece a las fórmulas de las que está consciente en s*. Con base en A se define un operador de conocimiento explícito "X",  $(M, s) \models X_i\Phi$  sii  $(M, s) \models A_i\Phi \ \& \ (M, s) \models K_i\Phi$ , y a su vez, los axiomas de  $S_5$  y  $KD_{45}$ , y sus respectivas propiedades, pueden reformularse bajo esta noción. Por ejemplo  $(K), (X_i\Phi \ \& \ X_i(\Phi \rightarrow \Psi) \ \& \ A_i\Psi) \rightarrow X_i\Psi$ .

Esta función (A) tiene las mismas desventajas que el enfoque sintáctico respecto al problema de Omnisciencia Lógica. Sin embargo, es una propuesta flexible e interesante que permite modelar la disponibilidad del conocimiento de diversas maneras, y que la lógica de justificaciones modifica y recupera (sección 4.2).

Como puede observarse, en estas propuestas surgen dos problemas. Uno es que debilitan en exceso la capacidad deductiva de los agentes, como es el caso de los estados imposibles, el enfoque sintáctico y el de consciencia (p. 3, Wang, 2010). El agente no puede inferir la clase de consecuencias lógicas que se restringen, aún si algunas de este tipo se encontraran al alcance de sus capacidades. El segundo problema es que la superación de la Omnisciencia es parcial (p, 2, Wang), como es el caso de la semántica de vecindad y la semántica de Kripke con lógica no estándar.

---

<sup>51</sup> De hecho, puede combinarse el enfoque de estados imposibles con el de lógica no estándar, de manera que los estados imposibles son estados no estándar donde suceden casos como  $\Phi$  y  $\sim\Phi$ . Al igual que en el enfoque de lógica no estándar, se introduce la implicación estricta para modelar la consecuencia lógica, por lo cual los agentes mantienen Omnisciencia Lógica respecto a ésta (pp. 359-362, Halpern, et.al.)

Adicionalmente, ninguna de estas posibles soluciones explica algunos casos de conocimiento insuficiente, por ejemplo, cuando una computadora no es capaz de resolver una tarea, aún si se programó para llevarla a cabo (p. 3, Artemov & Kuznets). En esta situación, la computadora podría no ser capaz de resolverla porque sus recursos espaciales y/o temporales no se lo permiten<sup>52</sup>. Pero como en estas propuestas la capacidad epistémica se define a través de filtros de racionalidad preestablecidos, no se consideran los recursos que cada agente tiene disponibles para cumplir con sus tareas.

#### **4.1.3 (c) Conocimiento algorítmico**

Para enfrentar la Omnisciencia Lógica, el conocimiento se define como la información que el agente es capaz de inferir por medio de su algoritmo. Con esta estrategia se pretenden diseñar *lógicas de conocimiento algorítmico* para distinguir, a diferencia de las propuestas anteriores, las consecuencias lógicas que puede inferir de las que no<sup>53</sup>.

Este es el enfoque alternativo al de la semántica de Kripke con lógicas modales, en el que Halpern, et.al., Parikh, y otros autores sobre inteligencia artificial, han pensado para modelar el conocimiento, especialmente de bases de datos. La idea es que el agente pueda determinar si por medio de su algoritmo y la información que tiene disponible puede derivar cierta fórmula (p. 394-396, Halpern et.al.):

---

<sup>52</sup> La noción de *recursos espacio-temporales* pertenece a la *Teoría de Complejidad Computacional*. En esta área se abordan preguntas del tipo *¿qué problemas son eficientemente computables?* (p. 261, Aaronson, 2013). Una importante incógnita en este campo es la cantidad de recursos espaciales (una hoja de papel, una cinta magnética) y temporales (segundos, años) que requiere una máquina, agente, o un ser humano, para resolver un problema. En este contexto, el empleo de una cantidad de recursos cuya medida corresponde a un polinomio (P) pertenece a problemas computables, mientras que si le corresponde una medida exponencial (EXP) se trata de un problema no computable (p. 264.).

Dado que la relación entre tiempo y espacio en Complejidad Computacional es profunda e intrincada, se hará referencia específicamente al tiempo cuando surja la noción de medida polinomial. Solo queda mencionar que el estudio de Sistemas Distribuidos es un área en la que interseca la complejidad computacional y la reflexión filosófica sobre el conocimiento (p. 263).

<sup>53</sup> Un algoritmo es eficiente si da soluciones en tiempo P, e ineficiente si es EXP (Aaronson, p. 265).

- $(SI, r(m)) \models X_{BD}\Phi$  sii  $A(\Phi, I) = Si$ , para  $A = alg_{BD}(r, m)$  y  $I = data_{BD}(r, m)$

Donde  $X$  denota conocimiento explícito,  $I$  el estado local de BD,  $alg_{BD}(r, m)$  su algoritmo en  $(r, m)$ , y  $data_{BD}(r, m)$  su información disponible en  $(r, m)$ . Técnicamente, el operador  $X$  se define en virtud de una función de consciencia lógica, con relación a lo que es posible inferir con base en los recursos  $alg_{BD}(r, m)$  y  $data_{BD}(r, m)$ .

Acerca de este enfoque, Artemov y Kuznets lo consideran una buena estrategia para cuantificar y medir la complejidad de las inferencias lógicas con relación a los recursos algorítmicos e informáticos de agentes (p. 4). Aunque por otro lado, observan que esta se restringe a un contexto semántico, ya que *“no es claro cómo computar una lógica para un conjunto de algoritmos con capacidad delimitada, dado que las medidas cuantitativas permanecen fuera del lenguaje de la teoría”* (trad.)<sup>54</sup>. Para solucionar esta restricción hace falta una sintaxis con la que pueda expresarse el costo que tiene la inferencia de fórmulas para cada nivel de capacidad epistémica.

Con el objetivo de ofrecer claridad respecto a las insuficiencias que emergen en las propuestas anteriormente presentadas, el resto de este capítulo estará dedicado a la *lógica de justificaciones*. Este es un formalismo derivado de las lógicas modales epistémicas y doxásticas, equipada con una función de consciencia lógica que relaciona fórmulas con términos que representan evidencia epistémica.

Una de las principales insuficiencias a las que se ha atendido por medio de esta lógica es la que surge con las propuestas semánticas de lógica epistémica y doxástica: ofrecer un método formal para representar agentes no omniscientes, procurando

---

<sup>54</sup> *“It is not clear how to compute the logic of a particular set of complexity-bound algorithms given that the quantitative measures remain outside of the language of the theory.”* (cita original).

eliminar el mayor número de situaciones donde los agentes tengan capacidades epistémicas idealizadas<sup>55</sup>, pero evitando también constreñir en exceso su capacidad deductiva.

Este método consiste en la introducción de una noción formal de justificación epistémica en el lenguaje, representada por *términos de justificación*, de tal manera que toda fórmula antecedida por estos términos se considera respaldada (justificada) por evidencia relevante. Para enfrentar la Omnisciencia Lógica, la noción de justificación se emplea como criterio para determinar cuál es la información que los agentes están justificados a conocer (creer), y cuál es la que mantienen solamente en forma de conocimiento (creencia) implícito.

La lógica de justificaciones también atiende al principal problema que enfrenta la lógica de conocimiento algorítmico de Halpern et.al, a saber, la falta de expresividad para introducir una notación al lenguaje que permita representar costos a las inferencias de los agentes. Se verá que la lógica de justificaciones tiene un lenguaje lo suficientemente rico para medir la derivación de fórmulas, en términos de complejidad computacional (sección 4.4.1).

Antes de presentar formalmente la lógica de justificaciones, es importante mencionar que esta supone solamente uno de los diversos desarrollos lógicos que abordan el problema de agentes omniscientes. Algunas de estas lógicas son la *deductiva modal* de Pucella (2006) y la *epistémica temporal (t-MEL)* de Wang (2010), donde en ambas se hace uso particular de la función de consciencia. En la primera se introduce un operador explícito de conocimiento, construido con base en esta función, y

---

<sup>55</sup> Se verá que la lógica de justificaciones mantiene algunas idealizaciones, por ejemplo, el conocimiento respecto a tautologías.

en la segunda, la función relaciona fórmulas con literales que representan momentos del tiempo<sup>56</sup>, este es, el tiempo que los agentes utilizan para llevar a cabo sus inferencias. Recientemente también se han estudiado diversos métodos para modelar agentes no omniscientes en lógica epistémica y doxástica.

Tanto la lógica t-MEL, como los desarrollos más recientes de agentes no omniscientes en lógica epistémica y doxástica, volverán a mencionarse con un poco de más detalle durante el siguiente capítulo, dada su gran importancia como objetos de estudio en futuras líneas de investigación en el marco de la Omnisciencia Lógica y la lógica de justificaciones. Sin embargo, un tratamiento profundo de estos trabajos se encuentra fuera de los límites de esta tesis, y una variedad de otras lógicas para agentes no omniscientes quedará sin mencionarse por las mismas razones<sup>57</sup>.

En este trabajo se ha elegido la lógica de justificaciones como herramienta formal para reflexionar sobre la Omnisciencia Lógica, porque permite abordar el problema con una dimensión que aún ha sido poco trabajada en las lógicas

---

<sup>56</sup>

- La *lógica deductiva modal* es muy cercano al enfoque de consciencia (sección 4.1.2.5). En ella hay dos operadores de conocimiento: K y X, donde K se mantiene como un operador de conocimiento implícito, y X, para el explícito. Para definir X se emplea una función de consciencia lógica (A), y se requiere que, para  $K\Phi$ , el agente pueda deducir sintácticamente  $\Phi$  de un conjunto de observaciones que representan su información actual (pp. 4-5).
- En t-MEL se introduce una función de consciencia para expresar hechos temporales sobre el conocimiento, con el objetivo de enfrentar la Omnisciencia Lógica. El autor parte de una idea intuitiva: *el conocimiento ocurre en el tiempo y no obtenemos todo el conocimiento en el mismo momento* (p. 3). Bajo esta idea, se propone representar el tiempo que le lleva a un agente inferir consecuencias lógicas, identificando las que le llevarían un tiempo fuera de su alcance, y las que podría inferir en un tiempo aceptable. La lógica temporal de Wang se abordará, aunque brevemente, en el 5º capítulo.

<sup>57</sup> Respecto a la mención de desarrollos de lógicas epistémicas con agentes no omniscientes, es pertinente citar a Benthem y Quesada (2010), quienes afirman que la Omnisciencia Lógica no es razón suficiente para tirar por la borda la lógica epistémica, lo que debe hacerse, más bien, es enriquecerla con un conjunto suficientemente rico de actitudes inferenciales de agentes que den cuenta de cómo es que obtienen conocimiento explícito (p. 6).

Los autores sostienen un argumento muy sólido respecto a la defensa de la lógica epistémica, y logran estimular su continuo desarrollo para enfrentar problemas como el de agentes omniscientes. Una vez más, esta tesis no tiene el espacio suficiente para tratar estos desarrollos, por lo que no queda más que remitir al lector a los respectivos textos que, sin lugar a dudas, hacen de la lógica epistémica un marco formal con un futuro muy prometedor e interesante por delante.

mencionadas anteriormente, la justificación. Con esta noción, además, se abre la posibilidad de construir definiciones de creencia y conocimiento explícito para estudiar actos de inferencia en agentes no omniscientes.

Además de la ventaja de ofrecer un punto de vista distinto al problema de Omnisciencia Lógica en el Sistema Distribuido *Listener-Teller*, en el siguiente capítulo se defenderá que con el empleo de la lógica de justificaciones existe la posibilidad de reflexionar sobre aspectos aún implícitos en modelos lógicos aún carentes de la noción de justificación. Por ejemplo, se propondrá con base en un modelo preliminar, que la representación explícita de justificaciones podría dar cuenta de un fenómeno intuitivo sobre el conocimiento: dadas las limitaciones de su memoria, los agentes como Listener podrían olvidar la información recibida con anterioridad, pero si ésta información está justificada, existe la posibilidad de que la agente la recuerde en el futuro al remitirse a la evidencia (i.e. términos) de donde proviene.

#### 4.2 ¿Qué es la lógica de justificaciones?

Para Fitting, si bien el enfoque de consciencia no es una solución definitiva a la Omnisciencia Lógica, “*es un formalismo muy general... y se trata, no de una herramienta, sino de una caja de herramientas con elementos útiles, y otros, poco útiles*” (p. 4, Fitting1, 2006, trad.)<sup>58</sup>. Con base en esta idea, el autor define la lógica de justificaciones como una extensión de este enfoque (pp. 1, 6-7, Fitting2, 2014).

Acorde a esta definición, la lógica de justificaciones, dice Artemov (p. 2, 2008), ofrece un mecanismo basado en evidencia para el análisis del conocimiento, a través

---

<sup>58</sup> “*The framework is, in fact, very general. It is more of a toolbox, able to contain many things useful or not, than a tool in itself.*” (cita original).

de una función de consciencia, enriquecida por una estructura de términos de justificación. En términos generales, esta lógica contiene una función que asigna literales a fórmulas, los cuales representan evidencia relevante para las fórmulas.

#### 4.2.1 Sintaxis<sup>59</sup>

$$\Phi, \Psi ::= p \mid \sim\Phi \mid \Phi \vee \Psi \mid t:\Phi$$

Donde  $\Phi$  es una fórmula que puede ser atómica o parte de una fórmula compuesta con las conectivas de lógica proposicional. Al igual que antes, mediante la negación y disyunción se definen el resto de las conectivas. Por último,  $t$  es un término de justificación, y  $t:\Phi$  es una fórmula justificada. Sintácticamente, la lógica de justificaciones es similar a  $S_5$  y  $KD_{45}$ , excepto que los términos reemplazan los operadores  $K$  y  $B$  (p. 7, Fitting2). Los términos de justificación, que anteceden fórmulas, representan evidencia que les respalda. “ $t:\Phi$ ” se lee como *t es justificación relevante para  $\Phi$* , o bajo una visión justificacionista, *el agente acepta t como justificación de  $\Phi$* .

Hay términos variables ( $x, y, z, \dots$ ) y constantes ( $a, b, c, \dots$ ); los primeros son arbitrarios y están sujetos a cambios, mientras que los segundos son fijos y justifican los axiomas del sistema. Los términos también pueden ser simples, como  $t:\Phi$ , o compuestos. Hay dos tipos de términos compuestos:

- $[t*s]:\Psi$ , donde “ $*$ ” es un símbolo de aplicación y representa la intersección de dos términos para derivar  $\Psi$ .
- $[t+s]:\Psi$ , donde “ $+$ ” es un símbolo de adición y representa la unión de dos términos.
- O también, combinaciones del tipo  $[q*r] + [s*t]:\Psi$  y  $[q+r] * [s+t]:\Psi$ .

#### 4.2.2 Semántica

---

<sup>59</sup> La sintaxis, semántica y axiomatización pertenece a *Why Do We Need Justification Logic?* de Artemov (2008), excepto donde se indique. Hay algunas modificaciones formales para evitar abuso de notación.

La semántica, propuesta por Melvin Fitting, está adaptada a la de Kripke. Sea un modelo Kripke-Fitting ( $M_F$ ) compuesto de los elementos  $(W, \rightarrow, f, \models)$ . Donde  $W$  es un conjunto de estados, " $\rightarrow$ " una relación entre estados, " $\models$ " una relación de consecuencia que preserva la condición semántica de conectivas booleanas, y  $f$  una *función de admisibilidad para evidencia*, tal que  $f(t, \Phi) \subseteq W$  para cualquier término  $t$ . Esta función determina el conjunto de estados posibles donde  $t$  se considera evidencia admisible para  $\Phi$ .

$f$  es un elemento de importancia para la condición semántica de justificaciones:

- $w \models t: \Phi$  sii:  $(w \models \Phi, \forall w' \text{ tal que } w \rightarrow w') \ \& \ (w \in f(t, \Phi))$

Además,  $f$  está cerrado bajo aplicación y suma:

- [\*]:  $f(s, \Phi \rightarrow \Psi) \cap f(t, \Phi) \subseteq f(s*t, \Psi)$
- [+]:  $f(s, \Phi) \cup f(t, \Phi) \subseteq f(s+t, \Phi)$

Es importante mencionar que la cerradura bajo aplicación y suma de la función  $f$ , no significa que el conocimiento sobre justificaciones del agente esté cerrado bajo estas acciones inferenciales. Lo que significa, más bien, para el caso de la cerradura bajo aplicación, es que si los términos  $s$  y  $t$  son evidencia relevante (admisible) para estar justificados en saber  $\Phi \rightarrow \Psi$  y  $\Phi$  respectivamente en cierto estado, entonces el término  $[s*t]$  es, consecuentemente, evidencia relevante para justificar  $\Psi$  en el mismo estado. El caso de la cerradura bajo suma es similar, si  $s$  y/o  $t$  son términos relevantes para justificar  $\Phi$  en un estado, entonces la suma de ambos también lo es.

Se constriñen condiciones para marcos de estados posibles de la misma manera en cómo se hace en lógica epistémica y doxástica (p. 15, Fitting2). Finalmente, es

relevante indicar que pueden emplearse modelos Kripke-Fitting para trabajar con operadores K y B, y términos de justificación simultáneamente.

### 4.2.3 Axiomatización

$J_0$  es la lógica de justificaciones básica y sus axiomas son los siguientes:

- (AJ1) Axiomas de lógica proposicional clásica y regla de *Modus Ponens*.
- (AJ2) Axioma de aplicación:  $s: (\Phi \rightarrow \Psi) \rightarrow (t: \Phi \rightarrow (s*t): \Psi)$ . (Por ende, no existe (K)).
- (AJ3) Axioma de suma:  $[s: \Phi \rightarrow (s+t): \Phi] \& [s: \Phi \rightarrow (t+s): \Phi]$ .

Al igual que en el caso de la cerradura bajo aplicación y suma de la función  $f$ , los axiomas AJ2 y AJ3 no implican que el conocimiento de justificaciones por parte del agente estén cerrados bajo estas acciones. AJ2 dice que si  $s$  y  $t$  son evidencia suficiente para que el agente sepa  $\Phi \rightarrow \Psi$  y  $\Phi$  respectivamente, entonces éste puede estar justificado en saber  $\Psi$ , en caso de que lleve a cabo una operación de aplicación con estos términos. El caso es similar para el axioma AJ3; por ejemplo, si el agente sabe  $\Phi$  justificadamente con el término  $s$ , entonces también lo estará si ejecuta una acción inferencial de suma del tipo  $[s+t]$ , lo cual también captura la monotonía respecto a los términos.

Se dice que la lógica  $J_0$  pertenece a un agente totalmente escéptico, pues solamente se pueden efectuar justificaciones relativas:  $(x: \Psi, y: \Gamma \dots) \rightarrow t: \Phi$ .

La lógica de justificaciones ofrece un mecanismo para representar distintos niveles de consciencia lógica, llamado *Conjunto de Especificación* (CS). Para cada lógica  $J$  hay un CS que contiene sus respectivos axiomas y términos constantes:  $J_0 + CS = J_{CS}$  (y si  $CS = 0$  entonces se trata de  $J_0$ ).

Formalmente, para cada lógica  $J$  hay un CS compuesto de los elementos:  $\{e_n:e_{n-1}:\dots:e_1:A\}$  ( $n \geq 1$ ). Donde  $A$  es axioma de  $J$  y  $e_1\dots e_n$  son términos constantes similares con índices. CS contiene todas las justificaciones intermedias:  $(\{e_n:e_{n-1}:\dots:e_1:A\} \in CS) \rightarrow (\{e_{n-1}:\dots:e_1:A\} \in CS)$ . La conexión entre la semántica de Kripke-Fitting y CS es la siguiente: *un marco  $M_F$  es consistente con CS en un estado  $w$  si  $w \in f(c, \Phi) \forall (c: \Phi) \in CS$ . Y es consistente en general si esto el caso para todos los estados del marco.*

Entre los tipos de CS se encuentran los que son *axiomáticamente apropiados*:  $\forall A \exists e_1$  tal que  $e_1:A \in CS$ , y si  $\{e_n:\dots:e_1:A\} \in CS$ , entonces  $\{e_{n+1}:e_n:\dots:e_1:A\} \in CS$ . Y los CS *totales*,  $\forall A \exists e_1\dots e_n, \{e_n:e_{n-1}:\dots:e_1:A\} \in CS$ . Un CS total es axiomáticamente apropiado.

En el teorema que se presentará a continuación se muestra que para todo Conjunto de Especificación (CS) que es axiomáticamente apropiado, su lógica correspondiente ( $J_{CS}$ ) satisface una propiedad llamada *internalización*. Que una lógica, por ejemplo  $J_{CS}$ , satisfaga internalización, significa que todos sus axiomas tienen un término constante de CS que los justifica, formalmente, significa que  $\models A \rightarrow \models t:A$ , para algún  $t \in CS$  de la lógica  $J_{CS}$ .

La satisfacción de esta propiedad se expresa mediante la *regla de internalización*, que es la siguiente:  $\forall CS$  es el caso que  $\forall A$ . de  $J_{CS} \exists e_1\dots e_n$  tal que  $\{e_n:\dots:e_1:A\}$ , donde  $A$  es un axioma y  $e_1\dots e_n$  términos constantes. A esta regla se le denominará R4.

Teorema:  $\forall CS$ , si CS es axiomáticamente apropiado entonces  $J_{CS}$  satisface la propiedad de internalización de R4. Prueba por hipótesis inductiva:  
Supóngase que CS es axiomáticamente apropiado y que  $\Phi$  es una fórmula,

- Si  $\Phi = A$ , entonces  $\exists e_1. e_1: A$ , pues  $A$  es axioma de  $J_{CS}$ , cuyo CS es axiomáticamente apropiado.
- Si  $\Phi \in CS$ , entonces  $\exists e_n. e_n: \Phi$ , pues  $\Phi$  es elemento de un CS axiomáticamente apropiado.
- Si  $(X \rightarrow \Phi)$ , entonces  $\exists s \exists t. \models s: (X \rightarrow \Phi) \ \& \ \models t: X$ , y por (AJ2),  $\models [s*t]: \Phi$

Como se ha mencionado anteriormente, a través de los CS, dependiendo de los axiomas justificados que contenga, se modelan distintos niveles la consciencia lógica de los agentes. Nótese además que R4 es una regla cercana a (G), de hecho, Artemov la define como una encarnación explícita de la misma (p. 6), pues todas las fórmulas válidas de una lógica de justificaciones particular tienen un término constante de justificación en su CS, tal que el agente está justificado en conocer estas fórmulas, sin llevar a cabo inferencia alguna.”

#### 4.3 ¿Cómo se expresa conocimiento y creencias en esta lógica?

Para expresar conocimiento y creencias se emplean lógicas de justificación análogas a las lógicas modales epistémica y doxástica. Esta idea está contenida en el *Teorema de Realización*, el cual afirma que cada instancia de los axiomas modales son convertibles a fórmulas con términos explícitos de justificación (p. 8, Fitting1). El método consiste en reemplazar los operadores de creencia y conocimiento que anteceden a tautologías, con términos constantes, y al resto, con términos variables, i.e., se recupera la información sobre la evidencia que respalda a cada componente de estas fórmulas<sup>60</sup>.

---

<sup>60</sup> La acción inversa, *Teorema de Olvido (forgetting)*, consiste en revertir el reemplazo, lo cual lleva a la pérdida de esta información.

La conversión de  $S_5$  tiene como resultado  $JT_{45}$ , que contiene todos los axiomas y reglas de  $J_0$ . Así también incluye (T)  $t: \Phi \rightarrow \Phi$  (facticidad), que es válido en marcos reflexivos. En este axioma se expresa que la justificación  $t$  es suficiente para que el agente sepa que  $\Phi$  es verdad (p. 12, Artemov). También, hay axiomas análogos a (4) y (5), los cuales son útiles para verificar pruebas que justifican fórmulas afirmadas mediante una operación de verificación positiva, y fórmulas negadas, mediante una de verificación negativa<sup>61</sup>.

$KD_{45}$  corresponde a  $JD_{45}$ , donde (T) no es axioma, y se añade el axioma (D):

- (D)  $t: \Phi \rightarrow \sim(t: \sim\Phi)$ , es decir, un término que justifica  $\Phi$  no puede justificar  $\sim\Phi$ . Válido en marcos seriales.

Un hecho intuitivo de  $JD_{45}$  es que las creencias justificadas no verdaderas no son conocimiento, dicho de otro modo, en esta lógica el axioma (T) no es válido.

Los teoremas derivados de los conjuntos axiomáticos  $JT_{45}$  y  $JD_{45}$  tienen una forma parecida a los de las lógicas modales. Véase por ejemplo la construcción de la cerradura bajo conjunción válida, que se convierte en la justificación de una conjunción a partir de la justificación de sus dos conyuntos (p. 7, Artemov). Considérese también que la siguiente demostración tiene el objetivo de presentar el funcionamiento del aparato sintáctico de la lógica de justificaciones:

i) Sea  $A \rightarrow (B \rightarrow (A \& B))$ , una tautología (TAUT) de lógica proposicional.

---

<sup>61</sup> Las expresiones axiomáticas correspondientes son las siguientes:

- (4)  $t: \Phi \rightarrow !t: (t: \Phi)$ , donde “!” es la instanciación de un término  $t$ . Válido en marcos transitivos.
- (5)  $\sim t: \Phi \rightarrow ?t: (\sim t: \Phi)$  donde “?” es la instanciación de un término  $t$  (p. 10, Fitting2). Válido en marcos euclidianos.

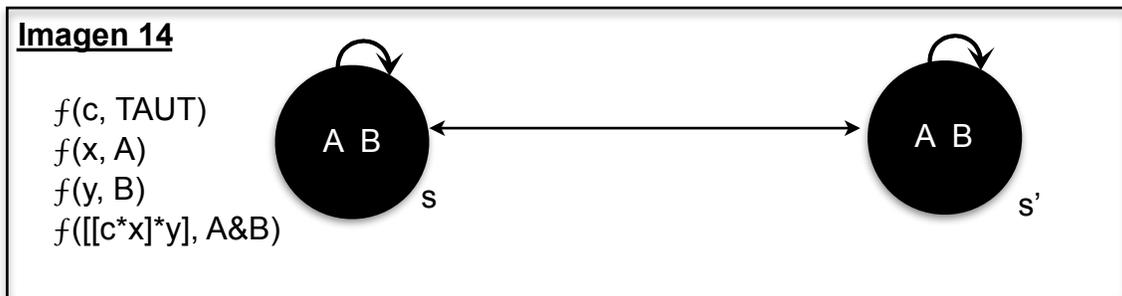
Donde “!” y “?” son operadores con los que se expresa meta-evidencia para la prueba de una fórmula justificada. Entiéndase por meta-evidencia la representación simbólica de la evaluación de una prueba; mediante esta evaluación se confirma que la prueba es correcta

En este sentido, “?” es un operador que expresa la verificación de una prueba que tiene como resultado la negación de una fórmula. Y “!” es un operador que expresa la verificación de una prueba que respalda la afirmación de una fórmula.

- ii)  $c: (A \rightarrow (B \rightarrow (A \& B)))$ , por internalización con R4 en (i), y donde  $\{c: \text{TAUT}\} \in \text{CS}$ .
- iii)  $x: A \rightarrow [c*x]: (B \rightarrow (A \& B))$  por AJ2 y *Modus Ponens* en (ii).
- iv)  $x: A \rightarrow (y: B \rightarrow [[c*x]*y]: (A \& B))$ , por AJ2 y *Modus Ponens* en (iii).
- v) Entonces  $(x: A \& y: B) \rightarrow [[c*x]*y]: (A \& B)$  dado que  $c: (A \rightarrow (B \rightarrow (A \& B)))$ . Cuya contraparte es  $(KA \& KB) \rightarrow K(A \& B)$  dado que  $K(A \rightarrow (B \rightarrow (A \& B)))$ <sup>62</sup>.

Todas las fórmulas proposicionales cobran una forma similar. Inclusive, no se admite la cerradura bajo equivalencias lógicas, a diferencia de algunas de las propuestas de la sección 4.1.2. De acuerdo con la intuición detrás de las justificaciones, las formulas complejas deberían ser más complicadas de inferir a diferencia de las más sencillas, aun si son lógicamente equivalentes (p. 7, Fitting2).

Para mostrar el funcionamiento del aparato semántico, y exponer con mayor claridad cómo la lógica de justificaciones permite un análisis del conocimiento y creencia más competente frente al problema de Omnisciencia Lógica, con relación a la lógica epistémica y doxástica, a continuación se presenta un modelo semántico del caso anterior:



El modelo se constituye de dos estados posibles, siendo  $s$  el actual. Las relaciones de acceso obedecen a los axiomas de JT<sub>45</sub>. A a la izquierda del estado actual se encuentran las fórmulas y los términos capturados por la función de evidencia ( $f$ ). Sea

<sup>62</sup> Mediante un razonamiento similar, empleando el axioma AJ3, se obtiene el siguiente resultado:  
 •  $(x: A \vee y: B) \rightarrow [a*x + b*y]: (A \vee B)$  dado que  $a: (A \rightarrow (A \vee B))$  y  $b: (B \rightarrow (A \vee B))$

c un término constante, y TAUT la tautología  $A \rightarrow (B \rightarrow (A \& B))$ , en la primera línea se expresa que el Conjunto de Especificación (CS) del agente contiene el término c, el cual, justifica el conocimiento de  $A \rightarrow (B \rightarrow (A \& B))$ .

En términos semánticos, el agente está justificado en conocer esta tautología en todo mundo posible, lo cual no es la excepción en s. Véase que las dos cláusulas de la condición semántica se satisfacen sin ningún impedimento; 1) TAUT se cumple en todo mundo posible por las mismas razones que en semántica de lógica epistémica, y 2) el agente está justificado en conocer TAUT en s, pues de acuerdo con  $f$ , c es suficiente en s para justificar TAUT, es decir  $f(c, TAUT)$ .

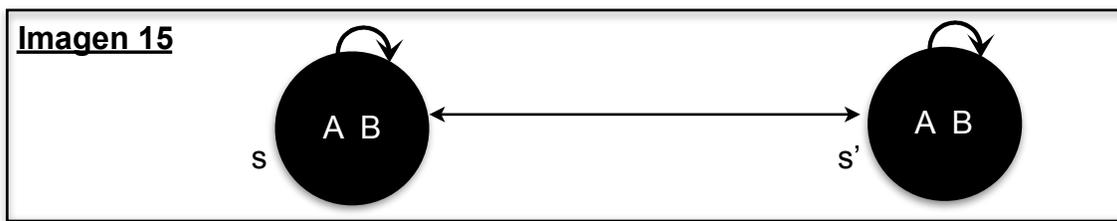
En la segunda y tercera línea se expresa que la función  $f$  relaciona los términos variables x, y, con las fórmulas A y B respectivamente. Obsérvese que ambas fórmulas se cumplen en los dos mundos a los que se tiene acceso, y además,  $f$  les ha asignado en s un término que las justifica, siendo x, y estos términos. A y B además, no solo se tratan de creencias, pues como A y B son verdaderas en s, y el modelo satisface reflexividad, se sigue  $(x: A \rightarrow A)$  y  $(y: B \rightarrow B)$ , tal que A y B son conocimiento.

Finalmente véase que  $(A \& B)$  se satisface en s, pues  $(A \& B)$  se cumple en todos los mundos accesibles desde s, y además, por medio de  $f$  se certifica que esta fórmula es una creencia justificada, en caso de que el agente efectúe dos operaciones de aplicación con los términos c, x, y, y obtenga el término complejo  $[[c*x]*y]$ , pues como se mostró en la construcción sintáctica,  $[[c*x]*y]: (A \& B)$ . Por razones similares a las anteriores  $(A \& B)$  tiene estatus de conocimiento.

Véase que axiomas como K y sus derivados no son válidos en el contexto de justificaciones, pues si bien  $c_n: (x: A \& y: B \rightarrow [x*y]: (A \& B))$  es una tautología que el

agente está justificado en saber con un término constante  $c_n$ , no es el caso que conocer el antecedente de la misma ( $x: A \& y: B$ ) implique conocer el consecuente  $[x^*y]$ : ( $A \& B$ ), pues hay una acción de inferencia de que el agente debe realizar “[ $x^*y$ ]” para llegar al consecuente. Como se sabe, esto no es así en  $S_5$ , ya que  $(KA \& KB) \rightarrow K(A \& B)$  es una validez. Con un método similar, es evidente que la omnisciencia por extensionalidad tampoco es válida en el contexto de las justificaciones.

En el caso de la lógica epistémica el agente conoce ( $A \& B$ ) porque la fórmula se satisface en todos los mundos a los que tiene acceso, pero no hay mayor explicación sobre cómo obtuvo esta información:



Bien la noción de creencia verdadera justificada puede ser criticada, y no considerarse adecuada para definir el conocimiento. Esta no es una discusión abordada en este trabajo. Lo que se quiere defender es que en lógica de justificaciones, la adición de la noción de justificación, a las de creencia y verdad, permite trabajar con una noción explícita de conocimiento (y creencia) para analizar situaciones epistémicas desde un punto de vista distinto que el de la lógica epistémica. Una visión en términos de justificación, evidencia, y en general, a partir de una distinción entre conocimiento implícito y explícito (entendiendo por conocimiento explícito una creencia verdadera justificada).

Esta defensa será reforzada en la siguiente sección, especialmente en la subsección 4.4.2, donde se emplea la lógica de justificaciones para analizar un contra-

ejemplo de la noción tripartita de conocimiento. Una vez más, el objetivo no será defender la noción tripartita, sino defender el hecho que la lógica de justificaciones es más competente en ciertos aspectos a diferencia de  $S_5$  y  $KD_{45}$ , gracias la introducción de términos de justificación en el lenguaje.

#### **4.4 ¿Qué ventajas tiene la lógica de justificaciones frente a $S_5$ y $KD_{45}$ ?**

Hay tres ventajas con relación a la expresión de hechos sobre el conocimiento y creencias, y con relación al análisis epistémico: la ausencia de agentes omniscientes<sup>63</sup>, la definición formal del conocimiento como *creencia verdadera justificada*, y una visión centrada en el agente donde se distingue su conocimiento actual del potencial.

##### **4.4.1 Pruebas de Omnisciencia Lógica**

En el contexto de las justificaciones explícitas, la omnisciencia se aborda como un problema de complejidad computacional, es decir, uno donde al agente se le adscribe un conocimiento que, dados sus recursos espaciales y temporales, no le es posible computar. Con la finalidad de detectar teorías epistémicas formales omniscientes, y cerrar la brecha entre racionalidad y ausencia total de consciencia lógica, Artemov y Kuznets han diseñado una estrategia que reemplaza los criterios de omnisciencia-ignorancia de propuestas anteriores (sección 4.1), por un mecanismo de prueba equipado con una graduación numérica (pp. 6 y 15), cuya estrategia consiste en medir

---

<sup>63</sup> Se recuerda al lector que cuando se dice que en lógica de justificaciones no hay agentes omniscientes, no se pretende afirmar que esta lógica es la solución definitiva a la Omnisciencia Lógica. Más bien, se quiere decir que con esta lógica se tiene una visión distinta de las habilidades inferenciales de los agentes; una visión en la que es posible dar cuenta, con un mecanismo de términos, cómo los agentes derivan por deducción las consecuencias lógicas de lo que saben. En lógica de justificaciones, si bien los agentes mantienen un nivel de consciencia lógica sobre tautologías de lógica proposicional (consciencia justificada por supuesto), su conocimiento no está cerrado bajo consecuencia lógica. Por ende las habilidades inferenciales de los agentes no se vuelven triviales como es el caso en las lógicas modales de conocimiento y creencia.

el tamaño de las pruebas con las que el agente deriva el conocimiento que su teoría epistémica le adscribe.

El uso principal que se le da a este mecanismo es demostrar que todos los agentes de lógicas epistémicas y doxásicas son omniscientes, mientras que los agentes de lógicas de justificación no lo son (p. 14). De acuerdo con los autores “...*la teoría propuesta demuestra dónde buscar soluciones prácticas para el problema de Omnisciencia Lógica: utilice lenguajes de justificación... ..con anotaciones concisas sobre las razones que representarían testigos fiables del conocimiento*” (p. 21, trad.)<sup>64</sup>.

La unidad de medida para evaluar el tamaño de las pruebas se representa con símbolos del lenguaje formal. Por ejemplo, sea  $k_n$  un término que representa un costo de recursos con valor 1:

- i) Si  $\Phi \rightarrow \Psi$  y  $K_i\Psi$ , entonces  $k_c[k_1(\Phi \rightarrow \Psi) \rightarrow ([k_2] \Phi \rightarrow [k_1*k_2] \Psi)]$
- ii) El tamaño total de la prueba tiene medida  $k_c + k_1 + k_2 + [k_1*k_2]$ . Si el uso de AJ2 tiene para  $i$  un costo con valor  $x$ , entonces la medida de complejidad para  $K_i\Psi$  es el resultado del polinomio  $1 + 1 + 1 + x$ , que representa el costo de recursos que invierte para inferir  $\Phi$ .
- iii) Los términos de justificaciones cumplen esta función. Esto se hace a través del CS, estipulando que, por ejemplo,  $\forall c_n \in CS$ ,  $c_n$  tiene un costo de  $3n$ , o cualquier otro.

Este mecanismo, para identificar teorías omniscientes, se compone de dos pruebas, una débil y otra fuerte (p. 8, Artemov & Kuznets). Sea  $L$  una teoría formal epistémica,  $KL$  una lógica de  $L$ , y  $A$  un axioma de  $KL$ .  $L$  supera la prueba débil si un

---

<sup>64</sup> “*The proposed theory demonstrates where to look for practical solutions to the logical omniscience problem: find justification languages, perhaps model specific, with concise annotations of reasons that would yield feasible witnesses of knowledge.*” (cita original).

algoritmo no determinista computa una demostración de tamaño polinomial<sup>65</sup>  $\forall KA$  en  $L$ . Y  $L$  supera la prueba fuerte si  $\forall KA$  existe un algoritmo determinista que reconstruye su demostración en tiempo polinomial. Superar la prueba fuerte implica superar la débil.

La lógica de justificaciones supera la prueba fuerte por (iii). Las lógicas modales no lo hacen porque en su lenguaje no se refleja la dificultad de obtener conocimiento, ya que los operadores toman el papel de los términos y no son suficientemente expresivos para esta tarea (p. 7, Artemov & Kuznets). El problema es que la única forma de medir la complejidad sería contabilizando sobre los operadores  $K$  y  $B$ . En  $(K\Phi \& K\Psi) \rightarrow K(\Phi \& \Psi)$ , conocer  $\Phi$  tendría el mismo costo que conocer  $(\Phi \& \Psi)$ .

#### **4.4.2 Conocimiento como creencia verdadera justificada**

Dado que el conocimiento se define como *creencia verdadera* en lógicas epistémica y doxástica, lo único que separa al conocimiento de la creencia es la verdad. La ausencia de justificaciones en estos lenguajes ha causado una brecha entre el análisis formal del conocimiento y la epistemología, donde el conocimiento se ha definido tradicionalmente como *creencia verdadera justificada* (p. 1, Artemov).

En lógicas como  $JT_{45}$  las creencias llevan al conocimiento porque son verdaderas y están justificadas, esto es, la definición del conocimiento bajo la noción tripartita. Así es como se habilita el análisis epistémico a contra-ejemplos de conocimiento, donde hay creencias verdaderas sin una genuina justificación, o creencias justificadas que no alcanzan un estatus de verdad.

---

<sup>65</sup> En complejidad computacional, los problemas polinomiales no deterministas (NP) son aquellos para los que puede encontrarse una solución en tiempo polinomial, pero una que es muy difícil de encontrar (p. 267, Aaronson). En ellos, el algoritmo revisa distintas posibles soluciones hasta encontrarla. En cambio, los algoritmos deterministas siguen un solo camino para llegar a la solución.

La lógica de justificaciones es una herramienta de gran utilidad para analizar problemas epistemológicos en torno a la tradicional noción tripartita del conocimiento. Uno de los problemas que ha capturado gran atención de la filosofía surge con la pregunta *¿es el conocimiento creencia verdadera justificada?* Con base en esta pregunta han surgido múltiples contra-ejemplos a esta definición; situaciones donde se muestra que las creencias verdaderas justificadas no siempre implican la posesión de conocimiento.

Entre los contra-ejemplos propuestos, en los que se ataca al análisis del conocimiento bajo la noción tripartita, los más conocidos son los *Casos I y II* de Edmund Gettier, quien en 1963 publica *Is Justified True Belief Knowledge?*, texto seminal en epistemología donde el autor muestra que, en ambos casos, las condiciones de creencia, verdad y justificación no son necesarias ni suficientes para establecer que algo es conocimiento.

Otros autores han propuesto diversos contra-ejemplos. Uno de ellos, llamado *Granero Rojo*, es propuesto por Saul Kripke en 1980, que posteriormente es retomado por Dretske en 2005. Artemov (2008, Artemov-2) presenta la formulación de Dretske, y procede a analizarlo con lógica epistémica y lógica de justificaciones. Dados los resultados de cada análisis y su respectiva comparación, el autor sugiere lo siguiente:

- 1) A diferencia de la lógica epistémica, la lógica de justificaciones es más apta para analizar este tipo de contra-ejemplos.

- 2) Mediante un análisis con lógica de justificaciones emergen aspectos sobre la justificación que anteriormente se encontraban ocultos, los cuales, ofrecen una visión más amplia de los problemas tratados en los contra-ejemplos<sup>66</sup>.
- 3) Finalmente, no necesariamente es la noción tripartita la que falla en caracterizar el conocimiento, sino que hay aspectos de la justificación que se ignoran, y que surgen cuando se tiene una visión más amplia de los contra-ejemplos.

#### **4.4.2.1 Granero Rojo<sup>67</sup>**

Supóngase que un conductor llega a un vecindario donde hay graneros rojos de dos tipos; unos son reales, y otros son falsos porque están contruidos con papel mache. Nuestro conductor, desde lo lejos, observa uno de los graneros falsos, y sin embargo, su visión no es lo suficientemente competente como para darse cuenta que está hecho de papel mache. En consecuencia, el conductor cree que se trata de un genuino granero rojo.

---

<sup>66</sup> No está demás recordar al lector que la lógica de justificaciones, como toda lógica, no está exenta de problemas. Uno de estos problemas se origina en el contexto del estudio sobre la noción tripartita, y es tratado en Halpern et.al.-2 (2009). Los autores afirman que en lógica de justificaciones no es posible una total definición del operador modal de conocimiento K, en términos de la noción tripartita de creencia verdadera justificada (p. 12). Este problema no afecta el argumento central defendido en este trabajo (*la lógica de justificaciones ofrece herramientas con las que puede abordarse la Omnisciencia Lógica en un SD como Listener-Teller, y así estudiar aspectos ignorados por las lógicas S<sub>5</sub> y KD<sub>45</sub>*), pero no puede dejar de mencionarse.

Omitiendo los aspectos más técnicos, Halpern et.al.-2 utilizan como criterio de definición la expresión  $K \leftrightarrow J$ , donde  $Jp = t: p$ , y donde p es una creencia, y t es un término que satisface facticidad, tal que p es una creencia verdadera justificada. De ser  $K \leftrightarrow J$  el caso, entonces habría una definición total de K en términos de J. Sucede que en lógica de justificaciones no se cumple  $K \leftrightarrow J$ , pues mientras  $J \rightarrow K$  es válido en esta lógica,  $K \rightarrow J$  no lo es.

No hay por el momento una respuesta o solución a este problema, y es uno al que la lógica de justificaciones debe atender, si se defiende la idea que esta es una herramienta útil para analizar adecuadamente casos de conocimiento bajo la noción tripartita. Sin embargo, es adecuado mencionar algunos motivos a favor de la no validez de  $K \leftrightarrow J$  en lógica de justificaciones. En primer lugar, véase que la validez de  $J \rightarrow K$  refleja un hecho intuitivo del conocimiento: el conocimiento explícito (entendido en este caso como creencia verdadera justificada) es siempre subconjunto del conocimiento implícito (K). Pero por el contrario, lo inverso no puede sostenerse, dado que es una contingencia que el conocimiento explícito llegara a coincidir con el implícito, y por ello, al menos parece razonable que  $K \rightarrow J$  no sea válido.

<sup>67</sup> El ejemplo pertenece a Artemov (2008, pp. 486-488, Artemov-2).

#### 4.4.2.1.1 Análisis con lógica epistémica

Sea G “el objeto frente a mi es un granero”, R “el objeto frente a mi es rojo”, e i representa nuestro conductor. Desde una perspectiva en tercera persona se sabe que i no sabe que hay un granero rojo en el vecindario, por lo cual lo siguiente es verdadero:

i.  $\sim K_i G$

Sin embargo, con base en la evidencia que le proporciona su percepción, nuestro conductor dice estar seguro de que hay un granero rojo en el vecindario:

ii.  $K_i(G \ \& \ R)$

Si en realidad nuestro conductor supiera que hay un granero rojo en el vecindario, es decir que (ii) es el caso, entonces podría deducirse, que él sabe que hay un granero.

Sin embargo, como lo expresa (i), nuestro conductor en realidad no sabe que hay un granero puesto que el que ha visto es falso.

Véase la misma situación pero en términos formales. Las expresiones (i) y (ii) entran en conflicto, pues  $K_i(G \ \& \ R) \rightarrow \sim K_i G$  no es una fórmula válida de lógica epistémica. Por el contrario,  $K_i(G \ \& \ R) \rightarrow (K_i G \ \& \ K_i R)$  sí es una fórmula válida:

iii.  $K_i(G \ \& \ R) \rightarrow (K_i G \ \& \ K_i R)$

iv.  $K_i G \ \& \ K_i R$

Donde (iv) se sigue por *Modus Ponens* con (ii) y (iii). Finalmente por eliminación de la conjunción se obtiene lo siguiente:

v.  $K_i G$

Evidentemente, hay una contradicción con (i) y (v), pues se concluye que el conductor sabe que hay un granero, pero al mismo tiempo no lo sabe.

La moraleja de este análisis es que con lógica epistémica es posible observar que hay un problema en esta situación, la cual es causa de la contradicción obtenida, sin embargo, no es capaz de decirnos cuál es.

#### **4.4.2.1.2 Análisis con lógica de justificaciones**

En este análisis es posible representar explícitamente de cada fórmula, creída o conocida. En este caso, “u” es el término de justificación que representará la evidencia relevante para creer G, es decir, la evidencia perceptual de ver un genuino granero en el vecindario. Por otro lado, “v” representará al evidencia perceptual de ver un granero rojo. Nótese que “v” representa la evidencia en posesión de nuestro conductor, producto de haber observado un granero rojo de papel mache:

i. u: G

ii. v<sub>i</sub>: (G & R)

A continuación emergerá un aspecto anteriormente oculto: el conductor no tiene conocimiento, no a causa de que las condiciones de creencia, verdad y justificación no sean necesarias y suficientes del mismo, sino porque la justificación en su posesión no lleva al conocimiento sobre la existencia de un granero. Véase en términos formales:

iii. (G & R) → (G)

iv. c<sub>i</sub>: ((G & R) → (G))

v. v<sub>i</sub>: (G & R) → [a\*v]: G

vi. [c\*v]<sub>i</sub>: G

En la línea (iii) se expresa un axioma de lógica proposicional (el mismo del análisis anterior), y nuestro conductor, para quien se asume en virtud de nuestra de lógica que posee en su CS todos los axiomas y teoremas de lógica proposicional, conoce la

validez de  $(G \& R) \rightarrow (G)$  con un término constante “c” (iv). Como nuestro conductor también está justificado en creer  $(G \& R)$  en virtud del término “v”, puede ejecutar una acción inferencial de aplicación (\*), y deducir la información G, que está justificada con el término complejo “[c\*v]”.

Ahora asumase el axioma de facticidad, (T) t:  $\Phi \rightarrow \Phi$ . Esto significa que todas las justificaciones llevan a conocimiento. Como se sabe que “u” es la única pieza de evidencia que certificaría el conocimiento de G, y por el contrario, “[c\*v]” no lleva a conocimiento porque está compuesta de un fragmento “v” de evidencia falsa, se concluye lo siguiente:

vii.  $\sim([c*v]: G \rightarrow G)$ ; de “[c\*v]” no se sigue la verdad de G.

viii.  $(u: G \rightarrow G)$ ; de “u” se sigue la verdad de G.

A diferencia del análisis anterior, el sistema no explota a causa de una contradicción, sino que muestra que tanto la pieza de evidencia “[c\*v]” como “u” pueden co-existir, y además es claro cuál de las dos es evidencia genuina de G. Aún más, se ha revelado que no hubo conocimiento, no porque las condiciones de creencia, verdad y justificación no fueran suficientes y necesarias, sino porque no fueron del todo satisfechas, específicamente, la condición de justificación<sup>68</sup>.

#### **4.4.3 Conocimiento Potencial VS Conocimiento Actual**

Bentham y Quesada (2010) mencionan que para abordar el problema de Omnisciencia

Lógica, los esfuerzos no deben centrarse en el operado K, sino en la construcción de

---

<sup>68</sup> A continuación se menciona otro problema de la lógica de justificaciones en el contexto de la noción tripartita. A pesar de que la lógica de justificaciones ha sido una herramienta útil para evaluar casos de conocimiento, como el de Granero Rojo, la versión estándar de esta lógica, que es la que se ha presentado en este capítulo, aún presenta problemas de *gettierización*, lo cual significa que aún hay contra-ejemplos a la noción tripartita que no es posible analizar con su aparato formal. Sin embargo, actualmente existen propuestas que buscan enriquecer la lógica de justificaciones estándar con más nociones de evidencia, cuya introducción permite enfrentar su *gettierización*. Este tema será tratado con más detalle en el siguiente capítulo.

nociones explícitas de conocimiento (p. 6). Como se ha mencionado antes, esta modalidad semántica está diseñada para representar el conocimiento potencial en agentes omniscientes.

Para los autores, *“las propiedades fuertes de omnisciencia en los agentes no deben verse como idealizaciones estáticas, sino como el resultado de procesos dinámicos donde los agentes interactúan”* (p. 5, trad.)<sup>69</sup>. Para dar cuenta de cómo los agentes obtienen nueva información mediante estos procesos, deben establecerse las acciones de inferencia con las que se infiere nuevo conocimiento explícito, a partir del conocimiento explícito previamente obtenido. Los autores formalizan este problema de la siguiente manera (p. 6):

- $Ex (\Phi \rightarrow \Psi) \rightarrow (Ex \Phi \rightarrow [ ]Ex \Psi)$

Donde “Ex” es un operador de conocimiento explícito y “[ ]” es un espacio que debe llenarse con una acción de inferencia, que explique cómo el agente ha llegado al conocimiento explícito de  $\Psi$ . Así pues, para estudiar este tipo de acciones inferenciales es necesario primeramente establecer una noción explícita de conocimiento, para conocer las condiciones bajo las cuales se obtiene<sup>70</sup>.

Como se ha mostrado, en el caso de la lógica de justificaciones emerge una noción explícita de conocimiento y creencia en virtud de la introducción de los términos de justificación;  $\Psi$  es creído explícitamente si hay un término que respalde  $\Phi$  con evidencia relevante, y además  $\Psi$  es conocido explícitamente si el término que lo

---

<sup>69</sup> *“Strong omniscience properties of agents should be seen not as static idealizations, but as the result of dynamic processes that agents engage in.”* (cita original).

<sup>70</sup> En el texto los autores estudian acciones inferenciales para conocimiento explícito, empleando lógica dinámica epistémica, y una noción explícita de conocimiento construida con base en una función de consciencia (A):  $K(\Phi \& A\Phi) = Ex \Phi$ . Desafortunadamente, al igual que el resto de las lógicas mencionadas, no podrá ser estudiada en este trabajo.

justifica certifica la verdad de la fórmula. Para llenar el espacio “[ ]” se utilizan operaciones de aplicación y suma en términos de justificación, por ejemplo si  $t: (\Phi \rightarrow \Psi)$  y  $s: \Phi$ , entonces el agente conoce (o cree)  $\Psi$  porque ha llevado a cabo una acción inferencial de aplicación, combinando los términos  $t$  y  $s$ :  $[t*s] \Psi$ .<sup>71</sup>

Las nociones explícitas de conocimiento y creencia conforman la tercera ventaja de la lógica de justificaciones frente a las modales que se identifica en esta tesis, pues con ellas se posibilita un análisis enfocado en la capacidad deductiva del agente. En la sección 3.4 se mencionaron dos problemas relacionados a la noción potencial del conocimiento y creencia adscrita a los operadores  $K$  y  $B$ . Hay grandes razones para considerar que estos problemas no deberían surgir en esta lógica, en virtud de sus nociones explícitas de conocimiento y creencia<sup>72</sup>.

Se revisaron algunas de las propuestas más relevantes para abordar la Omnisciencia Lógica en  $S_5$  y  $KD_{45}$ . También se mostró que la lógica de justificaciones supera algunos de los problemas que estas propuestas presentan, así como sus contribuciones para dar los primeros pasos hacia una caracterización de agentes no omniscientes. En el siguiente capítulo se reflexionará sobre cómo con esta lógica podrían abordarse los problemas surgidos en el análisis epistémico de *Listener-Teller* con  $S_5$  y  $KD_{45}$ .

---

<sup>71</sup> Una manera diferente de expresar la idea de Benthem y Quesada, aunque con una clara preferencia por las lógicas de justificación, es la de Artemov y Kuznets. Los autores aseguran que la estrategia para superar las pruebas de Omnisciencia Lógica (ver sección 4.4.1) es utilizar “*sistemas que contienen tanto la noción justificada del conocimiento así como la modal; usando operadores epistémicos para la cognoscibilidad potencial y aseveraciones justificadas de conocimiento para ‘conocimiento real’...*” (p. 22, trad.)

“*Consider systems that contain both justified and the usual modal-style presentation of knowledge; use modal epistemic operators for ‘potential knowability’ and justified knowledge assertions for ‘real knowledge’...*” (cita original).

<sup>72</sup> Estas razones serán tratadas con mayor detalle durante el siguiente capítulo. Además, se propondrá que la introducción de nociones explícitas de conocimiento y creencia podrían ser de gran ayuda para abordar el problema de las acciones que dependen de la información de los agentes como Listener.

## **Capítulo 5. Primeros pasos hacia una análisis de *Listener-Teller* bajo una perspectiva de la Lógica de Justificaciones**

Se propone la lógica de justificaciones como herramienta formal para analizar *Listener-Teller* bajo una perspectiva de agentes no omniscientes. Posteriormente se reflexiona sobre cómo podrían enfrentarse con esta herramienta los problemas relativos a la Omnisciencia Lógica, identificados en el modelo epistémico-doxástico. Estos problemas son i) la falta de claridad en la noción de incognoscibilidad que surge del operador de conocimiento potencial, y ii) la incapacidad de representar acciones dependientes a la información. Se verá que ambos problemas están muy relacionados, por lo que enfrentar uno de ellos implica enfrentar el otro.

Después se presentan tres de los retos que aún debe superar la lógica de justificaciones, para que cumpla con mayor satisfacción las tareas de representación y análisis de sistemas artificiales como *Listener-Teller*. Estos son la representación de una noción epistémica de tiempo, el empleo de otras acciones respecto a la información, y la introducción de nociones más específicas de evidencia. Finalmente, se propondrán vías de investigación que podrían emprenderse para enfrentar estos retos con mayor competencia.

### **5.1. ¿Qué se sugiere para llevar a cabo un análisis de *Listener-Teller* bajo una perspectiva de agentes no omniscientes?**

Siguiendo el hilo argumentativo de este trabajo, el paso más natural es sugerir un análisis epistémico de *Listener-Teller* con las herramientas de la lógica de justificaciones. El carácter de este análisis es exploratorio; no se considerará que el

modelo propuesto sea una solución al problema de Omnisciencia Lógica que presenta el Sistema Distribuido bajo la perspectiva de la lógica epistémica y doxástica, sino más bien, los primeros pasos hacia una reflexión sobre aspectos de la información que pasan desapercibidos bajo la misma.

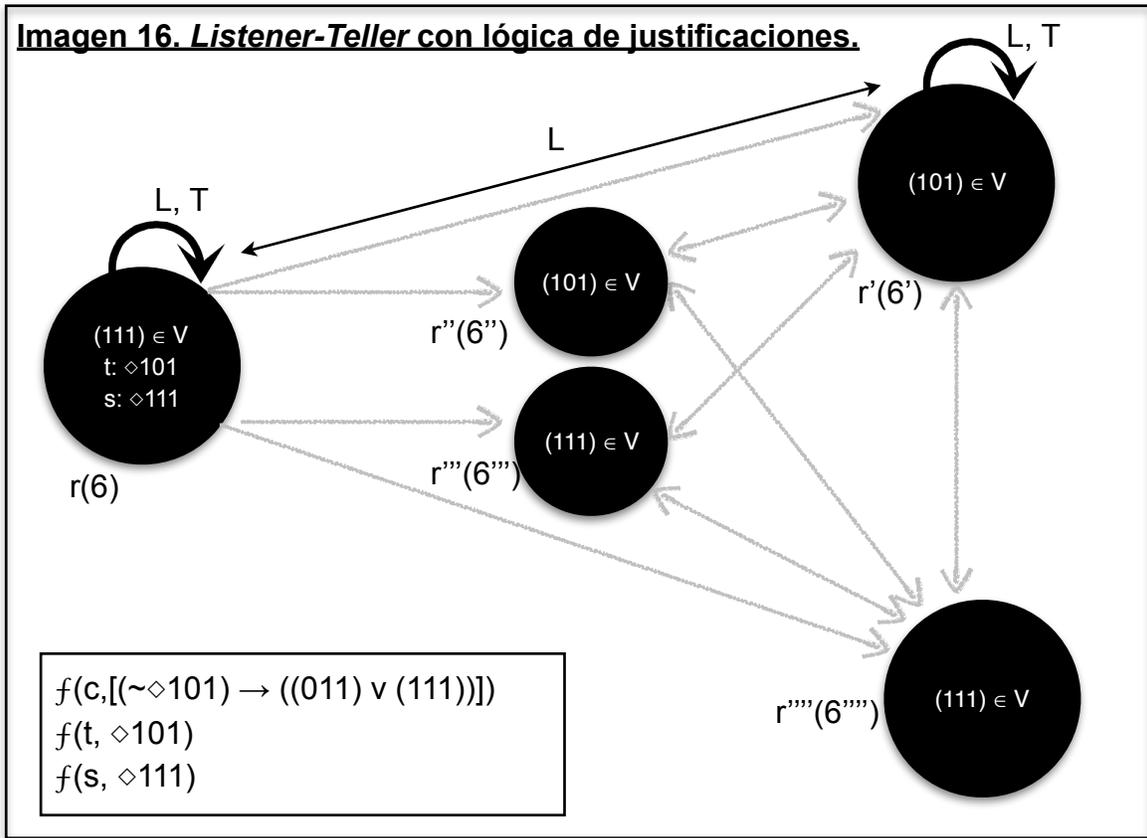
En términos formales, se propone un modelo semántico con marcos de Kripke-Fitting ( $M_F$ ) para razonar sobre la información implícita y explícita de Listener. Para la información implícita se mantienen los operadores  $K$  y  $B$ , mientras que la información explícita se trata de fórmulas con términos de justificación.

Sea  $CS_{L2}$  el Conjunto de Especificación de Listener cuando tiene dos espacios de memoria. Una vez más, Listener conoce el protocolo, y en este caso particular, conoce los axiomas de  $JT_{45} + D$ . Dado que  $CS_{L2}$  es axiomáticamente apropiado, Listener también conoce la regla  $R4$ . Supóngase que toda deducción con términos tiene un costo temporal de un punto del sistema, y que la complejidad se mide con relación al número de sub-fórmulas derivadas, como se muestra en la sección 4.4.1.

(Nótese que Listener podría modelarse con diferentes niveles de consciencia lógica e información inicial, dictaminando qué axiomas contendría su Conjunto de Especificación. El costo computacional de sus inferencias podría estipularse de la misma manera, dependiendo de sus habilidades inferenciales.).

La ilustración a continuación (Imagen 16) se trata del bosquejo de un modelo  $M_F$  para razonar simultáneamente sobre agentes omniscientes y aquellos con términos de justificación. Esta vez, las flechas negras representan las relaciones de acceso  $\rightarrow_K$  así como las relaciones  $\rightarrow_{JT45}$ , y las grises representan las relaciones  $\rightarrow_B$  y  $\rightarrow_{JD45}$ . Debajo del estado  $r_L(6)$  están las formulas que Listener está justificada a creer en el mismo.

Además de la adición de relaciones del tipo  $\rightarrow_{JT45}$  y  $\rightarrow_{JD45}$ , y la lista de fórmulas justificadas bajo la función  $f$ , el modelo es similar al de la Imagen 8 (Capítulo 2).



Antes de explicar con mayor detalle el modelo, nótese lo siguiente respecto a términos de justificación. Así como en lógica de justificaciones se establece, Listener está justificada, con términos constantes, en saber todas las tautologías proposicionales, y en este modelo en particular, hay tautologías del tipo  $[(\neg \diamond 101) \rightarrow ((011) \vee (111))]$  (la cual se explicará en el inciso ii). Esta clase de información está contenida en su CS, y se asume que Listener ya la sabe

Ahora bien, para la información que Teller le proporciona en cada ronda, Listener las justifica con términos variables y arbitrarios, pues no es información tautológica. Por ejemplo, si Teller le informa  $(\sim 111)$ , entonces hay un término variable  $t$  tal que  $t$ :

( $\sim\Diamond 111$ ). Véase también que Listener tiene creencias justificadas como t: ( $\Diamond 101$ ), esto es razonable si se observa que los términos como t, para el caso de fórmulas que representan creencias, son evidencia susceptible de contradecirse en otro momento, pues si Teller le informa ( $\sim 101$ ) habrá nueva evidencia, digamos r tal que r: ( $\sim\Diamond 101$ ), que refutaría la correspondiente al término t.

- Sea $(111) \in V$	<i>Hipótesis</i>
<b>(<math>r_L(6)</math>)</b>	
- ( $\sim\Diamond 101$ ) $\rightarrow$ ( $(011) \vee (111)$ )	<i>TAUT</i>
- c: [ $\sim\Diamond 101$ ] $\rightarrow$ ( $(011) \vee (111)$ )	<i>Internalización de TAUT por R4</i>
- (SI, $r_L(6)$ ) $\models$ <i>recibe</i> ( $\sim 101$ , T, L)	<i>Acto de Recepción</i>
- $\tau(r_L(6))$ ( <i>recibe</i> ( $\sim 101$ , T, L)) = ( $r_L(7)$ )	<i>Transición</i>
<b>(<math>r_L(7)</math>)</b>	
- u: ( $\sim 101$ )	<i>Justificación con término variable</i>
- c * u	<i>Acto Inferencial por Aplicación</i>
- $\tau(r_L(7))$ ([c*u]) = ( $r_L(8)$ )	<i>Transición</i>
<b>(<math>r_L(8)</math>)</b>	
- [c*u]: (111)	<i>Justificación con término compuesto</i>

- i) Sea c: [ $\sim\Diamond 101$ ]  $\rightarrow$  ( $(011) \vee (111)$ )  $\in$  CS<sub>1</sub>. Y sea también  $(111) \in V$  el caso, esto es, la verdadera la descripción del ambiente es (111). Los estados accesibles para Listener son indistinguibles para ella dado que en todos se cumple la incertidumbre sobre la verdad de las descripciones (111) y (101).
- ii) La fórmula justificada con “c” es tautológica por lo siguiente. (011) es la descripción verdadera cuando Teller informa ( $\sim 101$ ) a Listener durante la quinta ronda, pues como el protocolo lo indica, Teller no puede revelar directamente a Listener que (011) es verdadero, por lo cual, en lugar de que en la quinta ronda Teller le diga tal cosa, debe negar la descripción que le sigue de (011), en este caso (101), información cuyo envío originalmente debería ser durante la sexta ronda. O bien,

(111) es la descripción verdadera cuando Teller informa ( $\sim 101$ ) a Listener durante la sexta ronda para transitar a la séptima, pues en esta ronda no hay una descripción más que pueda ser verdadera. En este caso se cumple lo segundo; como la agente es sincrónica en  $R_2$  (sección 2.6.1.2) ella sabe que la ronda actual es la sexta, por lo que al recibir ( $\sim 101$ ) sabrá que  $(111) \in V$ .

- iii) Si  $(S_i, r_L(6)) \models \text{recibe}(\sim 101, T, L)$  entonces transita a  $r_L(7)$ , donde existe un término variable y arbitrario de justificación  $u$ , tal que  $(S_i, r_L(7)) \models u: (\sim 101)$ .
- iv) Luego  $(S_i, r_L(8)) \models [c^*u]: (111)$  por *A2* y *Modus Ponens*. El costo de esta inferencia (\*) es de un punto temporal, razón por la cual Listener transita a un estado  $r(8)$ .

A causa de sus dos espacios de memoria, podría pensarse que Listener solo puede inferir información cuya derivación le lleve solo dos momentos del tiempo. Quizá, sería trivial para ella derivar formulas con un mayor número de sub-fórmulas, ya que en tal proceso olvidaría la información anterior al transitar al siguiente estado, como sucede en el SDM formalizado con lógica epistémica y doxástica. Pero ello podría no suceder, en este caso, con la lógica de justificaciones, al menos suponiendo que los términos de justificación se comportan de una manera en particular.

Véase que, en  $S_5$ ,  $(S_i, r_L(7)) \models K_L(111)$  sería verdadero automáticamente por (G) y (K). Pero cuando su conocimiento está justificado con términos, puede, en  $(r, 8)$ , reconocer el origen de la evidencia que la respalda, como se observa con la expresión “[ $c^*u$ ]: (111)” en el inciso (iv), empleando lo que al parecer es un tercer bit de memoria.

Una posible explicación de este fenómeno, también susceptible de suceder en la corrida donde Listener tiene un solo espacio de memoria ( $R_1$ ), es que podría asumirse que el uso de justificaciones, a diferencia de los operadores modales, permiten

recuperar el conocimiento que pudo haberse olvidado por una insuficiencia en la retención de información. Sin embargo esto sucedería siempre y cuando Listener esté llevando a cabo una inferencia con estos términos, pues supóngase que el protocolo reanuda en la siguiente ronda, entonces ella olvidará “[c\*u]: (111)” en (r, 8) de no continuar usando esta información para otras inferencias.

Por razones evidentes, este fenómeno no se presentaría en una corrida donde Listener fuera una base de datos (BD). Sin embargo, al igual que una agente con capacidades limitadas de memoria, como sucede en el ejemplo anterior, BD también invertiría más puntos temporales del sistema para inferir consecuencias lógicas mediante derivaciones basadas en términos (considerando el mismo costo de recursos para las acciones inferenciales de aplicación y suma), a diferencia de los puntos que ella invierte en el modelo  $S_5$ -KD<sub>45</sub>.

*Pero ¿por qué se ha asumido que los términos derivados sobreviven a las transiciones?* No se dará una respuesta, y mucho menos una satisfactoria a esta pregunta. Lo que sí es posible hacer por el momento es reflexionar sobre tal fenómeno.

### **5.1.1 ¿Qué relación podría haber entre la evidencia y la memoria en Listener?**

En el modelo se asume que la información obtenida por actos de inferencia (\*, +) se olvida cuando el sistema transita de un punto al siguiente, lo cual ya es muy discutible, y da pie a un gran número de preguntas respecto a cómo se está entendiendo el carácter de la información. Además, queda una gran pregunta por abordar, a saber *¿es correcto asumir que los términos sobreviven a las transiciones del sistema?*

En el modelo propuesto se ha asumido que los términos producidos por actos de suma y aplicación sobreviven a las transiciones. Se adoptó el compromiso de que estos términos no tienen el mismo estatus que la información proposicional, porque se trata de registros de los actos inferenciales de Listener, que contienen codificada la información de las fórmulas que respaldan. La razón de esta decisión en el modelo es exploratoria, pues adoptar tal compromiso respecto a los términos lleva a discusiones interesante sobre otros aspectos epistémicos sobre la dinámica en la información en *Listener-Teller*, que de otra forma pasan desapercibidos con las lógicas  $S_5$  y  $KD_{45}$ .

Si los términos derivados por medio de aplicación y suma tienen el mismo estatus que las fórmulas, entonces lo más razonable es pensar que eventualmente Listener también los olvidará. Sin embargo, si se tratara de información con un estatus especial, podría pensarse que pertenecen a un conjunto distinto a las de las fórmulas, tal que pudieran ser almacenadas en un espacio de memoria distinto.

En palabras menos abstractas, puede imaginarse un detective muy olvidadizo quien registra toda la evidencia recolectada de un asesinato que investiga actualmente. Su pérdida de memoria no afecta la evidencia. Supóngase que durante la mañana ha trabajado en su oficina deduciendo posibles conclusiones sobre el arma usada en el homicidio, después de reflexionar sobre las distintas relaciones que puede haber entre las piezas de evidencia recogida en la escena del crimen, mismas que registra en una libreta. El detective regresa en la tarde, después de sufrir un episodio de pérdida de memoria, pero bien podría recordar las conclusiones a las que llegó durante la mañana, remitiéndose a la libreta donde figuran sus anotaciones.

Por otro lado, un caso distinto sería el de considerar que, si bien no comparten el mismo estatus, todos los términos y las fórmulas se encuentran en una relación, tal que lo que afecta a las fórmulas afecta a los términos, y viceversa. Este es un punto de vista muy factible atendiendo al hecho que en lógica de justificaciones hay una función de evidencia que relaciona términos con fórmulas. Supóngase que, lo que representa esta función es la capacidad mediante la cual Listener identifica la información que es respaldada por las piezas de evidencia que posee. Si esta capacidad es afectada por su pérdida de memoria, tal que olvida las relaciones que estableció previamente, entonces la agente no podría inferir que (111) es la verdadera descripción del mundo, aún si la evidencia fuera almacenada en un espacio distinto al de la información proposicional.

La pregunta por la relación que hay entre la memoria y la evidencia de Listener es solamente una de las diversas preguntas que surgen respecto a aspectos epistémicos que han pasado desapercibidos en un modelo epistémico-doxástico. Por el momento, esta discusión será concluida para continuar con la defensa de un modelo de lógica de justificaciones de *Listener-Teller*.

## **5.2 ¿Cómo podrían abordarse los dos problemas identificados en el modelo de *Listener-Teller* con Omnisciencia Lógica bajo la perspectiva que ofrece un modelo de justificaciones?**

Los problemas que se identificaron en el modelo epistémico-doxástico de *Listener-Teller* son dos: i) la falta de claridad en la noción de incognoscibilidad, producto de la negación del operador potencial de conocimiento ( $\sim K$ ), y ii) la incapacidad de

representar acciones que dependen de la información de agentes como Listener. Ambos problemas fueron brevemente mencionados en el contexto de la lógica de justificaciones durante la sección 4.4.3. Ahora se tratarán con mayor profundidad, especialmente el segundo.

Del primer problema se dijo que con la introducción de la noción de conocimiento explícito, por parte de la lógica de justificaciones, podrían tratarse casos donde no hay una idea clara sobre cómo interpretar la negación del operador  $K$ , atendiendo a la información explícita que el agente en cuestión posee. Hay quienes podrían argumentar contra esta propuesta, diciendo que el problema respecto a la interpretación de " $\sim K$ " no estaría resuelta, sino que solamente estaría evadiéndose con el uso de otras nociones. Esto es verdad, pero también es pertinente recordar que, como mencionan Benthem y Quesada, la mejor estrategia para abordar la Omnisciencia Lógica no es pretender hacer de  $K$  un operador de conocimiento explícito, sino construir nociones de conocimiento (y creencia) que sí lo sean.

Al haber una distinción entre conocimiento actual ( $t: \Psi$ ) y potencial ( $K\Psi$ ), resulta más intuitivo decir que una fórmula no es cognoscible cuando no existe un término que represente una demostración que se encuentre dentro de las capacidades inferenciales de Listener para deducirla. En consecuencia,  $\sim K_L \Psi$  podría mantener un significado estándar de Semántica de Kirpke, evitando así las confusiones que causa al adjudicarle una interpretación epistémica: *existe al menos un estado, accesible desde el estado presente de Listener, donde  $\Psi$  es falso*. Esta sería una de las posibles estrategias para evitar el problema respecto a la confusión originada en la interpretación de la negación del operador  $K$ .

Ahora bien, con la introducción de nociones explícitas de conocimiento y creencia surge la posibilidad de abordar el segundo problema identificado en el modelo epistémico-doxástico de *Listener-Teller*. Podría saberse el momento en el que un agente puede ejecutar una acción, cuando esta depende de su conocimiento. Por ejemplo, si “x” es una acción cuya ejecución depende de conocer  $\Psi$ , es posible pensar en modelos que representen situaciones donde *Listener ejecuta una acción x siempre que Listener está justificada en conocer  $\Psi$ , (t:  $\Psi$ )*.

En *Listener-Teller*, Listener no parece ejecutar acciones dependientes de su conocimiento, hasta el momento en el que se analiza bajo la lógica de justificaciones. Originalmente Listener parece ser una agente totalmente receptiva, pero bajo la perspectiva de justificaciones, lleva a cabo acciones inferenciales como la aplicación, y acciones basadas en evidencia, como la justificación de nueva información. Así pues, hay acciones que, en este contexto, son dependientes de la información que posee, por lo que la noción explícita de información “t:  $\Psi$ ” parece ser competente para saber cuándo éstas pueden ejecutarse. Es posible saber si Listener puede ejecutar una acción de aplicación para deducir la información (111), porque se sabe si está justificada, o no, en saber las fórmulas relevantes para llevar a cabo la inferencia.

Pero esta ventaja respecto a las acciones, en caso de que sea verdadera, lo sería solamente dentro del contexto de las justificaciones. De no adoptar tal modelo, no existen acciones dependientes de la información.

No hay duda respecto a esta afirmación, pero hay acciones en *Listener-Teller* cuya representación sería problemática, incluso sin considerar las acciones que introduce la lógica de justificaciones. Por ejemplo *¿qué puede decirse respecto a la*

*información que Teller posee sobre el mundo?* En el protocolo se asume, por simplicidad, que esta información ya está almacenada en su memoria, pero es fácil pensar en una versión más compleja donde Teller es *también Listener*, y recibe información de un tercer agente para comunicarla a otros agentes.

Para casos como estos los modelos de justificaciones resultarían de gran utilidad. Semánticamente, si la acción (*envía*,  $\Phi$ ) es ejecutable bajo la condición de que Teller conozca, o crea que  $\Phi$  es verdadero, entonces esta acción podría ser satisfecha solamente en los estados donde  $t: \Phi$  se cumpla para Teller. Habiendo satisfecho esta condición, puede entonces haber estados donde (*recibe*,  $\Phi$ ) es verdadero para Listener. Para representar en la semántica este tipo de acciones adecuadamente, es necesario que se consideren las condiciones y efectos de estas, tarea que supera a la lógica de justificaciones estándar. Por tal razón, durante la siguiente sección se sugerirá el estudio de acciones para agentes artificiales con una lógica de justificaciones más sofisticada.

### **5.3 ¿Qué retos tiene un modelo de justificaciones respecto al análisis y representación de *Listener-Teller* y qué vías de investigación se sugieren para enfrentarlos?**

Se analizan tres de los retos que enfrentaría un modelo de justificaciones, y se sugieren herramientas con las que podrían superarse. El primero es expresar con una lógica temporal la noción del tiempo en el que los agentes ejecutan acciones de inferencia, de manera que pueda diferenciarse del tiempo en el que corre el sistema. El segundo y el tercer reto se tratan, respectivamente, de estudiar otras acciones

inferenciales y nuevas nociones de evidencia. Para ambos se sugiere seguir recientes trabajos de lógica de justificaciones y de lógica dinámica epistémica.

### **5.3.1 Tiempo epistémico y tiempo del sistema**

Una de las muchas insuficiencias en el modelo de justificaciones anterior, es que la memoria de Listener se define a partir del reloj del sistema. Por ende, el costo temporal para la inferencia del conocimiento explícito obedece también a puntos temporales del sistema y no el tiempo que le llevaría a Listener inferir esta información. En aras de profundizar en esta dificultad que enfrenta la lógica de justificaciones respecto a Listener-Teller, se expondrá la lógica t-MEL.

En *Knowledge, Time, and the Problem of Logical Omniscience* Wang presenta una lógica epistémica temporal t-MEL (*timed-Modal Epistemic Logic*) para agentes no omniscientes, inspirada en la lógica de justificaciones. El tiempo, en este contexto, es un recurso epistémico empleado para la inferencia de nuevo conocimiento. En t-MEL se introducen superíndices a las sub-fórmulas para indicar el tiempo en el que son derivadas. (K) tiene la siguiente forma (p. 8):

- $K(\Phi \rightarrow \Psi)^i \rightarrow (K\Phi^j \rightarrow K\Psi^k)$ , donde  $i, j, k$  son diferentes momentos del tiempo, e  $i, j < k$ .

*Tiempo*, en el contexto de esta lógica, se entiende como un criterio para medir y discriminar el empleo de recursos epistémicos que el agente emplea, en términos de complejidad computacional, para derivar fórmulas.

Al igual que en lógica de justificaciones, se ha demostrado mediante el *Teorema de Realización* que los axiomas de las lógicas modales tienen también contrapartes en t-MEL (p. 4). De hecho,  $tS_4$  (de t-MEL) se define como la versión numérica de  $JT_4$ , y se han estudiado relaciones de correspondencia entre ambas (p. 5).

Su semántica es la de Kripke-Fitting, salvo algunas modificaciones, ya que se introduce una función sintáctica de consciencia lógica en la que se registra en cada estado, el tiempo que le ha llevado al agente inferir lo que conoce (p. 4). Esta función asocia fórmulas con números naturales, de manera que  $A(\Phi) = k$  significa que el agente está consciente de  $\Phi$  en tiempo  $k$ . Por ejemplo, si el agente conoce el antecedente de una implicación, usará una unidad de tiempo para ejecutar *Modus Ponens* (p. 6).

Una propuesta a desarrollar para el SDM *Listener Teller*, es introducir la noción temporal de t-MEL al modelo de justificaciones, con el objetivo es ganar expresividad para formular hechos sobre el empleo de recursos temporales para la inferencia de consecuencias lógicas respaldadas por evidencia relevante. En la sección de Anexos puede verse el ejemplo sobre una posible implementación de t-MEL (Imagen 17A, 17-B y 17-C) en el modelo de justificaciones previamente presentado. Dado que el ejemplo es solamente una adaptación *ad-hoc* de algunos componentes de t-MEL al modelo de justificaciones, se sugiere que se considere una ilustración con fines expositivos, y no un genuino modelo.

### **5.3.2 Otras acciones de carácter epistémico**

Las operaciones de aplicación y suma representan solamente dos de las acciones que agentes como Listener pueden ejecutar en un ambiente como *Listener-Teller*; hay un buen número de procedimientos disponibles para que este tipo de agentes administren la información que tienen. Su empleo, en un modelo de justificaciones, podría revelar fenómenos todavía desapercibidos, lo cual induciría a reflexiones interesantes sobre aspectos epistémicos de la información como se observa en la sección 5.1.1. En

general, la representación explícita de otras acciones permitiría un análisis más fino de los procesos dinámicos en Sistemas Distribuidos.

Actualmente, el uso de acciones provenientes de teorías epistémicas de la información, como la revisión de creencias y la lógica dinámica epistémica, han resultado exitosas. Su implementación ha iniciado con el desarrollo de lógicas epistémicas y doxásticas, donde se representan diversas actitudes epistémicas de agentes, entre ellas, actitudes no omniscientes respecto a la información.

Entre estos trabajos se encuentra el estudio de agentes *en términos de la información que poseen* (Nepomuceno, et.al., 2012), donde los autores utilizan el lenguaje de anuncios públicos de lógica dinámica epistémica para construir acciones inferenciales significativas para agentes omniscientes, así como para agentes no omniscientes. En el texto se demuestra además que estas acciones son paralelas a las operaciones respectivas que estudian en *revisión de creencias*<sup>73</sup>, por ejemplo, el anuncio que representa la acción de adición de información es equivalente a la operación de expansión (pp. 117-119).

---

<sup>73</sup> La revisión de creencias es mejor conocida como una de las teoría de cambio epistémico. Se trata de marco formal para representar y reflexionar sobre las condiciones y efectos que tienen las acciones epistémicas (expansión, contracción y revisión) en cuerpos de creencias. La revisión de creencias es campo fértil para profundizar en problemas epistémicos. Por ejemplo, una de sus líneas de investigación es la búsqueda de criterios que determinen las creencias que deben retractarse en determinadas situaciones (p. 2, Gärdenfors, 1992).

Quizá la acción epistémica que mejor representa esta teoría, dada su complejidad, es la revisión. Esta se trata de una acción combinada de una expansión y una contracción, y se usa para modificar bases de datos con información inconsistente. Formalmente, si la base cuenta con las fórmulas  $\{\sim q, p \rightarrow q\}$ , y ha llegado una nueva pieza de información  $p$ , la base puede expandirse añadiendo  $p$ , pero si este es el caso, ahora contendría información contradictoria, a saber,  $p$  y  $\sim p$ . Esta inconsistencia puede ser eliminada contrayendo la pieza de información  $(p \rightarrow q)$ , y así, ahora es posible que  $p$  forme parte de la base de datos. Como puede observarse, la revisión favorece a un modelo coherentista de la justificación, pues es útil para mantener la consistencia en un sistema de creencias; otras variantes de esta teoría favorecen un modelo fundacionista (p. 10).

Para modelar acciones de agentes no omniscientes se introducen dos nociones de información, explícita e implícita (p. 121), por ejemplo, la acción de inferencia deductiva, la cual es significativa solamente para este tipo de agentes (p. 123):

$$[+\chi] \text{Inf}_{\text{Ex}} \Phi \leftrightarrow \text{Inf}_{\text{Ex}} \Phi$$

Una instancia de propiedades de acciones para agentes no omniscientes es la expresión de arriba, la cual describe uno de los efectos que posee la acción deductiva “[+]”. En ella se dice que una deducción con la fórmula  $\chi$ , mediante la cual se infiere la información explícita  $\Phi$ , afecta solamente la información explícita. Esto representa un hecho intuitivo de la información: la información explícita es subconjunto de la implícita.

Utilizando la misma clasificación para distinguir entre tipos de información, Quesada (2014)<sup>74</sup> construye un modelo de lógica dinámica epistémica para representar creencias implícitas y explícitas. Con base en este modelo, el autor también reconstruye operaciones dinámicas de revisión de creencias, en este caso particular para creencias explícitas.

Recientemente, estas técnicas para construir acciones también han comenzado a estudiarse en el contexto de la lógica de justificaciones. Un desarrollo muy notable al respecto es la lógica de Baltag et.al. (Baltag et.al. 1, 2012), quienes proponen acciones dinámicas, esta vez respecto a la evidencia, tales como [t!] (*actualización del estado epistémico del agente con evidencia dura t, donde t proviene de una fuente infalible*) y [t↑] (*mejora del estado epistémico del agente con evidencia suave, proveniente de una fuente confiable mas no infalible*) (pp. 182-183).

### **5.3.3 Evidencia implícita y explícita**

---

<sup>74</sup> Velázquez-Quesada, Fernando. *Dynamic Epistemic Logic for Implicit and Explicit Beliefs* (pp. 107–140). En *Journal of Logic, Language and Information*, Vol. 23, No. 2 (junio, 2014).

La noción de evidencia, que surge de los términos de la lógica de justificaciones estándar, es aún muy general y abstracta como para representar actitudes epistémicas de agentes respecto a la evidencia. En este caso se sabe que los términos aportan evidencia relevante a la pieza de información que justifican, pero se ignora qué tipo de evidencia es (es buena o mala evidencia) o qué relación tiene respecto al agente, es decir, si se encuentra, o no, disponible para el mismo.

Acerca del carácter de la evidencia, Baltag y compañía (Baltag et.al. 1, 2012) expanden la lógica de justificaciones siguiendo una intuición sobre la justificación y la evidencia: *para que una pieza de información esté debidamente justificada, esta debe estar respaldada por buena evidencia y no cualquier evidencia*. Con esta idea en mente, los autores añaden una restricción a la definición semántica del conocimiento de la lógica de justificaciones (pp. 187-188).

En la lógica de justificaciones estándar, además de creer la fórmula en cuestión, el agente debe tener un término admisible “t” para ésta, pero para satisfacer la admisibilidad, cuando “t” es término compuesto, no se requiere que los términos que le constituyen sean también admisibles para las fórmulas que respaldan. En la versión de Baltag et.al. la noción de admisibilidad es mas restrictiva, en el sentido que el término “t” debe provenir de una cadena admisible de términos, donde cada término es admisible para la fórmula que respalda. En este sentido, los términos empleados en la derivación de nueva información explícita deben representan *buena* evidencia.

Los autores además enriquecen la lógica de justificaciones lo suficiente como para analizar contra-ejemplos a la noción tripartita del conocimiento, que a través de la versión estándar de esta lógica no pueden explicarse. Esto significa que el lenguaje de

Artemov y Fitting es *gettierizable*, i.e. a causa de su falta de expresividad, no es posible modelar situaciones donde se requiere identificar nociones epistémicas más finas. Para abordar el problema de la *gettierización* se identifican dos tipos de evidencia, la explícita y la implícita (p. 171).

Respecto al estado de la evidencia con relación al agente, el hecho que los autores identifiquen entre evidencia explícita e implícita permite modelar un aspecto que no se encuentra en la lógica de justificaciones estándar: la disponibilidad de la evidencia. En un trabajo posterior (2014, Baltag et.al. 2), se introduce una noción llamada calidad de la evidencia (*evidential goodness*): *t es buena evidencia actualmente disponible en un estado si y solo si el agente computó t y además, t es evidencia conclusiva en s* (p. 50).

Para modelar la noción de calidad de la evidencia se introduce una función de consciencia que liga términos de justificación con una criterio de disponibilidad, por lo que es posible que un término pueda no estar disponible para el agente. Así, para que un agente tenga evidencia explícita de  $\Phi$ , se requiere que 1) tenga evidencia implícita de  $\Phi$ , y 2) que  $t$  sea un termino que satisfaga la condición de calidad de la evidencia (E) (p. 54):  $t^e \Phi := t: \Phi \ \& \ E t$ .

En general, la introducción de acciones y nociones epistémicas más finas en la lógica de justificaciones, suponen poderosas herramientas con las que podrían estudiarse una diversidad de procesos y aspectos de la información en Sistemas Distribuidos como *Listener-Teller*, los cuales, podría decirse, han permanecido ocultos bajo la perspectiva de lógicas omniscientes como  $S_5$  y  $KD_{45}$ . Además, cabe la posibilidad de formalizar sistemas artificiales más complejos.

Supóngase un sistema en el que existe una corrida en la que Teller mantiene un nivel de incertidumbre acerca del mundo. En este tipo de situaciones, acciones como la mejora del estatus epistémico con evidencia falible, y nociones como la admisibilidad de la evidencia, serían de gran ayuda para representar y analizar agentes artificiales con bases de datos que contienen información derrotable.

Para concluir este trabajo, en el capítulo presente se propuso la lógica de justificaciones como herramienta formal para dar los primeros pasos hacia un modelo de *Listener-Teller* sin agentes omniscientes. Consecuentemente, sin defender una definición particular del conocimiento, se sugirió la posibilidad de abordar los problemas que presenta el análisis del SD con  $S_5$  y  $KD_{45}$ , bajo la perspectiva del conocimiento como *creencia verdadera justificada* que ofrece la lógica de justificaciones.

La lógica de justificaciones no está exenta de problemas, y aún queda mucho trabajo para enriquecer sus herramientas de análisis epistémico. Por estas razones se mencionaron algunos de los retos que tendría un modelo de *Listener-Teller* con justificaciones, así como las vías de investigación que pueden seguirse en cada caso.

## 6. Anexo

### 6.1 t-MEL en un modelo de justificaciones

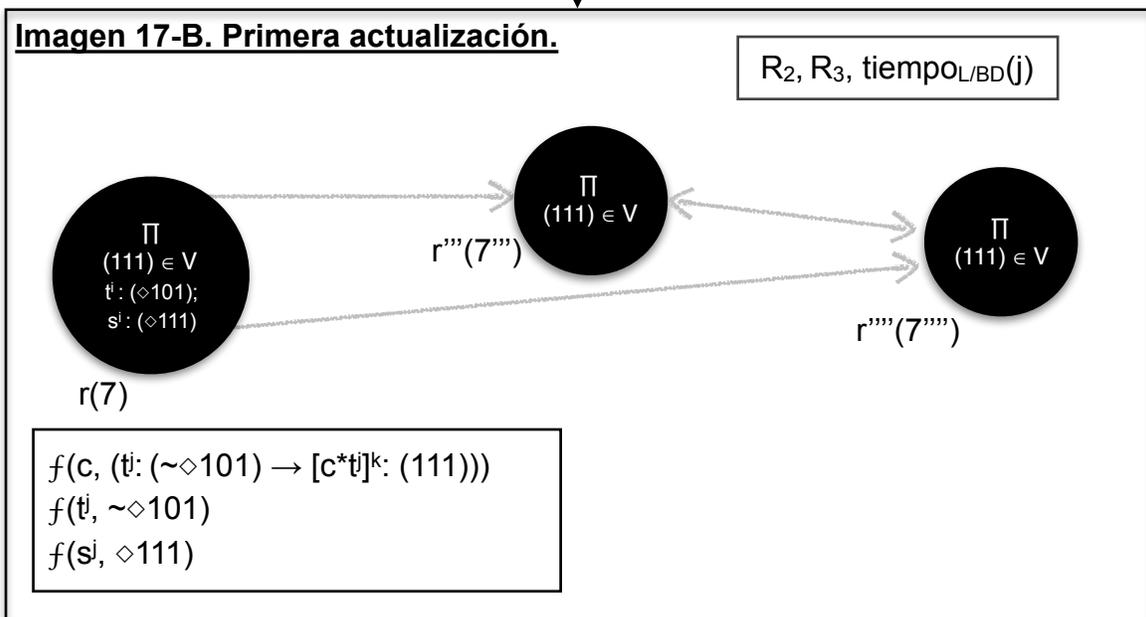
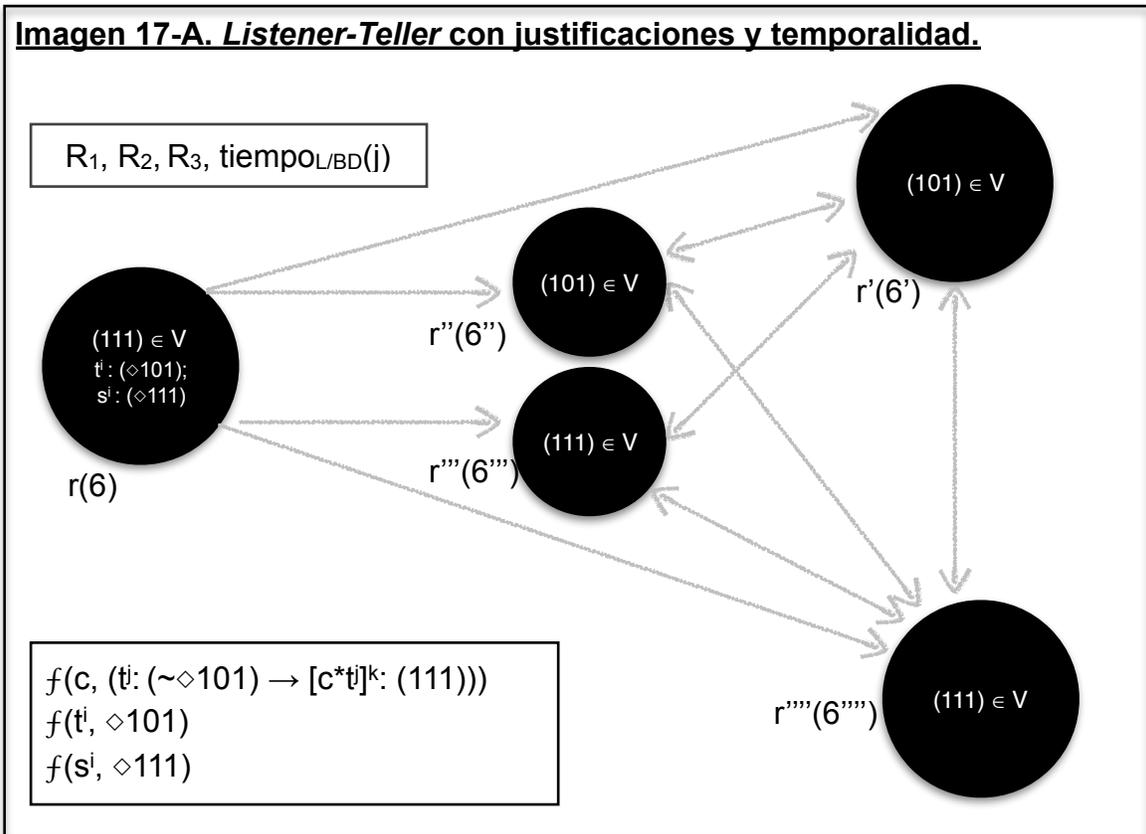
En este modelo Kripke-Fitting de  $JD_{45}$  se modela un ejemplo del SDM para las corridas  $R_1$ ,  $R_2$  y  $R_3$ . Con fines de claridad se han eliminado por completo las relaciones de acceso para  $S_5$  y  $KD_{45}$ , y solamente figuran las de justificaciones, sin embargo, estas relaciones pueden añadirse como en el modelo anterior. Con fines de sencillez, el ejemplo posee solamente relaciones  $JD_{45}$ . Al igual que en el modelo anterior, (111) es la verdadera descripción del ambiente.

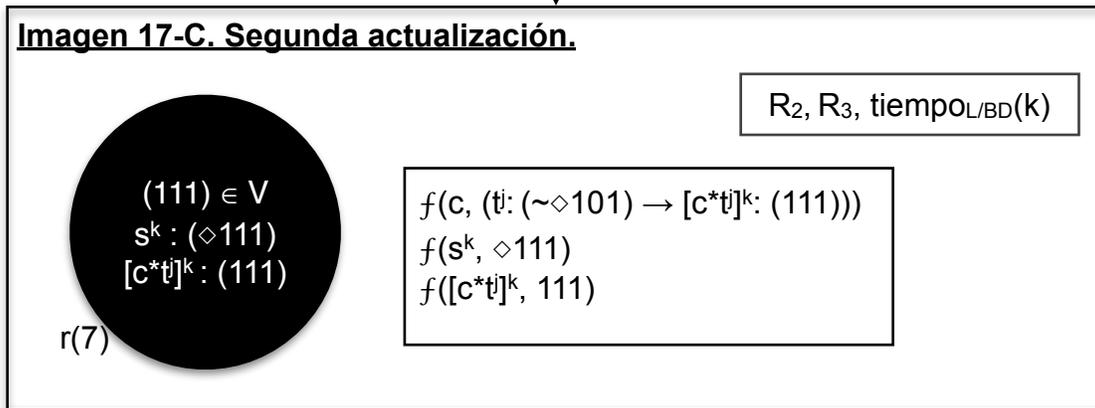
En el tiempo  $i$  el estado de conocimiento y creencias de Listener/BD es el mismo para los tres casos. Para el tiempo  $j$ , Listener de  $R_1$  ha olvidado lo que sabía en el tiempo anterior, dado que ha agotado el único espacio temporal del que se constituye su memoria al recibir la información de Teller. Por ello solamente se mantienen las corridas  $R_2$  y  $R_3$  para los tiempos posteriores, que son  $j$  y  $k$ .

Sean las unidades del tiempo epistémico de Listener ( $\text{tiempo}_{L/BD}(\cdot)$ ) expresadas con letras ( $i, j, k$ ). La memoria de Listener/BD se define con relación a este tiempo, invirtiendo una unidad cada vez que su información explícita se actualiza, o cada vez que realiza una inferencia del tamaño de una sub-fórmula (i.e. *Modus Ponens*, como sucede del  $\text{tiempo}_{L/BD}(j)$  al ( $k$ )). Nótese que la actualización de la información también se refleja en sus términos de justificación, por ejemplo, el término  $t$  del  $\text{tiempo}_{L/BD}(i)$  se actualiza en el  $\text{tiempo}_{L/BD}(j)$  para inferir ( $\sim\Diamond 101$ ), dado que la agente recibe ( $\sim 101$ ) (acción de recepción que se representa con la fórmula primitiva  $\square$ ).

La noción de *tiempo* respecto al sistema se define al igual que en el modelo  $S_5$ - $KD_{45}$  del 2º capítulo: como  $\forall(r, m), \forall(r, m+1). \tau(r, m) (\text{recibe}(\partial, T, L)) = (r, m+1)$ , el

sistema transita al siguiente punto siempre que Listener/BD recibe nueva información de Teller.





Las transiciones representadas en los modelos no tendrían sentido en lógica epistémica y doxástica. El conocimiento del mensaje ( $\sim 101$ ) llevaría a un modelo con un solo estado, donde (111) sería el caso.

En cambio, para este ejemplo se ha asumido que la agente Listener (y BD) requiere una actualización intermedia entre los modelos A y C. Tal idea resulta intuitiva si se piensa esta situación en términos de inferencia deductiva con subíndices temporales. En el tiempo  $\text{tiempo}_{L/BD}(i)$ , durante la sexta ronda del protocolo del sistema, Listener esta justificada en creer posibles las descripciones (111) y (101) con los términos  $t$  y  $s$  respectivamente. Así como en el modelo de la Imagen 16, en la Imagen 17-A, i.e. tiempo  $\text{tiempo}_{L/BD}(i)$ , Listener considera indistinguibles todos los estados a los que tiene acceso dado que en cada uno se satisface la incertidumbre sobre la verdad de (101) y (111).

En la misma ronda recibe el mensaje ( $\sim 101$ ) para transitar a la séptima ronda (Imagen 17-B), con lo cual invierte una unidad de tiempo para actualizar sus creencias, de manera que en el tiempo  $j$  cree justificadamente ( $\sim \diamond 101$ ) y ( $\diamond 111$ ), con los términos  $t$  y  $s$  actualizados al tiempo presente, es decir al tiempo  $j$ :  $t^j$ ,  $s^j$ . Los estados a los que la

agente tiene acceso le son indistinguibles porque en todos se satisface la recepción del mensaje ( $\sim 101$ ), i.e.  $\sqsupset$ , y en consecuencia también se satisface la imposibilidad de (101) así como la posibilidad de (111).

Finalmente, Listener invierte otra unidad de tiempo para ejecutar una acción inferencial de aplicación (\*) con los términos  $t^j$  y  $c$  (Imagen 17-C). Al efectuar esta inferencia la agente transita al tiempo  $k$ , y se concluye que Listener está justificada en creer (111) con el término compuesto  $[c^*t^j]$  en el tiempo  $k$ , formalmente,  $[c^*t^j]^k$ : (111).

En modelos de justificaciones y temporalidad, quizá sería posible establecer la satisfacción de las propiedades temporales de *sincronía* y *asíncrona* de dos maneras distintas, una con relación al tiempo del sistema y otra con relación al tiempo que a los agentes les lleva inferir información deductivamente. Sea *sincronía<sub>SD</sub>* la capacidad de percibir la transición del tiempo del sistema para cada punto  $(r, m)$ . Listener/BD es *sincronica<sub>SD</sub>* si identifica cada momento en el que Teller le comunica nueva información.

Más aún, en el ejemplo expuesto Listener es *asincrona<sub>SD</sub>* en  $R_1$ . No es necesariamente *sincronica<sub>SD</sub>* en  $R_2$  (ya que podría agotar sus dos espacios de memoria haciendo inferencias durante un solo estado, por lo cual, olvidará que Teller le dio información en tal punto del sistema). En cambio, sería suficiente que Listener fuera una base de datos ( $R_3$ ) para ser *sincronica<sub>SD</sub>*, ya que recordaría toda la información que recibiría de Teller.

Por otro lado, la sincronía con relación al tiempo epistémico podría ser definida con base en el tiempo epistémico de otros agentes, en caso de haberlos. Es intuitivo pensar que, de acuerdo a las capacidades que se tienen, cada agente ejecuta inferencias con distinta velocidad. En este sentido, Listener/BD sería *sincronica* con

relación a Teller si a éste también le lleva una unidad de tiempo recibir información o derivar una inferencia del tamaño de una sub-fórmula.

Nótese una diferencia y una similitud con relación al primero modelo de justificaciones. La diferencia es que ya no es necesario indicar que  $(\sim \diamond 101)$  debe cumplirse exactamente en  $(r, 7)$  para derivar  $(111)$ , pues con la expresión “ $t$ ” se especifica que durante el tiempo  $o_{L/BD}(j)$  es cuando el término  $t$  le es suficiente para estar justificada en saber  $(\sim \diamond 101)$ . La similitud es que, de continuar anidando términos de justificación para una inferencia, Listener podría rastrear la información que olvidó, en los casos  $R_1$  y  $R_2$ , atendiendo a los términos que representan las pruebas de fórmulas que está justificada en saber.

## Bibliografía:

- (1) Aaronson, Scott. *Why Philosophers Should Care about Computational Complexity* (pp. 261-327). En (Copeland, Posy & Shagrir eds.) Computability: Turing, Gödel, Church and Beyond (2013). MIT Press: Cambridge, Massachusetts.
- (2) Artemov, Sergei. (2008). *Why do we need justification logic?*. CUNY Academic Works TR-2008014. En [http://academicworks.cuny.edu/gc\\_cs\\_tr/319/](http://academicworks.cuny.edu/gc_cs_tr/319/). [En línea, recuperado el 10 de noviembre de 2015]
- (3) **[Artemov-2]** Artemov, Sergei. *The Logic of Justification* (477-513). En The Review of Symbolic Logic Vol.1, No. 4, (diciembre, 2008).
- (4) Artemov, Sergei & Kuznets, Roman. *Logical Omniscience As Infeasibility*. En Annals of Pure and Applied Logic Vol. 165, No. 1 (enero, 2014).
- (5) (Baltag.et.al.-1) Baltag, Alexandru; Baltag, Bryan & Semts, Sonja. *The Logic of Justified Belief Change, Soft Evidence and Defeasible Knowledge* (pp. 168-190). En Logic, Language, Information and Computation. 19th International Workshop, WoLLIC 2012, Buenos Aires, Argentina, September 3-6, 2012. Proceedings. Vol. 7456 de las series Lecture Notes in Computer Science.
- (6) (Baltag.et.al.-2) Baltag, Alexandru; Baltag, Bryan & Semts, Sonja. (2014). *The Logic of Justified Belief, Explicit Knowledge, and Conclusive Evidence* (pp. 49–81). En Annals of Pure and Applied Logic Vol. 165, No. 1 (enero, 2014).
- (7) Becutti, Marco; Franceschinis, Guiliana, et.al. (2015). *Modeling and Verification of Distributed Systems using Markov Decision Processes* (pp. 3-26). En Dario Bruneo y Salvatore Distefano (eds.) Quantitative Assessments of Distributed Systems. Wiley-Scrivener: Estados Unidos.
- (8) **[KRR]** Brachman J., Ronald & Levesque J., Hector. (2004). *Knowledge Representation and Reasoning* [3a edición]. Morgan Kaufmann Publishers: Estados Unidos.
  - (i) - Capítulos 1-4.
- (9) Dretske, Fred. *Is Knowledge Closed under Known Entailment?: The Case against Closure* [pp. 13-26]. En Steup, Matthias & Sosa, Ernest (eds.). *Contemporary Debates in Epistemology*. (2005). Wiley-Blackwell Publishing: Estados Unidos, Australia, Reino Unido.
- (10) Fagin, Ronald & Halpern, Joseph. (1994). *Reasoning About Knowledge and Probability*. Journal of the ACM, Vol. 41, num. 2. ACM: Nueva York.
- (11) **[Fitting1]** Fitting, Melvin. *Reasoning with Justifications* (p. 107-123). En Trends in Logic: Towards Mathematical Philosophy (2006). Springer: Países Bajos.

- (12) **[Fitting2]** Fitting, Melvin. *Justification Logics and Realization*. CUNY Academic Works TR-2014004. En [http://academicworks.cuny.edu/cgi/viewcontent.cgi?article=1394&context=gc\\_cs\\_tr](http://academicworks.cuny.edu/cgi/viewcontent.cgi?article=1394&context=gc_cs_tr). [En línea, recuperado el 15 de abril de 2016].
- (13) Gärdenfors, Peter. (1992). *Belief Revision*. Cambridge University Press: Reino Unido.  
 (i) - Belief Revision: An Introduction.
- (14) Gettier (1963). *Is Justified True Belief Knowledge?* En *Analysis* Vol. 23 No.6.
- (15) Halpern, Joseph. (2003). *Reasoning About Uncertainty*. The MIT Press: Londres<sup>75</sup>.  
 (i) - Chapter 2: Representing Uncertainty.  
 (ii) - Chapter 6: Multi-Agent Systems.
- (16) Halpern, Joseph, et.al. (2003). *Reasoning About Knowledge*. The MIT Press: Londres.  
 (i) - Chapter 1: Introduction and Overview.  
 (ii) - Chapter 4. Knowledge in Multi-Agent Systems.  
 (iii) - Chapter 5: Protocols and Programs.  
 (iv) - Chapter 9: Logical Omniscience.  
 (v) - Chapter 10: Knowledge and Computation.
- (17) **[Halpern, et.al.-2]** Halpern, Joseph, et.al. *Defining Knowledge in Terms of Belief: The Modal Logic Perspective* (pp. 469-487). En *The Review of Symbolic Logic* Vol. 2, No. 3 (septiembre, 2009).
- (18) Halpern, Joseph Y. & Puccella, Riccardo. *Dealing With Logical Omniscience* (pp. 220-235). En *Artificial Intelligence* Vol. 175, No. 1: John McCarthy's Legacy (enero, 2011). Elsevier: Estados Unidos<sup>76</sup>.
- (19) Hendricks, Vincent F. & Symons, John. *Where's the Bridge? Epistemology and Epistemic Logic* (pp. 137-167). En *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, Vol. 128, No. 1-8: Bridges between Mainstream and Formal Epistemology (marzo, 2006). Springer: Estados Unidos.
- (20) **[RWR]** Goertzel, Ben; Geisweiller, Nil (et al.). (2011). *Real World Reasoning: spatial, temporal and contextual inference under uncertainty*<sup>77</sup>. Atlantis Press: Paises Bajos.

<sup>75</sup> Esta es la primera edición en tapa blanda, y su contenido está actualizado hasta la fecha indicada. La edición original es de 1995.

<sup>76</sup> La versión citada pertenece al archivo digital de Joseph Halpern, ubicado en la página de la Universidad de Cornell. En <https://www.cs.cornell.edu/home/halpern/papers/pragmatics.ps>. [En línea, recuperado el 18 de diciembre de 2015].

<sup>77</sup> La edición citada es un borrador de la versión final del libro.

- (i) - Introduction.
- (21) Lacey, N. J. & Lee, M. H. *The Epistemological Foundations of Artificial Agents* (pp. 339-365). En Minds and Machines Vol. 13 (2003). Países Bajos: Kluwer Academic Publishers.
- (22) **[McCarthy1]** McCarthy, John. (1990). *Artificial Intelligence, Logic and Formalizing Common Sense*. En: <http://www-formal.stanford.edu/jmc/> [En línea, recuperado el 19 de diciembre de 2015].
- (23) **[McCarthy2]** McCarthy, John. (2007). *What is Artificial Intelligence?*. En: <http://www-formal.stanford.edu/jmc/whatisai.pdf> [En línea, recuperado el 9 de mayo de 2016].
- (24) Meyer, J.-J. Ch. *Epistemic Logic* (pp. 183-202). En Goble, Lou (Ed.). The Blackwell Guide to Philosophical Logic (2001). Blackwell Publishers: Reino Unido y E.U.A
- (25) Meyer, J.J.CH. y Hoek, W. van der. (1995). *Epistemic Logic for AI and Computer Science*. Cambridge University Press: Reino Unido.
- (26) Nepomuceno-Fernández, Ángel; Velázquez-Quesada, Fernando & Soler-Toscano, Fernando. *Dinámica de la información en agentes no omniscientes*. En Ensayos sobre lógica, lenguaje, mente y ciencia. Alfar: Sevilla.
- (27) Nilsson, Nils. *Logic and artificial intelligence* (pp. 31-56). En Artificial Intelligence Vol. 47, No. 1-3 (enero, 1991). Elsevier: Estados Unidos.
- (28) Pacuit, Eric. (2014). *Neighborhood Semantics for Modal Logic*. En: <http://web.pacuit.org/files/nbhd-v3-CH1.pdf>. [En línea, recuperado el 20 de mayo de 2016].
- (29) **[Parikh1]** Parikh, Rohit. *Logical Omniscience* (pp. 22-29). En Leviant, Daniel (ed.) Logic and Computational Complexity: international workshop ; selected papers / LCC \*94, Indianapolis, IN, USA, October 13 - 16, 1994. (1995). Springer: Nueva York, Berlín y Heidelberg.
- (30) **[Parikh2]** Parikh, Rohit. (2007). *Sentences, Propositions and Logical Omniscience, or What does Deduction tell us?*. En <http://www.sci.brooklyn.cuny.edu/cis/parikh/chicago.pdf>. [En línea, recuperado el 16 de octubre de 2015]. Dretske
- (31) Poole, David L., Mackworth, Alan K. (2010). *Artificial Intelligence: foundations of computational agents*. Cambridge University Press: Reino Unido.
- (i) - 13 Ontologies and Knowledge-Based Systems.

- (32) Puccella, Riccardo. *Deductive Algorithmic Knowledge* (pp. 287-309). En Journal of Logic and Computation Vol. 16, No. 2 (abril, 2006)<sup>78</sup>.
- (33) Russell, Stuart & Norving, Peter. (2010). *Artificial Intelligence: a modern approach*. Prentice Hall: Estados Unidos.  
 (i) - 7 Logical Agents.
- (34) Schiaffonati, Viola. *A Framework for the Foundation of the Philosophy of Artificial Intelligence* (pp. 537-552). En Minds and Machines Vol. 13 (2003). Países Bajos: Kluwer Academic Publishers
- (35) Sergot, Marek. (2008). *“Interpreted Systems” and Epistemic Logic*. Notas del curso impartido por Alessio Lomuscio: *Modal and Temporal Logic Grounded Semantics for Multi-agent Systems*. Department of Computing Imperial College: Londres.
- (36) Shoham, Yoav & Leyton-Brown, Kevin. (2010). *Multiagent Systems. Algorithmic, Game- Theoretic, and Logical Foundations*. De distribución libre, con permiso de los autores, en: <http://www.masfoundations.org/mas.pdf> Ubicado en [En línea, recuperado el 23 de marzo del 2015].
- (37) Sloman, Aaron. *Epistemology and Artificial Intelligence*. En Michie (ed.) Expert Systems in the Microelectronic Age (1979). Escocia: Edinburgh University Press.
- (38) Stalnaker, Robert. (1999). *Context and Content. Essays on Intentionality in Speech and Thought*. Oxford University Press: Nueva York.  
 (i) - 13 The Problem of Logical Omniscience I.  
 (ii) - 14 The Problem of Logical Omniscience II.
- (39) Thomason H., Richmond. *Logic and Artificial Intelligence* (pp. 848-902). En: Haaparanta, Leila (Comp.). The Development of Modern Logic. (2009). Oxford University Press: Nueva York.
- (40) van Benthem, Johan. (2010). *Modal Logic for Open Minds*. Center for the Study of Language and Information.  
 (i) - 12 Epistemic logic.  
 (ii) - 13 Doxastic and conditional logic.
- (41) van Benthem, Johan & Velázquez-Quesada, Fernando. The Logic of Awareness (pp. 5–27). Synthese Vol. 177, Supplement 1, (diciembre, 2010).
- (42) Wan, Wei. (2014). *Probabilistic and Epistemic Model Checking for Multi-Agent Systems*. Disertación Doctoral de Filosofía. University of Concordia: Québec.

---

<sup>78</sup> La versión usada se citó del archivo digital de la Universidad de Cornell. En <http://arxiv.org/pdf/cs/0405038v3.pdf>. [En línea, recuperado el 18 de diciembre de 2015].

- (43) Wang, Ren-June. *Knowledge, Time, and the Problem of Logical Omniscience* (pp. 1-18). En Fundamenta Informaticae Vol. XX (2010). IOS Press: Amsterdam.
- (44) Wooldridge, Michael. *Intelligent Agents* (pp. 27-77). En Weiss (ed.) Multiagent Systems. A Modern Approach to Distributed Artificial Intelligence (1999, tercera edición: 2001). Estados Unidos: MIT Press.