Quaderns de filosofia i ciència, 38, 2008, pp. 7-21

Lógicas Epistémicas y Sistemas Multiagentes: los agentes nacen y mueren

José P. Úbeda Rives
Jose.P.Ubeda@uv.es
Universidad de Valencia

Abstract

In this paper we study the propositional epistemic logics with operators to analize the knowledge of a group, and its applications to the multi-agent systems when the number of agents can vary between different worlds or states.

Keywords: Epistemic Logic, Multi-agent Systems.

1. Introducción

La atribución de conocimiento a los agentes y la consiguiente aplicación de las Lógicas Epistémicas a los sistemas multiagentes iniciada en 1984 por Lehmann [8] y Halpern & Moses [4][1] plantean una serie de problemas teóricos entre los que pueden citarse los siguientes:

- Definir formalmente los sistemas multiagentes *epistémicos* que consisten en un sistema multiagente junto con una descripción de cómo se asigna conocimiento a los agentes y/o grupos de agentes.
- Establecer rigurosamente las relaciones entre los sistemas multiagentes epistémicos y los marcos de Kripke sobre los que se basa la semántica más usual de las lógicas epistémicas. Es conocido que cada sistema multiagente epistémico puede identificarse con un marco de Kripke. Pero, ¿podemos asociar con cada marco de Kripke un sistema multiagente epistémico de forma que se conserve la verdad de los enunciados?
- Modificar los marcos de Kripke para tratar situaciones en las que cada mundo tiene asociado un conjunto de agentes, los que *viven* en él, y definir adecuadamente su semántica, lo que exige definir el conocimiento de un agente en un mundo en el que no existe.
- Modificar los marcos de Kripke para tratar situaciones en las que el grupo de agentes asociado con un nombre puede variar en los diferentes mundos, es decir, los nombres de los grupos no son necesariamente rígidos. Definir adecuadamente su semántica y establecer una definición uniforme de los operadores de conocimiento para grupos.[2]

En este artículo se presentan propuestas para los tres primeros problemas. En el apartado 2, después de una breve descripción de la sintaxis y semántica kripkeana de la lógica proposicional con operadores epistémicos para agentes y grupos de agentes, se establecen

[1]Estos últimos presentaron una versión previa en 1984.
[2]Véase una solución en [3].

las propiedades características de los operadores epistémicos de un grupo cuando éste es vacío o finito. En el apartado 3 se da una formalización de los sistemas multiagentes y los n-sistemas epistémicos que constituyen una semántica diferente para las lógicas epistémicas. Además, se establecen las relaciones entre los marcos de Kripke y los sistemas multiagentes y se estudian los sistemas multiagentes en los que el entorno no existe. En el apartado 4 se definen los marcos *vivos* de Kripke para tratar las situaciones en las que el conjunto de agentes puede variar de mundo a mundo y, se analizan diversas definiciones semánticas de los operadores de conocimiento. Finalmente en el apartado 5 se definen los sistemas multiagentes de forma que el número de agentes puede variar en el tiempo.

2. Lógica epistémica para un conjunto de agentes

El estudio formal de los conceptos relacionados con el conocimiento fueron iniciados por von Wright [12]. El primer estudio sistemático se debe a Hintikka [7] que analiza la modalidad epistémica "el agente a sabe o conoce que φ" (que simboliza como $K_a\varphi$) y otras modalidades asociadas.

El estudio de los aspectos sociales del conocimiento fue iniciado por D. Lewis [9] que introduce la noción de *conocimiento común* para explicar las convenciones sociales. Desde entonces se han analizado otros conceptos relacionados con el aspecto social del conocimiento o el conocimiento de un grupo. Si G es un grupo de agentes, los "conocimientos" de G más estudiados son:

a) "todos en G saben φ" (simbolizado por $E_G\varphi$);

b) "alguien en G sabe que φ" (simbolizado por $S_G\varphi$);

c) "φ es un conocimiento común en G" (simbolizado por $C_G\varphi$) que, intuitivamente, equivale a la conjunción infinita: $E_G\varphi \wedge E_G E_G\varphi \wedge \cdots$ y

d) "φ es un conocimiento distribuido en G" (simbolizada por $F_G\varphi$) que expresa que φ sería conocido por alguien que combina el conocimiento de todos los agentes en G.

Las lógicas epistémicas para analizar el conocimiento de cada agente perteneciente al conjunto $\mathcal{A}$ y de cada grupo G de agentes perteneciente a un conjunto $\mathcal{G}$ de subconjuntos[3] de $\mathcal{A}$ se construyen sobre el lenguaje $\mathbf{L}_{\mathcal{A},\mathcal{G}}$ cuyo vocabulario consta de un conjunto V (usualmente numerable) de *variables enunciativas*, de un conjunto funcionalmente completo de conectivas (por ejemplo, la conjunción $\wedge$ y negación $\neg$; el condicional $\rightarrow$, la disyunción $\vee$ y el bicondicional $\leftrightarrow$ se consideran definidas), los operadores K_a para $a \in \mathcal{A}$ y, para cada $G \in \mathcal{G} \subseteq \wp(\mathcal{A})$, los operadores E_G, C_G, S_G y F_G.

Un *enunciado* se define recursivamente como:

a) una variable enunciativa es un enunciado (*atómico*),

b) si φ y ψ son enunciados, entonces $\neg\varphi$, $(\varphi \wedge \psi)$ y, para cada $i \in \mathcal{A}$, $K_i\varphi$ son enunciados.

c) si φ es un enunciado, entonces, para cada $G \in \mathcal{G}$, $E_G\varphi$, $C_G\varphi$, $S_G\varphi$ y $F_G\varphi$ son enunciados.

Nótese que en caso de que el conjunto $\mathcal{A}$ de los agentes sea vacío y $\mathcal{G}$ no sea vacío (consta del conjunto $\emptyset$), se tienen los operadores $E_\emptyset$, $S_\emptyset$, $C_\emptyset$ y $F_\emptyset$, aunque no haya ningún operador K. Si $\mathcal{A}$ consta de n agentes y $\mathcal{G}$ es vacío tenemos el lenguaje de la lógica epístemica con n agentes. Si $\mathcal{A}$ y $\mathcal{G}$ son vacíos tenemos el lenguaje de la lógica proposicional.

La semántica de estas lógicas se basa en la idea de que un agente a sabe o conoce φ si y sólo si φ es verdadero en todos los mundos posibles para a; es decir, los que considera que pueden ser el mundo real en base a lo que sabe. Por ejemplo, con los enunciados p y q, diremos que a sabe $p \vee q$ si y sólo si a considera posibles

[3] $\mathcal{G} \subseteq \wp(\mathcal{A})$ donde $\wp(\mathcal{A})$ es el conjunto potencia o de las partes de $\mathcal{A}$.

a) mundos donde p y q son verdaderos;

b) mundos en los que p es verdadero y q es falso y

c) mundos en los que p es falso y q es verdadero.

Y no considera posibles los mundos donde p es falso y q es falso.

Formalmente, su semántica se basa en los *marcos de Kripke* sobre $\mathcal{A}$ y $\mathcal{G}$ que se definen como pares ordenados

$$\mathfrak{E} = \langle W, \mathcal{K} \rangle$$

tal que W es un conjunto no vacío de mundos y $\mathcal{K}$ es una función

$$\mathcal{K} : \mathcal{A} \longrightarrow W \times W$$

que a cada agente $a \in \mathcal{A}$ le hace corresponder una relación binaria $\mathcal{K}_a$ (denominada de *accesibilidad*) sobre el conjunto de los mundos W.[4]

Los *modelos de Kripke* sobre $\mathcal{A}$ y $\mathcal{G}$ se definen como pares ordenados

$$\langle \mathfrak{E}, \pi \rangle$$

donde $\mathfrak{E}$ es un marco de Kripke $\langle W, \mathcal{K} \rangle$ sobre $\mathcal{A}$ y $\mathcal{G}$ y π (*asignación de valores de verdad*) es una función

$$\pi : W \times V \longrightarrow \{\mathbf{t}, \mathbf{f}\},$$

que a cada mundo y variable enunciativa le asigna un valor de verdad o $\mathbf{t}$ (*verdadero*) o $\mathbf{f}$ (*falso*).

La relación $\mathfrak{M}, w \vDash \varphi$ (φ es *verdadero en el mundo* w del modelo $\mathfrak{M}$) se define por medio de las siguientes cláusulas, donde $a \in \mathcal{A}$ y $G \in \mathcal{G}$:

a) Si φ es una variable enunciativa, $(\mathfrak{M}, w) \vDash \varphi$ si y sólo si $\pi(w, \varphi) = \mathbf{t}$

b) $(\mathfrak{M}, w) \vDash \varphi \wedge \psi$ si y sólo si $(\mathfrak{M}, w) \vDash \varphi$ y $(\mathfrak{M}, w) \vDash \psi$.

c) $(\mathfrak{M}, w) \vDash \neg\varphi$ si y sólo si no se tiene $(\mathfrak{M}, w) \vDash \varphi$

d) $(\mathfrak{M}, w) \vDash K_a\varphi$ si y sólo si, para todo $s \in W$ tal que $w\mathcal{K}_a s$, $(\mathfrak{M}, s) \vDash \varphi$

e) $(\mathfrak{M}, w) \vDash E_G\varphi$ si y sólo si, para todo $a \in G$, $(\mathfrak{M}, w) \vDash K_a\varphi$

f) $(\mathfrak{M}, w) \vDash C_G\varphi$ si y sólo si, para todo $k > 0$, $(\mathfrak{M}, w) \vDash E_G^k\varphi$
donde E_G^k se define como sigue:

$$\begin{array}{rcl} E_G^0\varphi & = & \varphi \\ E_G^{k+1}\varphi & = & E_G(E_G^k\varphi). \end{array}$$

g) $(\mathfrak{M}, w) \vDash S_G\varphi$ si y sólo si hay un $a \in G$ tal que $(\mathfrak{M}, w) \vDash K_a\varphi$

h) $(\mathfrak{M}, w) \vDash F_G\varphi$ si y sólo si, para todo $s \in W$ tal que $\langle w, s \rangle \in \bigcap_{a \in G} \mathcal{K}_a$, $(\mathfrak{M}, s) \vDash \varphi$

A partir de esta relación se definen, en la forma usual:

a) el enunciado φ es *válido* en el modelo $\mathfrak{M}$, $\mathfrak{M} \vDash \varphi$, si y sólo si, para todo $w \in W$, $(\mathfrak{M}, w) \vDash \varphi$;

b) el enunciado φ es *válido* en el marco $\mathfrak{E}$, $\mathfrak{E} \vDash \varphi$, si y sólo si, para todo modelo $\mathfrak{M}$ sobre $\mathfrak{E}$ (es decir, tal que $\mathfrak{M} = (\mathfrak{E}, \pi)$ para una asignación π), $\mathfrak{M} \vDash \varphi$;

[4] Si $\mathcal{A} = \{1, \ldots, n\}$ un marco se define frecuentemente como el $(n+1)$-tuplo

$$\mathfrak{E} = \langle W, \mathcal{K}_1, \mathcal{K}_2, \ldots, \mathcal{K}_n \rangle .$$

Si $\mathcal{A}$ es vacío, $\mathcal{K}$ es la función vacía.

c) un enunciado φ es *válido* en la clase $\mathfrak{C}$ de marcos (modelos), $\mathfrak{C} \models \varphi$, si y sólo si, para todo marco (modelo) $\mathfrak{E} \in \mathfrak{C}$, $\mathfrak{E} \models \varphi$.

Las clases de marcos y modelos más estudiadas son las siguientes: la clase que consta de todos los marcos (modelos) y las clases que constan de marcos (modelos) en los que todas las relaciones de accesibilidad son reflexivas, simétricas, transitivas, euclídeas y/o seriales.[5] En especial y en relación con su aplicación a los sistemas multiagentes, ha merecido un estudio más detallado la clase de los marcos (modelos) cuyas relaciones de accesibilidad son relaciones de equivalencia.[6]

Sintácticamente y con relación a los operadores K_a se definen unas lógicas entre las que destaca $\mathbf{K}_{\mathcal{A}}$, la "menor" de todas, basada en los axiomas:

A1. Todas las tautologías proposicionales

A2. $K_a\varphi \wedge K_a(\varphi \to \psi) \to K_a\psi$, para $a \in \mathcal{A}$

y las siguientes reglas de inferencia, donde $a \in \mathcal{A}$:

MP. De φ y $\varphi \to \psi$, infiérase ψ (*Modus ponens*); ψ se obtiene por **MP** de φ y $\varphi \to \psi$;

N. De φ, infiérase $K_a\varphi$ (*Necesitación*); $K_a\varphi$ se obtiene por **N** de φ.

Otras lógicas se obtienen de $\mathbf{K}_{\mathcal{A}}$ al añadir nuevos axiomas (éstas se denotan por **K** seguida de los nombres de los axiomas y el subíndice $\mathcal{A}$) y/u otras reglas de inferencia. Los axiomas más usuales tienen las siguientes formas, donde $a \in \mathcal{A}$ y $\perp$ es lo *falso*.

T. $K_a\varphi \to \varphi$

4. $K_a\varphi \to K_aK_a\varphi$

5. $\neg K_a\varphi \to K_a\neg K_a\varphi$

D. $\neg K_a\perp$, que es equivalente en $\mathbf{K}_n$ a $K_a\varphi \to \neg K_a\neg\varphi$.

B. $\varphi \to K_a\neg K_a\neg\varphi$.

Si **L** es cualquier lógica, se dice que un enunciado φ es *derivable* en **L**, $\vdash_{\mathbf{L}} \varphi$, si y sólo si existe una sucesión de enunciados

$$\psi_1, \ldots, \psi_m,$$

tal que φ es ψ_m y, para cada $j = 1, \ldots, m$, ψ_j es un axioma o se obtiene de enunciados anteriores en la sucesión por **MP** o por **N** (u otra regla de **L**).

Las relaciones entre la semántica y sintaxis se establece en los siguientes teoremas (véase Chellas [1] para el caso de un agente):

Teorema 1 *Si $\mathfrak{C}$ es la clase de todos los marcos o todos los modelos (reflexivos, transitivos, euclídeos, seriales, simétricos, de equivalencia), entonces, para todo enunciado φ, $\vdash_{\mathbf{K}_{\mathcal{A}}} \varphi$ ($\vdash_{\mathbf{KT}_{\mathcal{A}}} \varphi$, $\vdash_{\mathbf{K4}_{\mathcal{A}}} \varphi$, $\vdash_{\mathbf{K5}_{\mathcal{A}}} \varphi$, $\vdash_{\mathbf{KD}_{\mathcal{A}}} \varphi$, $\vdash_{\mathbf{KB}_{\mathcal{A}}} \varphi$, $\vdash_{\mathbf{KT45}_{\mathcal{A}}} \varphi$, respectivamente) si y sólo si $\mathfrak{C} \models \varphi$.*

Teorema 2 *Un marco $\mathfrak{E}$ es reflexivo (transitivo, euclídeo, serial, simétrico, de equivalencia) si y sólo si en $\mathfrak{E}$ son válidos todos los enunciados derivables en $\boldsymbol{KT}_{\mathcal{A}}$ ($\boldsymbol{K4}_{\mathcal{A}}$, $\boldsymbol{K5}_{\mathcal{A}}$, $\boldsymbol{KD}_{\mathcal{A}}$, $\boldsymbol{KB}_{\mathcal{A}}$, $\boldsymbol{KT45}_{\mathcal{A}}$, respectivamente).*

[5] Estos tipos de relaciones se definen como sigue:

- $\mathcal{K}_a$ es *reflexiva* sobre W si y sólo si $\forall w(w \in W \to w\mathcal{K}_a w)$.
- $\mathcal{K}_a$ es *simétrica* sobre W si y sólo si $\forall x\forall y(x \in W \wedge y \in W \wedge x\mathcal{K}_a y \to y\mathcal{K}_a x)$.
- $\mathcal{K}_a$ es *transitiva* sobre W si y sólo si $\forall x\forall y\forall x(x \in W \wedge y \in W \wedge z \in W \wedge x\mathcal{K}_a y \wedge y\mathcal{K}_a z \to x\mathcal{K}_a z)$.
- $\mathcal{K}_a$ es *euclídea* sobre W si y sólo si $\forall x\forall y\forall z(x \in W \wedge y \in W \wedge z \in W \wedge x\mathcal{K}_a y \wedge x\mathcal{K}_a z \to y\mathcal{K}_a z)$.
- $\mathcal{K}_a$ es *serial* sobre W si y sólo si $\forall x(x \in W \to \exists w(w \in W \wedge x\mathcal{K}_a w))$.

[6] Es decir, son reflexivas, simétricas y transitivas. O reflexivas y euclídeas. O simétricas, transitivas y seriales. (Véase para estas equivalencias Fagin y otros [2].)

El teorema 2 no vale si sustituimos “marco” por “modelo”, ya que hay modelos en los que son válidos los enunciados derivables en $\mathbf{KT}_{\mathcal{A}}$ y no son reflexivos. Análogamente para los otros casos.

Si K_a satisface **T**, **4** y **5** (**D** y **B** son derivables), se dice que K_a es un operador de *conocimiento* (por **T**) *introspectivo* tanto *positiva* (por **4**) como *negativamente* (por **5**). La lógica $\mathbf{KT45}_{\mathcal{A}}$ se denomina $\mathbf{S5}_{\mathcal{A}}$.

Respecto a los operadores de conocimiento “social” distinguimos tres casos según G sea vacío, finito o infinito. Si G es vacío es fácil deducir las siguientes consecuencias:

- $E_\emptyset\varphi$ es verdadero para cada enunciado φ en cada mundo w de cada modelo $\mathfrak{M}$, ya que $a \in \emptyset$ es falso;
- $C_\emptyset\varphi$ es verdadero para cada enunciado φ en cada mundo w de cada modelo $\mathfrak{M}$, ya que $E^k_\emptyset\varphi$ es verdadero para cada $k > 0$;
- $S_\emptyset\varphi$ es falso para cada enunciado φ en cada mundo w de cada modelo $\mathfrak{M}$, ya que $a \in \emptyset$ es falso;
- $F_\emptyset\varphi$ es verdadero en un mundo w de un modelo $\mathfrak{M}$ si y sólo si φ es verdadero en cada mundo de $\mathfrak{M}$, ya que $\bigcap_{a\in\emptyset} \mathcal{K}_a$ es $W \times W$ donde W es el conjunto de mundos de $\mathfrak{M}$.

Así, $E_\emptyset$, $C_\emptyset$ y $S_\emptyset$ cumplen los siguientes axiomas: $E_\emptyset\varphi$, $C_\emptyset\varphi$ y $\neg S_\emptyset\varphi$. $F_\emptyset$ cumple los axiomas de la lógica **S5**. Así, la lógica de los operadores $E_\emptyset$ y $C_\emptyset$, que es la misma que la de la conectiva monádica que siempre da como valor lo verdadero, satisfacen trivialmente axiomas de la forma **A1**, **A2**, **4**, **5**, **B** y las reglas **MP** y **N**, pero no los axiomas **T** y **D** (cuando en los axiomas se sustituye K_a por $E_\emptyset$ o $C_\emptyset$). La lógica de $S_\emptyset$, que es la misma que la de la conectiva monádica que siempre da como valor lo falso, satisface **A1**, **A2**, **T**, **4** y **D** y la regla **MP**, pero no satisface **5** ni **B** ni la regla **N**, aunque si la regla:

M. De $\varphi \to \chi$ infiérase $S_\emptyset\varphi \to S_\emptyset\chi$ (*Monotonía*)

De las cláusulas para definir la verdad de E_G y S_G, se sigue, si $G = \{a_1, \ldots, a_m\}$ es finito y no vacío, que todo lo expresable con los operadores E_G y S_G es expresable sin ellos

$$E_G\varphi \leftrightarrow K_{a_1}\varphi \wedge \cdots \wedge K_{a_m}\varphi,$$

$$S_G\varphi \leftrightarrow K_{a_1}\varphi \vee \cdots \vee K_{a_m}\varphi.$$

Ya que entonces las cláusulas para definir la verdad de E_G y S_G pueden tomar la forma:

- $(\mathfrak{M}, w) \models E_G\varphi$ si y sólo si $(\mathfrak{M}, w) \models K_{a_1}\varphi \wedge \cdots \wedge K_{a_m}\varphi$
- $(\mathfrak{M}, w) \models E_G\varphi$ si y sólo si $(\mathfrak{M}, w) \models K_{a_1}\varphi \vee \cdots \vee K_{a_m}\varphi$

Esto no vale si existe un número infinito de agentes (véase Halpern & Shore [6]) ya que habría infinitos términos en la conjunción o disyunción. Los operadores C_G y F_G, aunque haya un número finito de agentes, no son definibles a partir de los operadores K_a.

Sintácticamente, el significado de estos operadores se especifica añadiendo a los axiomas y reglas relativos a las conectivas y los operadores K_a, nuevos axiomas y reglas. Para cada $G = \{a_1, \ldots, a_m\}$ no vacío se añaden los siguientes axiomas:[7]

A7. $E_G\varphi \leftrightarrow K_{a_1}\varphi \wedge \cdots \wedge K_{a_m}\varphi$

A8. $C_G\varphi \to E_G(\varphi \wedge C_G\varphi)$

A9. $S_G\varphi \leftrightarrow K_{a_1}\varphi \vee \cdots \vee K_{a_m}\varphi$

A10. $F_{\{a\}} \leftrightarrow K_a$, para $a \in \mathcal{A}$

[7] Véase Fagin & otros [2] para los axiomas relativos a E_G, C_G y F_G y la demostración de su adecuación y completud.

A11. $F_{G'}\varphi \to F_G\varphi$, si $G' \subseteq G$

A2$_F$. $F_G\varphi \wedge F_G(\varphi \to \psi) \to F_G\psi$

C. De $\varphi \to E_G(\psi \wedge \varphi)$, infiérase $\varphi \to C_G\psi$.

Si la lógica a la que se añaden es $\mathbf{KT}_{\mathcal{A}}$ ($\mathbf{K4}_{\mathcal{A}}$, $\mathbf{K5}_{\mathcal{A}}$, $\mathbf{KB}_{\mathcal{A}}$) o $\mathbf{S5}_{\mathcal{A}}$ debe añadirse el axioma **T** (**4**, **5**, **B**) o **T**, **4** y **5** sustituyendo K_a por F_G.

De **A10**, **A2**$_F$ y, posiblemente, otros axiomas para F_G, se deduce que F_G es un operador modal cuya relación de accesibilidad es

$$\mathcal{K}_{a_1} \cap \cdots \cap \mathcal{K}_{a_m}.$$

Respecto al operador C_G puede demostrarse que es un operador modal cuya relación de accesibilidad es el cierre transitivo de

$$\mathcal{K}_{a_1} \cup \cdots \cup \mathcal{K}_{a_m}.$$

Las lógicas de C_G y F_G con G no vacío y finito dependen de la lógica de los operadores K_a, para $a \in G$. En todos los casos, al ser transitiva $\mathcal{C}_G$, C_G cumple el axioma **4**; además, si los operadores K_a, para $a \in G$, cumplen el axioma **D** (**T**, **B**), entonces C_G cumple **D** (**T**, **B**); sin embargo, puede suceder que los K_a cumplan **5**, pero C_G no. Por ejemplo, si $W = \{1,2,3\}$, $\mathcal{K}_1 = \{\langle 1,2\rangle, \langle 2,2\rangle\}$ y $\mathcal{K}_2 = \{\langle 1,3\rangle, \langle 3,3\rangle\}$ (Figura 1), $\mathcal{C}_{\{1,2\}} = \{\langle 1,2\rangle, \langle 1,3\rangle, \langle 2,2\rangle, \langle 3,3\rangle\}$ (Figura 2) no es euclídea y no cumple **5**. Para F_G, si los K_a, para $a \in G$, cumplen el axioma **T** (**4**, **5**, **B**), entonces F_G cumple **T** (**4**, **5**, **B**); sin embargo, puede suceder que los K_a cumplan **D**, pero F_G no. Por ejemplo, si $W = \{1,2\}$, $\mathcal{K}_1 = \{\langle 1,1\rangle, \langle 2,2\rangle\}$ y $\mathcal{K}_2 = \{\langle 1,2\rangle, \langle 2,1\rangle\}$ (Figura 3), $\mathcal{F}_{\{1,2\}} = \emptyset$ (Figura 4) no es serial y no cumple **D**.

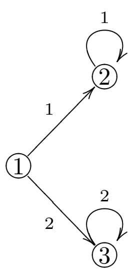

Figura 1

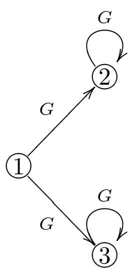

Figura 2

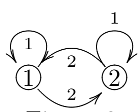

Figura 3

Figura 4

Además deben añadirse los axiomas y reglas que hacen referencia al caso de que G sea vacío:

$\mathbf{A}_{E\emptyset}$. $E_\emptyset\varphi$

$\mathbf{A}_{S\emptyset}$. $\neg S_\emptyset\varphi$

$\mathbf{A}_{C\emptyset}$. $C_\emptyset\varphi$,

los axiomas **A2**, **T**, **4**, **5** aplicados a $F_\emptyset$:

$\mathbf{A2}_{F_\emptyset}$. $F_\emptyset\varphi \wedge F_\emptyset(\varphi \to \psi) \to F_\emptyset\psi$

$\mathbf{T}_{F_\emptyset}$. $F_\emptyset\varphi \to \varphi$

$\mathbf{4}_{F_\emptyset}$. $F_\emptyset\varphi \to F_\emptyset F_\emptyset\varphi$

$\mathbf{5}_{F_\emptyset}$. $\neg F_\emptyset\varphi \to F_\emptyset\neg F_\emptyset\varphi$

y la regla de necesitación para $F_\emptyset$:

$\mathbf{N}_{F_\emptyset}$ De φ, infiérase $F_\emptyset\varphi$.

Para $\mathcal{A}$ infinito y los $G \in \mathcal{G}$ infinitos, véase Halpern & Shore [6].

3. Sistemas multiagentes

Para especificar y analizar los sistemas multiagentes se han propuesto diversos modelos formales. Fagin & otros [2] analizan un modelo que intenta dar cuenta de situaciones como la de los "niños con la frente sucia", la computación distribuida entre varios procesadores, los protocolos de comunicación, etc. Su modelo presupone que

(1) los componentes de un sistema se catalogan, según los intereses del que cataloga, como *agentes* o como *entorno*; el entorno es único, mientras que el número de agentes es 1 o más; si el número de agentes es n tenemos un *n-sistema*;

(2) el entorno y cada uno de los agentes se caracterizan por medio de *estados*;

(3) en cada instante, el n-sistema con el conjunto de agentes $\mathcal{A} = \{a_1, \ldots, a_n\}$, está en un *estado global* que es un tuplo $s = (s_0, s_1, ..., s_n)$, donde s_i es el estado del agente a_i, su *estado local*, que incorpora toda la información que le es accesible, y s_0 es el estado del entorno, que incorpora toda la información relevante no incluida en los estados de los agentes;

(4) la descripción de cómo cambia el estado global con el tiempo se hace por medio de funciones, denominadas *procesos*, que a cada instante le hace corresponder un estado global;

(5) un *sistema* se define como un conjunto no vacío de procesos, cada uno de los cuales es una posible evolución del sistema.

Con estos presupuestos, la descripción formal de un n-sistema es la siguiente.

(1) Sea $\mathbf{L}_0$ el conjunto de los estados posibles del entorno y, para cada $i = 1, 2, ..., n$, $\mathbf{L}_i$ el conjunto de estados posibles del agente a_i.

(2) El conjunto de los *estados globales posibles* es

$$\mathbf{G} = \mathbf{L}_0 \times \mathbf{L}_1 \times \cdots \times \mathbf{L}_n.$$

Los estados locales del entorno y de los agentes en $s \in \mathbf{G}$ son $s_0, s_1, ..., s_n$, respectivamente.

(3) Un *proceso* sobre $\mathbf{G}$ es una función $f : \mathbf{T} \longrightarrow \mathbf{G}$, donde $\mathbf{T}$ es el conjunto de los instantes temporales.

(4) Un *sistema* $\mathcal{R}$ sobre $\mathbf{G}$ es un conjunto de procesos sobre $\mathbf{G}$.

(5) El conjunto $W_\mathcal{R}$ de los estados globales de $\mathcal{R}$ es

$$W_\mathcal{R} = \{r(m) : m \in \mathbf{T} \wedge r \in \mathcal{R}\}.$$

La lógica epistémica para n agentes se aplica a un n-sistema $\mathcal{R}$ después de

(1) elegir un conjunto $\mathbf{\Phi}$ de enunciados atómicos para describir los hechos básicos;

(2) asociar con cada estado global de $\mathcal{R}$ el conjunto de enunciados atómicos verdaderos en él, lo que se hace con funciones de *asignación* $\pi : \mathbf{W}_\mathcal{R} \longrightarrow \wp(\mathbf{\Phi})$, tales que, para cada $s \in \mathbf{W}_\mathcal{R}$, $\pi(s)$ es el conjunto de enunciados atómicos verdaderos en s;

(3) definir para cada agente a una relación $\sim_a$ binaria sobre los estados globales y

(4) establecer que el agente a sabe el enunciado φ en el estado s si y sólo si φ es verdadero en todos los estados s' tales que $s \sim_a s'$.

Esta aplicación es general y da lugar al concepto de *conocimiento externo* que obedece una lógica modal normal.

Un sistema multiagente junto con una atribución de conocimiento a los agentes, que se establece al asociar con cada agente una relación binaria entre estados globales, lo denominaremos *sistema multiagente epistémico.*

Sea $\mathcal{R}$ un n-sistema epistémico, $\mathbf{\Phi}$ un conjunto de enunciados atómicos y π una función de asignación. Entonces, para cada $s \in W_{\mathcal{R}}$ y cada enunciado φ, "*φ es verdadero en el estado global s de $\mathcal{R}$ para π*", en símbolos $(\mathcal{R},\pi, s) \Vdash_n \varphi$, se define por inducción:

a) si $\varphi \in \mathbf{\Phi}$, $(\mathcal{R},\pi, s) \Vdash_n \varphi$ si y sólo si $\varphi \in \pi(s)$;

b) $(\mathcal{R},\pi, s) \Vdash_n \varphi \wedge \psi$ si y sólo si $(\mathcal{R},\pi, s) \Vdash_n \varphi$ y $(\mathcal{R},\pi, s) \Vdash_n \psi$;

c) $(\mathcal{R},\pi, s) \Vdash_n \neg\varphi$ si y sólo si no se tiene $(\mathcal{R},\pi, s) \Vdash_n \varphi$;

d) $(\mathcal{R},\pi, s) \Vdash_n K_a\varphi$ si y sólo si, para todo $s' \in W_{\mathcal{R}}$ tal que $s \sim_a s'$, $(\mathcal{R},\pi, s') \Vdash_n \varphi$.

A partir de esta relación se pueden definir:

a) $(\mathcal{R},\pi) \Vdash_n \varphi$ si y sólo, para todo $s \in W_{\mathcal{R}}$, $(\mathcal{R},\pi, s) \Vdash_n \varphi$ (φ es *válido para* $\mathcal{R}$ y π);

b) $(\mathcal{R}, s) \Vdash_n \varphi$ si y sólo si, para toda $\pi : \mathbf{W}_{\mathcal{R}} \longrightarrow \wp(\mathbf{\Phi})$, $(\mathcal{R},\pi, s) \Vdash_n \varphi$ (φ es *válido para* $\mathcal{R}$ *en el estado global* s);

c) $\mathcal{R} \Vdash_n \varphi$ si y sólo si, para todo $s \in W_{\mathcal{R}}$, $(\mathcal{R}, s) \Vdash_n \varphi$ (φ es *válido en* $\mathcal{R}$);

d) para cada clase $\mathcal{C}$ de n-sistemas, $\mathcal{C} \Vdash_n \varphi$ si y sólo si, para todo $\mathcal{R} \in \mathcal{C}$, $\mathcal{R} \Vdash_n \varphi$ (φ es *válido en la clase* $\mathcal{C}$). Si $\mathcal{C}$ es la clase de todos los n-sistemas, se escribe $\Vdash_n \varphi$ en lugar de $\mathcal{C} \Vdash_n \varphi$ (φ es *válido*).

Conceptos estrechamente relacionados con estos y que no tienen análogo en los modelos de Kripke tienen que ver con los *procesos.* Así podemos definir para cada $r \in \mathcal{R}$:

e) $(\mathcal{R}, r) \Vdash_n \varphi$ si y sólo si, para todo $m \in \mathbf{T}$, $(\mathcal{R}, r(m)) \Vdash_n \varphi$ (φ es *válido* para $\mathcal{R}$ en el proceso r).

Los operadores E_G, C_G, S_G y F_G se aplican a los sistemas multiagentes usando cláusulas análogas a las vistas para modelos de Kripke.

Entre los sistemas multiagentes epistémicos destacan aquellos en los que las relaciones binarias asociadas con los agentes son relaciones de equivalencia tales que para cada agente a_i y cada $s, s' \in W_{\mathcal{R}}$, $s \sim_{a_i} s'$ si y sólo si $s_i = s'_i$. A estos sistemas los denominaremos sistemas multiagentes epistémicos *regulares.*

La relación entre modelos de Kripke y sistemas multiagentes es muy estrecha como lo muestra el

Teorema 3 *Sea $\mathfrak{C}$ el conjunto de todos los modelos de Kripke para $\mathbf{L}_{\mathcal{A},\emptyset}$ cuyas relaciones de accesibilidad son relaciones de equivalencia y $\mathfrak{R}$ el conjunto de los sistemas multiagentes epistémicos regulares. Entonces para cada $\varphi \in \mathbf{L}_{\mathcal{A},\emptyset}$, $\mathfrak{C} \vDash \varphi$ si y sólo si $\mathfrak{R} \Vdash_n \varphi$. Es decir, el conjunto de enunciados demostrables en $\boldsymbol{S5}_n$ es el mismo que el de los enunciados válidos en sistemas epistémicos regulares con n agentes.*

Demostración. $\Longrightarrow$) Con cada n-sistema $\mathcal{R}$ sobre el conjunto de agentes $\mathcal{A} = \{a_1, \ldots, a_n\}$ y cada asignación π se puede asociar el modelo de Kripke

$$\mathfrak{M} = \langle\langle W_{\mathcal{R}}, \mathcal{K}_{a_1}, \ldots, \mathcal{K}_{a_n}\rangle, \pi_1\rangle,$$

donde cada $\mathcal{K}_{a_i}$ es tal que, para todo $r, s \in W_{\mathcal{R}}$,

$$r\mathcal{K}_{a_i}s \Longleftrightarrow r \sim_{a_i} s$$

y π_1 es tal que, para todo $u \in W_{\mathcal{R}}$ y enunciado atómico α, $\pi_1(u,\alpha) = \mathbf{t}$ si $\alpha \in \pi(u)$ y $\pi_1(u,\alpha) = \mathbf{f}$ si $\alpha \notin \pi(u)$.
Es fácil mostrar que, para cada $s \in W_{\mathcal{R}}$, $(\mathcal{R}, \pi, s) \Vdash_n \varphi$ si y sólo si $\mathfrak{M}, s \vDash \varphi$.
Esta parte de la demostración nos dice que todo n-sistema epistémico, sea o no regular, es un modelo de Kripke.
$\Longleftarrow$) Con cada modelo de Kripke con n agentes $\mathfrak{M} = \langle \mathfrak{E}, \pi \rangle$, donde

$$\mathfrak{E} = \langle W, \mathcal{K}_{a_1}, \mathcal{K}_{a_2}, ..., \mathcal{K}_{a_n} \rangle\,,$$

y cada $\mathcal{K}_{a_i}$ $(i = 1, \ldots, n)$ es una relación de equivalencia sobre W, se asocia un n-sistema regular multiagente $\mathcal{R}$ y la función $\pi_1 : W_{\mathcal{R}} \longrightarrow \wp(\mathbf{\Phi})$ como sigue:

a) A cada $w \in W$ se asocia el tuplo $t^w = (t^w_1, ..., t^w_n)$ tal que, para $i = 1, ..., n$,

$$t^w_i = w/\mathcal{K}_{a_i}$$

donde $w/\mathcal{K}_{a_i} = \{u | u \in W \wedge w\mathcal{K}_{a_i}u\}$ (la clase de equivalencia de w respecto a $\mathcal{K}_{a_i}$).

b) Si la función, que a cada $w \in W$ le asigna el n tuplo t^w, es inyectiva, entonces con cada $w \in W$ asociamos el $n + 1$ tuplo $s^w = (0, t^w_1, ..., t^w_n)$.

c) En caso de que no sea inyectiva, sea $z_1, ..., z_p$ una enumeración sin repetición del conjunto

$$\{t^w : w \in W\}.$$

Para cada $i = 1, ..., p$, sea $v^0_i, \ldots, v^{k_i}_i$ una enumeración sin repetición del conjunto

$$C_i = \{w : w \in W \wedge t^w = z_i\}.$$

Entonces, si $w \in C_i$ y $w = v^j_i$, definimos $s^w = (j, t^w_1, ..., t^w_n)$. Así, la función que asocia con cada $w \in W$ el $n + 1$ tuplo s^w es una función inyectiva.

d) El sistema $\mathcal{R}$ que asociamos con el marco $\mathfrak{E}$ ha de cumplir que $W_{\mathcal{R}} = \{s^w : w \in W\}$. El conjunto $\mathbf{\Phi}$ ha de contener todas las variables enunciativas que ocurren en φ y la función π_1 es tal que, para todo $s^w \in W_{\mathcal{R}}$,

$$\pi_1(s^w) = \{p : p \in \mathbf{\Phi} \wedge \pi(w, p) = \mathbf{t}\}.$$

Es fácil mostrar que $\mathfrak{M}, w \vDash \varphi$ si y sólo si $(\mathcal{R}, \pi_1, s^w) \Vdash_n \varphi$. ■

El teorema anterior puede generalizarse estableciendo que con cada modelo de Kripke para n agentes puede asociarse un n-sistema epistémico $\mathcal{R}$. La idea de la demostración consiste en lo siguiente:

- Sea $\mathfrak{M} = \langle W, \mathcal{K}_{a_1}, \ldots, \mathcal{K}_{a_n} \rangle$ un marco de Kripke.
- Para cada $w \in W$, sea $w^* =< \underbrace{w, \ldots, w}_{n+1 \text{ veces}} >$.
- El conjunto de los estados globales de $\mathcal{R}$ es $W_{\mathcal{R}} = \{w^* | w \in W\}$.
- Definimos $\sim_{a_i}$ para $i = 1, \ldots, n$ por $s^* \sim_{a_i} t^*$ si y sólo si $s\mathcal{K}_{a_i}t$.
- Finalmente, si consideramos un modelo, hacemos que cada enunciado atómico sea verdadero en w^* si y sólo si es verdadero en w.

Clases interesantes de sistemas multiagentes epistémicos regulares son las clases $\mathcal{SC}_n$ de *sistemas cerrados*. Un n-sistema $\mathcal{R}$ es *cerrado* si y sólo si, para cada $r, s \in \mathcal{R}$ y $m, p \in \mathbf{T}$,

$$\forall i(1 \leq i \leq n \wedge [r(m)]_i = [s(p)]_i) \rightarrow r(m) = s(p).$$

Es decir, dos estados globales son iguales si los estados locales de cada agente son iguales en ambos estados globales. Sus estados globales no dependen del entorno (como si el entorno no existiera) y puede considerarse que son n-tuplos $(s_1, ..., s_n)$.

Si $\mathcal{R} \in \mathcal{SC}_n$, entonces la relación $\sim$ tal que, para cada $s, t \in W_{\mathcal{R}}$,

$$s \sim t \Longleftrightarrow s \sim_{a_1} t \wedge \cdots \wedge s \sim_{a_n} t,$$

es de equivalencia y todas las clases de equivalencia respecto de $\sim$ son unitarias. Con está relación puede asociarse un operador que coincide con F_G, $G = \{a_1, ..., a_n\}$. En $\mathcal{SC}_n$ son válidos todos los enunciados de la forma:

$$\varphi \leftrightarrow F_G\varphi.$$

Es decir, el conocimiento distribuido de todos los agentes en un estado global s coincide con todo lo que es verdadero en s. Es el conocimiento completo.

4. Marcos de Kripke con número de agentes variable

Las lógicas vistas hasta ahora se basan en marcos y sistemas multiagentes en los que el conjunto de agentes que existen en cada mundo o en cada estado global es el mismo. Así, no sirven para analizar situaciones donde nacen y/o mueren agentes, es decir, se crean y/o destruyen agentes. Una posible generalización de los marcos de Kripke para tratar estas situaciones se consigue asociando con cada mundo un conjunto de agentes.[8] Formalmente, un *marco vivo de Kripke* sobre el conjunto $\mathcal{A}$ de *posibles* agentes y $\mathcal{G} \subseteq \wp(\mathcal{A})$ tiene la forma

$$\mathfrak{V} = \langle W, f, \mathcal{K} \rangle \, ,$$

donde W es un conjunto no vacío de *mundos*, $f : W \longrightarrow \wp(\mathcal{A})$ es una función que asigna a cada mundo el conjunto de agentes que viven en él y $\mathcal{K} : \mathcal{A} \longrightarrow \wp(W \times W)$ una función que a cada agente le asigna una relación binaria sobre W. En lugar de $f(w)$ y $\mathcal{K}(a)$ escribiremos A_w y $\mathcal{K}_a$ respectivamente.

Los *modelos vivos de Kripke* sobre $\mathcal{A}$ y $\mathcal{G}$ son pares ordenados

$$\mathfrak{M} = \langle \mathfrak{V}, \pi \rangle$$

donde $\mathfrak{V}$ es un marco vivo de Kripke y π una asignación de valores de verdad. La relación $\vDash$ puede definirse como en el apartado 2.

Para el estudio de las Lógicas generadas por los marcos vivos de Kripke es útil introducir las constantes $\mathbf{v}_a$ para $a \in \mathcal{A}$, con el propósito de que $\mathbf{v}_a$ sea verdadero en un mundo w si y sólo si a vive en w. Es decir, se tiene

$$\mathfrak{M}, w \vDash \mathbf{v}_a \text{ si y sólo si } a \in f(w).$$

El lenguaje apropiado para estos marcos vivos se construye con el vocabulario de $\mathbf{L}_{\mathcal{A},\mathcal{G}}$ aumentado con los símbolos $\mathbf{v_a}$ para $a \in \mathcal{A}$, añadiéndose a la definición de enunciados de $L_{\mathcal{A},\mathcal{G}}$ la cláusula

0) para cada $a \in \mathcal{A}$, $\mathbf{v}_a$ es un enunciado (atómico).

Un marco de Kripke $\langle W, \mathcal{K} \rangle$ sobre $\mathcal{A}$ y $\mathcal{G}$ es un caso particular de marco vivo, donde f asigna $\mathcal{A}$ a cada $x \in W$. Así, en cada mundo w es verdadero que el conjunto de agentes que hay en w es $\mathcal{A}$ y, por tanto, es conocimiento común. Así, $\mathbf{v}_a$ para $a \in \mathcal{A}$ serían válidos, es decir, verdaderos en cada mundo y para cada asignación.

[8]Sugerencia planteada por Grove & Halpern [3].

¿Debemos exigir algunas condiciones a las funciones f y/o $\mathcal{K}$ si queremos usar dichos marcos para analizar el conocimiento[9] de los agentes? En concreto, un agente a que no exista en el mundo w, ¿qué conocimiento tiene en w? ¿qué mundos le parecen posibles desde w? ¿desde qué mundo es posible w?

Si no exigimos nada, entonces, desde el punto de vista de la lógica epistémica en la que no ocurran las constantes $\mathbf{v}_a$, da lo mismo un marco vivo $\langle W, f, \mathcal{K}\rangle$ que un marco $\langle W, \mathcal{K}\rangle$ ambos sobre $\mathcal{A}$ y $\mathcal{G}$. Y, así, el conocimiento de un agente a en un mundo w no depende de que $a \in A_w$ o de que $a \notin A_w$. Puede darse el caso de que en un mundo w donde a no vive a supiera que vive, es decir, que es verdadero $K_a\mathbf{v}_a$. Igualmente podría suceder que en un mundo donde a vive, a supiera que no vive, es decir, $K_a\neg\mathbf{v}_a$. Estos casos se tienen, por ejemplo, en un modelo con dos mundos w_1 y w_2 tal que $a \notin f(w_1)$, $a \in f(w_2)$, $\langle w_1, w_2\rangle \in K_a$ y $\langle w_2, w_1\rangle \in K_a$ (Figura 5). Esto no es muy intuitivo, así, debe exigirse algo.

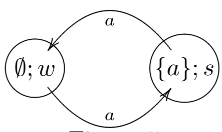

Figura 5

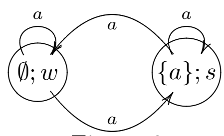

Figura 6

En el caso de que las relaciones $\mathcal{K}_a$ sean relaciones de equivalencia podemos tener que un agente a no sepa en un mundo w si está o no vivo, es decir, en w puede tenerse $\neg K_a\mathbf{v}_a \wedge \neg K_a\neg\mathbf{v}_a$ (Figura 6). Sin embargo, si las relaciones son reflexivas no se tiene nunca que a sepa que está vivo en un mundo donde no vive, es decir, si $a \in A_w$, entonces en w no se tiene $K_a\neg\mathbf{v}_a$. Tampoco puede tenerse que a sepa que vive en un mundo donde no vive, es decir, si $a \notin A_w$, entonces en w no se tiene $K_a\mathbf{v}_a$.

Dado que la relación $\mathcal{K}_a$ sirve para establecer el conocimiento de a, parece razonable suponer que si dos mundos están en la relación $\mathcal{K}_a$, entonces o en ambos existe a o en ambos está ausente. Formalmente,

$(\alpha)\qquad \forall w\forall s(w\mathcal{K}_a s \Longrightarrow a \in A_w \cap A_s \vee a \notin A_w \cup A_s).$

Si no suponemos esta restricción podemos tener los siguientes casos (Figura 5):

1) $s\mathcal{K}_a w$ tal que a vive en s y a no vive en w.

   Entonces en s sería verdadero que a no sabe que vive en s, es decir, $\neg K_a\mathbf{v}_a$. Además, si para todo w tal que $s\mathcal{K}_a w$, se tiene que a no vive en w, entonces en s se tiene que a sabe que no vive, es decir, $K_a\neg\mathbf{v}_a$

2) $w\mathcal{K}_a s$ tal que a no vive en w y a vive en s.

   Entonces en w sería verdadero que a no sabe que no vive en w, es decir, $\neg K_a\neg\mathbf{v}_a$, lo que puede parecer razonable ya que los que no viven no saben nada. Sin embargo, si para todo s tal que $w\mathcal{K}_a s$, se tiene que a vive en s, entonces en w el agente a sabe que vive, es decir, $K_a\mathbf{v}_a$. Esto no parece razonable.

Con la restricción (α) como tenemos la cláusula $K_\models$:

$$(\mathfrak{M}, w) \models K_a\varphi \text{ si y sólo si, para todo } s \in W \text{ tal que } w\mathcal{K}_a s,\ (\mathfrak{M}, s) \models \varphi$$

se tiene que, si w es tal que $a \notin A_w$, entonces existen enunciados φ tal que

$$(\mathfrak{M}, w) \models K_a\varphi;$$

en concreto cuando φ es una tautología. Así, el agente a sabe algo en el estado w en el que no existe, lo que no parece muy intuitivo ni razonable. Además, si no se imponen más condiciones podría suceder que el agente a supiera cosas diferentes en los diferentes mundos donde no exista.

[9]El conocimiento de un agente a en un mundo w del modelo $\mathfrak{M}$ puede identificarse con el siguiente conjunto de enunciados

$$\{\varphi : \mathfrak{M}, w \models K_a\varphi\}.$$

La restricción (α) elimina la posibilidad de que los mundos donde a no existe influyan en el conocimiento de a en un mundo donde existe e igualmente de que los mundos donde a vive no influyan en lo que sabe en los mundos donde no vive.

Con esta restricción (α), un agente a sabría en un mundo w en el que vive que vive, es decir, $K_a\mathbf{v}_a$ es verdadero en w. Igualmente, sabría en un mundo w en el que no vive que no vive, es decir, $K_a\neg\mathbf{v}_a$ sería verdadero en w. Aunque, podría suceder que también fueran verdaderos $K_a\neg\mathbf{v}_a$ y $K_a\mathbf{v}_a$ respectivamente, si no hay ningún mundo accesible desde w.

Grove & Halpern [3] proponen las siguientes restricciones para cada $a \in \mathcal{A}$:

(β) $\forall x\forall y(x\mathcal{K}_a y \to x \in W_a)$

(γ) $\forall x\forall y(x\mathcal{K}_a y \to y \in W_a)$

(δ) $\mathcal{K}_a$ es una relación de equivalencia sobre W_a

donde $W_a = \{w : a \in A_w\}$.[10] La restricción (β) nos dice que siempre que $x\mathcal{K}_a y$, a vive en x, es decir, el dominio de $\mathcal{K}_a$ es W_a. La restricción (γ) nos dice que siempre que $x\mathcal{K}_a y$, a vive en y, es decir, el rango de $\mathcal{K}_a$ es W_a. La restricción (δ) la entendemos en el sentido de que $\mathcal{K}_a$ restringida a W_a es una relación de equivalencia.

Si $\vDash$ se define usando la cláusula $K_{\vDash}$, se tiene

1) De (β) se sigue que $K_a\varphi$ es verdadero para cada enunciado φ en un mundo donde a no vive , así a sabe tanto que vive como que no vive. Sería válido el axioma $\neg\mathbf{v}_a \to K_a\bot$ o equivalentemente, $\neg K_a\bot \to \mathbf{v}_a$. Es decir, si a no sabe algo en un mundo, entonces vive en ese mundo; sin embargo podría suceder que a viviera en w y supiera todo.[11]

2) De (γ) se sigue que a sabe en un mundo donde vive que vive, es decir, se cumple $K_a\mathbf{v}_a$. Ahora bien, también se tiene que a sabe que vive en los mundo donde no vive. Así, con (γ) en todo mundo cada agente sabe que vive. Es decir, en todo mundo se tiene $K_a\mathbf{v}_a$. Puede darse el caso de que un agente que vive en w sepa en w el absurdo, es decir, se cumpla $\mathbf{v}_a \to K_a\bot$.

3) Con la restricción (γ) puede darse el caso de que un agente a sepa que vive en un mundo w donde no vive, $K_a\mathbf{v}_a$ y no sepa en w que no vive, es decir, $\neg K_a\neg\mathbf{v}_a$. Lo último ocurre cuando se tiene $x\mathcal{K}_a y$ y a no vive en x.

4) Con la restricción (δ) (es suficiente que $\mathcal{K}_a$ sea reflexiva sobre W_a o que para cada $w \in W_a$ exista un mundo s tal que $w\mathcal{K}s$) se tiene que en un mundo donde vive el agente a no se tiene que a sabe que no vive, es decir $\mathbf{v}_a \to \neg K_a\neg\mathbf{v}_a$. Pero, puede darse el caso de que tampoco sepa que vive, por ejemplo, si $w\mathcal{K}_a x$ y a no vive en x.

5) Con las restricciones (β), (γ) y (δ) se tiene que un agente en un mundo donde vive sabe que vive y no sabe que no vive, es decir, se tiene

$$\mathbf{v}_a \to K_a\mathbf{v}_a \wedge \neg K_a\neg\mathbf{v}_a.$$

Se elimina la posibilidad de que en un mundo donde vive a sepa tanto que vive como que no vive, ya que siempre hay un mundo accesible desde cualquier otro mundo. En donde no vive sabe tanto que vive como que no vive, es decir,

$$\neg\mathbf{v}_a \to K_a\mathbf{v}_a \wedge K_a\neg\mathbf{v}_a.$$

Con las tres restricciones, que implican la restricción (α), en los mundos w tales que $a \notin A_w$, para todo enunciado φ,

$$(\mathfrak{M}, w) \vDash K_a\varphi,$$

[10] W_a es el conjuntos de mundos en los que vive a.

[11] Hay que tener en cuenta que $K_a\bot \to K_a\varphi$ para cualquier enunciado φ. Si el absurdo forma parte del conocimiento de un agente, todo enunciado forma parte de dicho conocimiento.

ya que, para tales w no existe ningún s tal que $w\mathcal{K}_a s$. Así, el conjunto de mundos donde el agente a existe es

$$\{w|(\mathfrak{M}, w) \models \neg K_a\bot\}$$

de lo que se sigue fácilmente que $\mathbf{v}_a \leftrightarrow \neg K_a\bot$.

En esta propuesta de Grove & Halpern en la que $\mathbf{v}_a$ son definibles, se cumplen los axiomas **A1**, **A2**, **4**, **5** y, en lugar de **T**, el axioma **T´**

T´. $\neg K_a\bot \wedge K_a\varphi \rightarrow \varphi$, para $a \in \mathcal{A}$

y la regla de inferencia **MP** y, para $a \in \mathcal{A}$, la regla **N**.

¿Qué propiedades tienen en este caso los operadores E_G, C_G, S_G y F_G cuando G no es vacío y $G \subseteq \mathcal{A}$? Es fácil ver que, para todo mundo w y todo enunciado φ, son verdaderos en w

$$\begin{array}{rcl} E_G\varphi & \leftrightarrow & E_{G\cap A_w}\varphi \\ C_G\varphi & \leftrightarrow & C_{G\cap A_w}\varphi \end{array}$$

es decir, el conocimiento de los agentes de G que no existen en w no influyen en el comportamiento de E_G y C_G, lo cual parece razonable. Sin embargo, si hay un agente $a \in G$ que no existe en w, entonces, para cada enunciado φ, $S_G\varphi$ y $F_G\varphi$ son verdaderos en w, lo que no parece razonable.

Por supuesto en esta propuesta un agente a sabe *todo* (lo verdadero y lo falso) en los mundos donde no existe. Aunque, parece más intuitivo y razonable suponer que en ese mundo no tiene conocimiento, es decir, no sabe *nada*.

Si queremos que un agente a no tenga conocimiento en los mundos donde no existe, deberemos cambiar la definición de la relación $\models$ sustituyendo la cláusula $K_\models$ por la siguiente $K_\exists$:

$(\mathfrak{M}, w) \models K_a\varphi$ si y sólo si $a \in A_w$ y, para todo $s \in W$ tal que $w\mathcal{K}_a s$, $(\mathfrak{M}, s) \models \varphi$.

Si distinguimos los operadores K que siguen $K_\models$ y los operadores K' que siguen[12] $K_\exists$, es fácil ver que

$$K'_a\varphi \leftrightarrow \mathbf{v}_a \wedge K_a\varphi.$$

Esta equivalencia puede sustituirse por

$$K'_a\varphi \leftrightarrow \neg K_a\bot \wedge K_a\varphi$$

cuando las relaciones $\mathcal{K}_a$ son seriales sobre W_a (lo que ocurre si $\mathcal{K}_a$ son de equivalencia o reflexivas en W_a) y su dominio y rango coinciden con W_a.

De la cláusula $K_\exists$ se sigue que a no sabe$_\exists$ nada en los mundos donde no existe: si $a \notin A_w$, entonces, para todo enunciado φ, $(\mathfrak{M}, w) \models \neg K'_a\varphi$. De esto se sigue que, en los mundos w tal que $a \notin A_w$, se tiene, para K'_a, **A1** y, trivialmente, **A2**, **T**, **4** y **D**, pero no se tiene **5**, ni en general **B**. Además, ya no es adecuada la regla de necesitación **N** para K'_a. Aunque si es adecuada la regla de monotonía **M** (De $\varphi \rightarrow \psi$, infiérase $K'_a\varphi \rightarrow K'_a\psi$). Así, si, para cada $a \in \mathcal{A}$, $\mathcal{K}_a$ es una relación de equivalencia, se tendrá que la lógica para estas estructuras y los nuevos operadores K'_a cumple **A1** y, para todo $a \in \mathcal{A}$, los axiomas **A2**, **T, 4** y **D** y las reglas de inferencia **MP** y, para cada $a \in \mathcal{A}$, la regla **M**.

En este caso, podemos definir los operadores E'_G, C'_G, S'_G y F'_G a partir de los operadores K'_a de forma análoga a como se definieron E_G, C_G, S_G y F_G a partir de los operadores K_a, en la forma

- $(\mathfrak{M}, w) \models E'_G\varphi$ si y sólo si, para todo $a \in G$, $(\mathfrak{M}, w) \models K'_a\varphi$
- $(\mathfrak{M}, w) \models C'_G\varphi$ si y sólo si, para todo $k > 0$, $(\mathfrak{M}, w) \models {E'}^k_G\varphi$

  donde ${E'}^k_G$ se define como sigue:

$$\begin{array}{rcl} {E'}^0_G\varphi & = & \varphi \\ {E'}^{k+1}_G\varphi & = & E'_G({E'}^k_G\varphi). \end{array}$$

[12] $K'_a\varphi$ lo leeremos como el agente a sabe$_\exists$ φ.

- $(\mathfrak{M}, w) \models S'_G\varphi$ si y sólo si hay un $a \in G$ tal que $(\mathfrak{M}, w) \models K'_a\varphi$
- $(\mathfrak{M}, w) \models F'_G\varphi$ si y sólo si, para todo $s \in W$ tal que $\langle w, s\rangle \in \bigcap_{a\in G\cap A_w} \mathcal{K}_a$, $(\mathfrak{M}, s) \models \varphi$.

Entonces, los operadores E'_G, C'_G, S'_G y F'_G tienen las siguientes propiedades. Para todo mundo w y todo enunciado φ, son verdaderos en w

$$\begin{array}{rcl} S'_G\varphi & \leftrightarrow & S'_{G\cap A_w}\varphi \\ F'_G\varphi & \leftrightarrow & F'_{G\cap A_w}\varphi \end{array}$$

es decir, el conocimiento$_\exists$ de los agentes de G que no existen en w no influyen en el comportamiento de S'_G y F'_G. Sin embargo, si hay un $a \in G$ que no existe en w, entonces, para cada enunciado φ, $\neg E'_G\varphi$ y $\neg C'_G\varphi$ son verdaderos en w.

¿Cual de los dos operadores, K_a o K'_a, representa mejor el saber de un agente cuando hay mundos donde el agente no vive? En los mundos donde un agente vive, $K_a\varphi$ es equivalente a $K'_a\varphi$.

5. Sistemas con número variable de agentes

La definición de sistema multiagente que se presentó en el apartado 2, puede modificarse de la siguiente forma. Sea $\mathbf{L}$ el conjunto de todos los posibles estados locales tanto del entorno como de los agentes. Sea $\mathbf{G} = \mathbf{L}^{n+1}$, el conjunto de todos los estados globales posibles (que pueden ser más que en el apartado 2). Cada $s \in \mathbf{G}$ puede interpretarse como una función $s : \{0, 1, ..., n\} \longrightarrow \mathbf{L}$, con $s(0)$ el estado del entorno y $s(i)$, para $i = 1, ..., n$, el estado del agente a_i. El resto es igual.

Apoyándonos en esta nueva presentación de un sistema, se definen sistemas tales que en cada $t \in \mathbf{T}$ pueden existir diferentes agentes. Así, sea $\mathcal{A}$ el conjunto de todos los agentes que pueden existir, $\mathbf{L}$ el conjunto de los estados locales en los que pueden hallarse los agentes o el entorno, $\mathbf{G}$ el conjunto de los estados globales posibles, es decir de las funciones parciales $s : \mathcal{A} \cup \{e\} \rightharpoonup \mathbf{L}$, donde $e \notin \mathcal{A}$ es el entorno y $s(e)$ está definido. Si s no está definida para el agente a, en símbolos $s(a)\!\uparrow$, entonces el agente a no existe en el estado global s. En lugar de a existe en el estado global s, diremos que $s(a)$ converge o $s(a)\!\downarrow$. Si $s(a)\!\downarrow$, $s(a)$ es el estado local de a en el estado global s. Entonces, un sistema $\mathcal{R}$ es un conjunto de funciones de $\mathbf{T}$ en $\mathbf{G}$. El conjunto $W_\mathcal{R}$ de los estados globales de $\mathcal{R}$ es el subconjunto de $\mathbf{G}$:

$$W_\mathcal{R} = \{f(t) | t \in T \wedge f \in \mathcal{R}\}.$$

$W_\mathcal{R}$ consta de funciones parciales $u : \mathcal{A} \cup \{e\} \rightharpoonup \mathbf{L}$.

En $W_\mathcal{R}$ pueden definirse de forma natural, para cada $a \in \mathcal{A}$, las relaciones $\sim_a$ y $\approx_a$ tal que, para todo $v, w \in W_\mathcal{R}$,

$$\begin{array}{rcl} v \sim_a w & \Longleftrightarrow & v(a)\!\downarrow \wedge w(a)\!\downarrow \wedge v(a) = w(a), \\ v \approx_a w & \Longleftrightarrow & v \sim_a w \vee (v(a)\!\uparrow \wedge w(a)\!\uparrow). \end{array}$$

$\sim_a$ es una relación de equivalencia sobre el conjunto $\{v | v \in W_\mathcal{R} \wedge v(a)\!\downarrow\}$, es decir, sobre los estados donde a existe; $\approx_a$ es de equivalencia sobre $W_\mathcal{R}$.

Si se usa la siguiente cláusula para definir la verdad de $K_a\varphi$

- $(\mathcal{R}, \pi, s) \Vdash_n K_a\varphi$ si y sólo si, para todo $s' \in W_\mathcal{R}$ tal que $s \sim_a s'$, $(\mathcal{R}, \pi, s') \Vdash_n \varphi$,

entonces, si $s(a)\!\uparrow$, $(\mathcal{R}, \pi, s) \Vdash_n K_a\varphi$ para todo enunciado φ. Y, si $s(a)\!\downarrow$, K_a cumple los axiomas de **S5**. (Coincide con la propuesta de Grove & Halpern para modelos de Kripke vivos).

Si se usa la siguiente cláusula para definir la verdad de $K_a\varphi$

- $(\mathcal{R}, \pi, s) \Vdash_n K_a\varphi$ si y sólo si, para todo $s' \in W_\mathcal{R}$ tal que $s \approx_a s'$, $(\mathcal{R}, \pi, s') \Vdash_n \varphi$, entonces, si $s(a)\uparrow$, a conoce en s cualquier enunciado φ tal que, para todo $t \in W_{\mathcal{R}}$ con $t(a)\uparrow$, $(\mathcal{R},\pi,t) \Vdash_n \varphi$, es decir, a conoce en s lo que es verdadero a la vez en todos los mundos donde no vive y, así, conoce lo mismo en todos los mundos donde no existe y al menos conoce todo teorema. Y, si $s(a)\downarrow$, K_a cumple los axiomas de **S5**. En general, se cumplen los axiomas de la lógica **S5**, exista o no el agente a en un mundo.

Si se usa la siguiente cláusula para definir la verdad de $K_a\varphi$

- $(\mathcal{R},\pi,s) \Vdash_n K_a\varphi$ si y sólo si $s(a)\downarrow$ y, para todo $s' \in W_{\mathcal{R}}$ tal que $s \sim_a s'$, $(\mathcal{R},\pi,s') \Vdash_n \varphi$

entonces se tienen para K_a las mismas propiedades que cuando se usaba la cláusula $K_\exists$ para modelos de Kripke vivos.

Referencias

[1] Chellas, B. F.: *Modal Logic. An introduction,* Cambridge University Press, 1980.

[2] Fagin, R., Halpern, J. Y., Moses, Y. & Vardi, M. Y.: *Reasoning about knowledge*, The MIT Press, Cambridge, Massachussetts, London, England, 1995.

[3] Grove, A. J. & Halpern, J. Y.: Naming and Identity in Epistemic Logics. Part I: The Propositional Case, *Journal of Logic and Computation* 3 (4): 345-378 (1993).

[4] Halpern, J. Y. & Moses, Y.: Knowledge and common knowledge in a distributed environment, *Journal of the ACM* 37(3), pp. 549-587 (1990).

[5] Halpern, J. Y. & Moses, Y. (1992): A Guide to Completeness and Complexity for Modal Logics of Knowledge and Belief, *Artificial Intelligence* 54, pp. 319-379.

[6] Halpern, J. Y. & Shore, R. A.: Reasoning About Common Knowledge with Infinitely Many Agents, *Proceeding of the 14th IEEE Symposium on Logic in Computer Science*, (1999), 384-393.

[7] Hintikka, J. (1962): *Knowledge and Belief*, Cornell University Press, Ithaca, NY. Traducción española con Prologo de J. L. Acero: *Saber y creer. Una introducción lógica de las dos nociones*, Tecnos, Madrid, 1979.

[8] Lehmann, D. J.: Knowledge, common knowledge, and related puzzles, In *Proc. 3rd ACM Symp. on Principles of Distributed Computing*, pp. 62-67 (1984).

[9] Lewis, D.: *Convention: A Philosophical Study.* Harvard U. P., Cambridge, Mass., 1969.

[10] Úbeda, J. P.: Epistemic Logics allowing different sets of agents in each world, en *ILCLI*, Universidad del Pais Vasco, San Sebastián, 2004, pp. 169-174.

[11] Úbeda, J. P.: Marcos de Kripke y Sistemas multiagentes, en P. A. Castillo Valdivieso y A. Córdoba (Eds): *Actas del I Simposio de Sistemas Complejos*, Paraninfo, 2005, pp. 39-43.

[12] von Wright, G.: *An Essay in Modal Logic*, North-Holland, 1951. Traducción española de A. A. Demarchi: *Ensayo de Lógica Modal*, Santiago Rueda, 1970.