DOI: https://doi.org/10.29092/uacm.v20i53.1036

Conocimiento y creencia en lógica epistémica dinámica

Fernando Soler-Toscano*

Resumen. Este trabajo presenta una introducción a la lógica epistémica dinámica a través de diversos sistemas formales que permiten representar conocimientos y creencias de uno o varios agentes, así como acciones epistémicas que pueden modificarlos. Concretamente, presentamos la lógica de anuncios públicos, modelos de acción y modelos de plausibilidad.

Palabras clave. Lógica epistémica dinámica; anuncios públicos; modelos de acción; modelos de plausibilidad; conocimiento y creencia.

Knowledge and belief in dynamic epistemic logic

Abstract. This paper presents an introduction to dynamic epistemic logic through some formal systems that allow to represent knowledge and beliefs of one or more agents, as well as epistemic actions that can modify them. Specifically, we present the public announcements logic, action models and plausibility models.

Key words. Dynamic epistemic logic; public announcements; action models; plausibility models; knowledge and belief.

* Profesor Investigador en la Universidad de Sevilla, España. Pertenece al Grupo de Lógica, Lenguaje e Información. Correo electrónico: fsoler@us.es

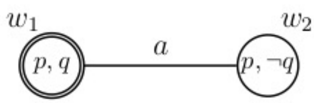
Introducción: lógica del conocimiento

Este trabajo se centra en varios sistemas de lógica epistémica dinámica, es decir, sistemas lógicos que permiten representar y razonar con información relativa al conocimiento y creencia de uno o varios agentes, así como estudiar de qué modo esta información puede cambiar mediante acciones epistémicas (Alchourrón *et al.*, 1985; van Benthem, 2007; van Benthem, 2011) que implican observación, comunicación, cambio de creencias, etc. Concretamente, presentaremos y discutiremos nociones elementales de lógica de anuncios públicos (en que los agente pueden ofrecer información abierta a todos los participantes), modelos de acción (que permiten formas más elaboradas de comunicación) y modelos de plausibilidad (que integran conocimiento y creencia).

Comenzamos presentando algunas nociones de lógica epistémica, de momento sin elementos dinámicos. Lo hacemos con un ejemplo que nos servirá de hilo conductor durante todo el trabajo. Ana y Belén han quedado en ir a casa de Carmen, que celebra su fiesta de cumpleaños. A diferencia de otras veces, no han acordado cuál de las dos buscará un regalo para Carmen. Así que camino de casa, sin poder comunicarse con Belén, Ana decide parar en una librería y comprar un libro para Carmen. Del mismo modo, sin poder comunicarse con Ana, Belén entra en una floristería y compra un ramo de flores para Carmen.

Conocimiento individual

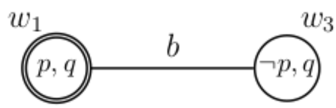
Podemos utilizar diagramas para representar el conocimiento de Ana y Belén por separado acerca de los regalos que han comprado para Carmen. Simplificando, sea p la proposición que representa "Ana ha comprado un regalo para Carmen", por lo que $\neg p$, la negación de p, representa que "Ana no ha comprado un regalo para Carmen". Del mismo modo, q puede representar "Belén ha comprado un regalo para Carmen". El conocimiento de Ana se puede representar mediante el siguiente modelo:



DOI: https://doi.org/10.29092/uacm.v20i53.1036

Vemos dos estados, w_1 y w_2, que representan, respectivamente, dos situaciones, la de que ambas Ana y Belén han comprado regalos para Carmen (en w_1 tanto p como q son verdaderas) como que solo lo ha comprado Ana pero no Belén (en w_2 solo p es verdadera, q es falsa). Ambas situaciones son indistinguibles para Ana, lo que representamos mediante la línea que une w_1 y w_2. El doble círculo que rodea w_1 significa que ese es el estado real, cosa que Ana no sabe, pues no es capaz de determinar qué estado w_1 o w_2 es el real. Ana sabe que p es verdad, dado que se verifica en w_1 y w_2, pero no conoce la verdad de q.

Como veremos posteriormente, en el diagrama anterior están todos los elementos propios de un modelo de Kripke para la lógica epistémica. Es interesante observar que la línea que une w_1 con w_2 (accesibilidad) indica indistinguibilidad epistémica, es decir, las opciones que están vinculadas por a son indistinguibles para la información de Ana, aunque otros agentes (Belén, por ejemplo, pueden distinguirlas).Del mismo modo, el siguiente modelo representa la información que tiene Belén:

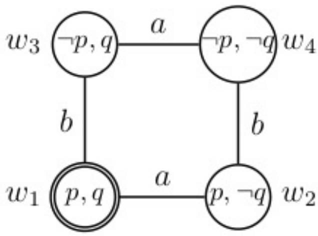


Aquí, podemos observar que el estado real, w_1, sigue siendo el mismo. Sin embargo, Belén no puede distinguirlo de otro w_3 (diferente del w_2 anterior) en el que esta vez es Ana la que no compró regalo para Carmen. Decimos que Belén conoce q pero no p. Así que cada una de nuestras amigas sabe que ella misma lleva un regalo para Carmen, pero no sabe si lo llevará la otra.

Conocimiento de dos agentes

Ahora, veamos cómo podemos representar en un modelo, simultáneamente, la situación epistémica de Ana y Belén:

DOI: https://doi.org/10.29092/uacm.v20i53.1036



Vemos los estados w_1, w_2 y w_3 presentes en los modelos anteriores. Pero ha aparecido un nuevo estado w_4, donde resulta que ni Ana ni Belén han comprado regalos para Carmen. Pero ese estado, ¿por qué está ahí? Veamos que desde w_1 no existe una línea rotulada con a ni con b hacia w_4, porque ninguna de las dos amigas confunde el estado real (en que cada una de ellas ha comprado un regalo para Carmen) con la posibilidad de que ninguna lo hubiera comprado. Para cada una de ellas, solo resulta indistinguible del estado real w_1 el caso en que sea la otra la que no ha comprado regalo para Carmen. Pero dado que Ana no puede distinguir la situación real w_1 de w_2 en que Belén no ha comprado regalo para Carmen (por ello la línea rotulada con a entre w_1 y w_2), Ana considera posible que el estado real fuese w_2, y en ese caso, si Belén no hubiera comprado un regalo para Carmen, Belén consideraría posible w_4, es decir, que no lo hubiese hecho ninguna de las dos. De modo que Ana concibe que Belén conciba w_4, donde ninguna compró un regalo para Carmen. Del mismo modo, Belén, no distingue el estado real w_1 de w_3, porque no sabe si Ana compró o no un regalo, de modo que Belén considera posible el caso w_3 en que Ana consideraría la posibilidad w_4. Así que el estado w_4 no es algo que ninguna de las dos amigas considere posible, pero cada una de ellas considera posible que la otra lo crea posible.

Lenguaje y semántica de la lógica epistémica

Mientras Ana y Belén llegan a casa de Carmen podemos aprovechar para introducir algunos detalles formales de la lógica epistémica (van Ditmarsch *et al.*, 2007). Consideremos un lenguaje formal construido a partir de un conjunto de proposiciones (que incluye y) y un conjunto de agentes A (que incluye a Ana y Belén), cuyas fórmulas se construyen mediante la siguiente gramática:

DOI: https://doi.org/10.29092/uacm.v20i53.1036

$$\varphi ::= \lambda \mid \neg\varphi \mid \varphi\wedge\varphi \mid \varphi\vee\varphi \mid \varphi\rightarrow\varphi \mid K_e\,\varphi$$

para cualesquiera $\lambda \in P$ y $e \in A$. Se trata de una extensión del lenguaje de la lógica proposicional donde hemos introducido un nuevo operador $K_e\,\varphi$ para indicar que "el agente e sabe que φ".

En nuestro ejemplo, $K_a\, p \wedge \neg K_a$ q expresa que Ana sabe que p (ella misma ha comprado un regalo para Carmen) pero no sabe que q (que Belén también compró un regalo). Podemos introducir el operador $\hat{K}_c\,\varphi$ que leemos como "el agente c considera φ como una posibilidad", o bien "el conocimiento de c no descarta φ". Semánticamente tendremos que $\hat{K}_c\,\varphi$ equivaldrá a $\neg K_c \neg\varphi$. En nuestro ejemplo, pese a ser p y q verdaderas en el estado real w_1, tenemos que se verifica (en dicho estado) $\hat{K}_a\,\hat{K}_b\,(\neg p \wedge \neg q)$, es decir, Ana considera posible que Belén considere posible $\neg p \wedge \neg q$.Estamos comenzando a escribir fórmulas que resultan verdaderas en un modelo, o en un estado de dicho modelo, sin haber ofrecido aún los criterios que determinan la verdad de una expresión de la lógica epistémica. Utilizamos modelos epistémicos para ello. Un modelo epistémico se compone de los siguientes elementos:

- S, que es un conjunto de estados. En los ejemplos anteriores, w_1, w_2, etc., son estados. Los estados tienen el papel de los mundos posibles en lógica modal.
- (R_e) es una familia de relaciones de accesibilidad. Es decir, para cada agente $e \in A$ existe una relación de accesibilidad R_e que es un subconjunto de pares ordenados de $S \times S$. En los ejemplos de Ana y Belén, tenemos que $(w_1, w_2) \in R_a$ *o* $(w_1, w_3) \in R_b$.
- V es una función que asigna a cada una de las proposiciones de P un subconjunto de S, que se corresponde con aquellos estados en que la proposición es verdadera. Así, en el último modelo, $V(q)$={w1, w3}.

Como vemos, un modelo epistémico especifica cuáles son los estados posibles (S), qué es verdadero en cada estado (V), y qué estados son indistinguibles para cada agente e (R_e). Si queremos especificar cuál es el estado real de un cierto modelo epistémico, utilizamos la notación (M,w) con $w \in S$ para especificar que w es el estado distinguido de M. En nuestros modelos

DOI: https://doi.org/10.29092/uacm.v20i53.1036

anteriores, el estado distinguido era w_1.

Dado el modelo $M=\langle S,(R_e),V\rangle$, para comprobar si una fórmula φ es verdadera en un estado distinguido $(M,w\vDash\varphi)$ o no $(M,w\nvDash\varphi)$, utilizamos las siguientes reglas:

- Para cada proposición $\lambda\in P$, $M,w\vDash\lambda$ si y solo si $w\in V(\lambda)$. Es decir, la verdad de las proposiciones en un determinado estado depende de la definición de V en el modelo epistémico.
- Para la negación, $M,w\vDash\neg\varphi$ si y solo si $M,w\nvDash\varphi$, es decir, la negación de cierta fórmula es verdadera cuando la fórmula sin negar es falsa.
- Para la conjunción, $M,w\vDash\varphi\wedge\psi$ si y solo si $M,w\vDash\varphi$ y $M,w\vDash\psi$.
- En el caso de la disyunción, $M,w\vDash\varphi\vee\psi$ si y solo si $M,w\vDash\varphi$ o bien $M,w\vDash\psi$. Utilizamos la disyunción inclusiva que es verdadera cuando ambas φ y ψ son verdaderas.
- Para el condicional, $M,w\vDash\varphi\rightarrow\psi$ si no es el caso de que $M,w\vDash\varphi$ *y* $M,w\nvDash\psi$.
- En el caso del operador de conocimiento, $M,w\vDash K_c\,\varphi$ si y solo si para cada w' tal que $(w,w')\in R_c$, se verifica $M,w'\vDash\varphi$.

El último punto es el característico de la lógica epistémica: el conocimiento de φ por parte del agente c requiere que en todos los estados que c considera epistémicamente posibles (indistinguibles de w), la fórmula φ sea verdadera. Para el operador $\hat{K}$, dual de K, el criterio para que $\hat{K}_c\,\varphi$ sea verdadera en (M,w) resulta: $M,w\vDash\hat{K}_c\,\varphi$ si y solo si existe $w'\in S$ tal que $(w,w')\in R_c$ y $M,w'\vDash\varphi$. De este modo, $\hat{K}_c\,\varphi$ requiere que haya algún estado accesible para c donde φ sea verdadera, lo que equivale a $\neg K_c\neg\varphi$.

En un modelo $M=\langle S,(R_e),V\rangle$ puede ocurrir que una cierta fórmula φ sea verdadera en cierto estado $w\in S$ $(M,w\vDash\varphi)$ y falsa en otro estado w' $(M,w'\nvDash\varphi)$. En el ejemplo anterior, considerando el modelo de los cuatro estados, $M,w_1\vDash p\wedge q$ pero $M,w_2\nvDash p\wedge q$.

Cuando una fórmula φ es verdadera en todos los estados de un modelo M decimos que es válida en el modelo y lo indicamos como $M\vDash\varphi$. ¿Hay alguna fórmula válida en nuestro modelo de Ana y Belén? En cada uno de los estados hay una interpretación diferente de p y q, por lo que ninguna fórmula proposicional (salvo las tautologías) será verdadera en todos los estados. Sin embargo, sí que hay fórmulas epistémicas válidas en el mo-

DOI: https://doi.org/10.29092/uacm.v20i53.1036

delo. Para poder justificarlo necesitamos hacer explícitas las relaciones de accesibilidad para a y b. Entendemos para ello que son reflexivas $((w,w)\in R_c$ para cada $w\in S$ y $c\in A)$ y que cada línea que une dos estados establece que ambos son mutuamente accesibles. Con ello, $R\_a=\{(w_1,w_1), (w_2,w_2), (w_3,w_3), (w_4,w_4), (w_1,w_2), (w_2,w_1), (w_3,w_4), (w_4,w_3)\}$ y de forma análoga se define R_b. Entonces, algunas fórmulas válidas en el modelo (es un buen ejercicio comprobarlo) son:

- $K_a p\vee K_a\neg p$, que indica que Ana sabe si ella ha comprado un regalo o no. Esta fórmula no equivale a la tautología $K_a (p\vee\neg p)$. De hecho, Belén sabe que Ana ha comprado un regalo o no lo ha comprado (Belén sabe que $p\vee\neg p$, es decir, $K_b (p\vee\neg p))$, pero no podemos afirmar que Belén sabe p o sabe $\neg p$ (no se verifica $K_b p\vee K_b\neg p$).
- $K_b (K_a p\vee K_a\neg p)$, es decir, Belén sabe que Ana sabe si Ana ha comprado o no un regalo para Carmen. Es decir, aunque Belén no sabe si es el caso de que p, sabe que Ana lo sabe.

El conocimiento y sus propiedades

Cuando hemos tenido que evaluar fórmulas en el modelo epistémico de Ana y Belén, hemos incluido en la relación de accesibilidad de cada una de las agentes cada par de estados (w,w). Con ello, hemos hecho que la relación de accesibilidad fuera reflexiva. Tiene sentido, si accesibilidad equivale a indistinguibilidad epistémica, que cada estado sea indistinguible de sí mismo, dado que ningún agente puede encontrar diferencia epistémica entre una situación determinada y la misma situación. Además, cada línea que en el diagrama unía dos estados w y w' diferentes hemos entendido que indicaba accesibilidad mutua, incluyendo ambos pares (w,w') y (w',w) en la accesibilidad de la agente correspondiente. Con ello, hacemos que la relación de accesibilidad sea simétrica. La intuición tras esta decisión es que si un agente no puede distinguir w de w' entonces tampoco puede distinguir w' de w. Es una intuición más débil que la que nos hizo asumir la reflexividad de las relaciones de accesibilidad, pues alguien puede objetar que si w fuera el estado real, entonces w' sería indistinguible para cierto agente, pero si el estado real fuera w', ya sí encontraría alguna diferencia con w. Dejamos de momento

DOI: https://doi.org/10.29092/uacm.v20i53.1036

esta observación de lado. Otra asunción habitual respecto a las relaciones de accesibilidad R_c es que sean transitivas. Si un agente no puede distinguir w de w' ni tampoco w' de w'' entonces no podrá distinguir w de w''. Cuando en un modelo epistémico se utilizan relaciones de accesibilidad reflexivas, simétricas y transitivas (es decir, relaciones de equivalencia), tenemos la lógica epistémica llamada *S5*, que es la que vamos a caracterizar axiomáticamente y sobre la que nos vamos a cuestionar algunos aspectos.

Tomar las relaciones de accesibilidad R_c como relaciones de equivalencia tiene la consecuencia de que los siguientes esquemas serán válidos (verdaderos en cualquier estado de cualquier modelo) dado cualquier agente c y fórmula φ:

- Verdad: $K_c \varphi \rightarrow \varphi$. Si c conoce φ, entonces φ es verdadera. Parece una propiedad deseable que tradicionalmente distingue el conocimiento de la creencia.
- Introspección positiva: $K_c \varphi \rightarrow K_c K_c \varphi$. Si c conoce φ, entonces c conoce que conoce φ. Ya comenzamos a ver que K_c modela un tipo de conocimiento idealizado
- Introspección negativa: $\neg K_c \varphi \rightarrow K_c \neg K_c \varphi$. Si el agente c no conoce φ, entonces c conoce que no conoce φ. Esta propiedad del conocimiento es aún más cuestionable que la anterior, pues implica que el agente conoce todos los límites de su conocimiento.

A pesar de que algunas de estas propiedades pueden no parecer muy intuitivas, no suelen ser problemáticas si trabajamos con modelos con unas pocas proposiciones (o si queremos diseñar protocolos de comunicación segura usando lógica epistémica como en Cordón-Franco *et al.*, 2013). Así, dado que Ana no sabe si Belén ha comprado un regalo para Carmen ($\neg K_a q$), no tenemos mucho problema en aceptar que Ana sabe que ella misma no sabe si Belén compró un regalo: $K_a \neg K_a q$.

Modificando las propiedades de las relaciones de accesibilidad R_c cambiamos las propiedades del conocimiento. De todos modos, mientras nos movamos en el terreno de los sistemas modales normales, vamos a tener dos importantes propiedades que implican lo que conocemos como omnisciencia lógica:

DOI: https://doi.org/10.29092/uacm.v20i53.1036

- Cada agente conoce todas las tautologías: $\vDash\varphi$ implica $\vDash K_c\varphi$.
- Cada agente conoce todas las consecuencias de su conocimiento: $\vDash K_c(\varphi\rightarrow\psi)\rightarrow(K_c\varphi\rightarrow K_c\psi)$. Es decir, que si el agente c conoce $\varphi\rightarrow\psi$ y conoce φ, entonces conoce ψ.

Un agente lógicamente omnisciente conoce todas las tautologías y además su conocimiento es cerrado bajo consecuencia lógica, es decir, cada proposición que se siga lógicamente del conocimiento de c es conocida por c. En determinadas aplicaciones de la lógica epistémica se considera que la omnisciencia lógica es un problema (Soler-Toscano, Velázquez-Quesada, 2014), dado que los agentes reales no lo son ni pueden serlo dados los problemas de intratabilidad computacional que ello supondría. Por ello, existen propuestas de sistemas de lógica epistémica que evitan la omnisciencia lógica (Velázquez-Quesada, 2010), pero quedan fuera del alcance de este trabajo.

Conocimiento de grupos

En algunas de las fórmulas con las que anteriormente hemos representado el conocimiento de Ana y Belén, hemos anidado operadores de conocimiento para expresar lo que una de ellas conoce acerca de lo que la otra conoce. Así, $K_b\neg K_a$ q expresa que Belén sabe que Ana no sabe que q. También podemos introducir en el lenguaje ciertos operadores que nos permiten hablar de diferentes modos acerca del conocimiento de Ana y Belén como grupo de agentes (Baltag *et al.*, 2018). Por ejemplo, en el estado que hemos considerado real, w_1, tanto p como q son verdaderas (ambas compraron regalos para Carmen), pero ninguna de las dos amigas sabe $p\wedge q$. Sin embargo, uniendo su conocimiento podrían llegar a conocer $p\wedge q$. Decimos que esta fórmula es conocimiento distribuido para el grupo de agentes $\{a,b\}$ formado por Ana y Belén. Además, aunque ninguna de las dos sabe que $p\wedge q$, ambas conocen $p\vee q$ (Ana porque conoce p y Belén porque conoce q). Decimos que $p\vee q$ es conocimiento general para $\{a,b\}$, dado que las dos amigas lo conocen.

Formalmente definimos los siguientes operadores:

- $E_b\varphi$ expresa que φ es conocimiento general en el grupo B. Tenemos que $M,w\vDash E_b\varphi$ si y solo si para todo w',

DOI: https://doi.org/10.29092/uacm.v20i53.1036

- $(w,w') \in \cup_{c \in B} R_c$ implica $M,w' \vDash \varphi$.
- $D_b\,\varphi$ expresa que φ es conocimiento distribuido en el grupo B. Tenemos que $M,w \vDash D_b\,\varphi$ si y solo si para todo w',

$$(w,w') \in \cap_{c \in B} R_c \quad \text{implica} \quad M,w' \vDash \varphi.$$

Como vemos, los operadores de conocimiento general y distribuido se evalúan como el conocimiento K_c de un solo agente, pero tomando como relación de accesibilidad la unión o la intersección, respectivamente, de las relaciones de accesibilidad de todos los agentes del grupo. Así, para que algo sea conocimiento general tiene que ser verdadero en todos los estados que sean accesibles al menos para algún agente del grupo, de modo que si algún agente no conoce una cierta fórmula φ, entonces es que φ es falsa en cierto estado accesible para tal agente y por ello no puede ser conocimiento general. En el caso del conocimiento distribuido se atiende a la intersección de las relaciones de accesibilidad. Intuitivamente, es como juntar el conocimiento de todos los agentes del grupo B, pues se atiende solo a los estados w' indistinguibles de w para todos y cada uno de los agentes.

En nuestro modelo, la fórmula $D_{\{a,b\}}\,(p \wedge q)$ es verdadera en w_1 pues si atendemos solo a los estados accesibles tanto para Ana como para Belén desde w_1 tenemos solo el propio w_1 en que $p \wedge q$ es verdadera. Sin embargo, como en el conocimiento general atendemos a todos los estados accesibles a través de la unión de las relaciones de accesibilidad, en w_1 solo es conocimiento general aquello que sea verdadero tanto en w_1 como en w_2 y w_3. Por ello en w_1 es verdadera $E_{\{a,b\}}\,(p \vee q)$.

Otra forma mucho más fuerte de conocimiento es el conocimiento común. Intuitivamente, φ es conocimiento común de Ana y Belén cuando ambas lo conocen $K_a\,\varphi$ *y* $K_b\,\varphi$ pero también ambas conocen que ambas lo conocen $(K_a\,K_b\,\varphi$ *y* $K_b\,K_a\,\varphi)$ y así sucesivamente. No hay forma de que una de ellas contemple la posibilidad de que la otra contemple la posibilidad de que … que la otra contemple la posibilidad de $\neg\varphi$.

El conocimiento común es un modo muy fuerte y en cierto modo idealizado de conocimiento. Si Ana consiguiera mandar un mensaje a Belén para indicarle que compró un regalo para Carmen (*p*), Belén podría no haberlo leído, por lo que Ana no sabe que Belén sabe que p. Belén puede mandar a Ana una confirmación de lectura del mensaje que ha recibido,

DOI: https://doi.org/10.29092/uacm.v20i53.1036

y entonces Ana sabrá que Belén sabe que p ($K_a K_b p$) pero entonces Belén sigue sin saber que Ana recibió la confirmación, por lo que $\neg K_b K_a K_b p$. Obtener conocimiento común en el sentido que vamos a ver requeriría de un intercambio infinito de confirmaciones de lectura entre Ana y Belén. Los agentes reales nos conformamos por lo general con un par de pasos de confirmación. Formalmente, $C_b \varphi$ expresa que φ es conocimiento común para el grupo de agentes *B*, y $M,w \vDash C_b \varphi$ se verifica si y solo sí, para toda w',

$$(w,w') \in (\cup_{c \in B} R_c)^t \quad \text{implica} \quad M,w' \vDash \varphi.$$

donde R^t representa el cierre de equivalencia de la relación *R*. Es decir, se añaden los mínimos pares a *R* que hacen que la relación resultante sea una relación de equivalencia. De este modo, φ será conocimiento común para *B* cuando sea verdadera en todos los estados a los que se puede acceder iterando cualquier número arbitrario de accesibilidad para todos los agentes de *B*. La relación $\cup_{c \in B} R_c$ es la que se usa para el conocimiento general, sirve para modelar aquello que conocen todos los agentes. Ahora, al hacer el cierre de equivalencia, tan solo será conocimiento común aquello que todos conocen que todos conocen que ... que todos conocen.

Terminamos mostrando el sistema axiomático para la lógica *S5C*, que presupone relaciones de equivalencia para la accesibilidad y emplea el operador de conocimiento común:

Todas las tautologías proposicionales

$K_a(\varphi \to \psi) \to (K_a \varphi \to K_a \psi)$	distribución de K_a sobre $\to$
Si φ y $\varphi \to \psi$, entonces ψ	Modus ponens
Si φ, entonces $K_a \varphi$	necesitación de K_a
$K_a \varphi \to \varphi$	Verdad
$K_a \varphi \to K_a K_a \varphi$	Introspección positiva
$\neg K_a \varphi \to K_a \neg K_a \varphi$	Introspección negativa
$C_B(\varphi \to \psi) \to (C_B \varphi \to C_B \psi)$	distribución de C_B sobre $\to$

DOI: https://doi.org/10.29092/uacm.v20i53.1036

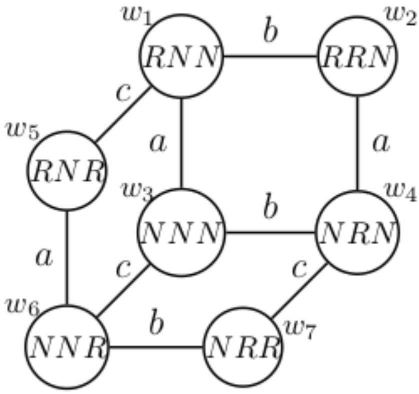
$C_B\,\varphi \rightarrow (\varphi \wedge E_B\,C_B\,\varphi)$	Mezcla
$C_B\,(\varphi \rightarrow E_B\,\varphi) \rightarrow (\varphi \rightarrow C_B\,\varphi)$	Inducción de C_B
Si φ, entonces $C_B\,\varphi$	necesitación de C_B

Anuncios públicos

Mientras nos entreteníamos con los formalismos, Ana y Belén han llegado a casa de Carmen. Ahora ya saben que las dos llevaban regalos para Carmen. Su conocimiento ha cambiado, ya ninguna de las dos considera que ni p (Ana ha comprado un regalo para Carmen) ni q (Belén ha comprado un regalo para Carmen) puedan ser falsas. Saben que $p \wedge q$ es verdadera, y además eso es conocimiento común. El mejor modelo epistémico para representar tal conocimiento tendría un solo estado con p y q verdaderas. Pero, ¿cómo han llegado Ana y Belén a tener ese conocimiento? ¿La lógica ofrece herramientas que permitan modelar cómo va cambiando su conocimiento? Desde luego, a eso se dedica la lógica epistémica dinámica. Dinámica porque estudia no el conocimiento de forma estática (un modelo epistémico fijo) sino el modo en que el conocimiento de uno o varios agentes cambia mediante acciones epistémicas. En esta sección vamos a presentar la acción epistémica más sencilla (pero no por ello poco importante), que es el anuncio público.

Resulta que Carmen ha preparado una sorpresa para sus amigas. Mientras las esperaba había estado leyendo un libro de acertijos lógicos (van Ditmarsch, Kooi, 2015) y ha decidido poner uno en práctica. Se conoce como "Los tres reyes sabios", pero, ¿por qué no plantearlo como "las tres amigas sabias"? Así que muestra a sus amigas tres sombreros negros y dos rojos. Las tres amigas se taparán los ojos y, una vez mezclados los sombreros, cada una de ellas se colocará uno al azar. Los sombreros sobrantes se guardan sin que sean vistos por ninguna de las tres. Al destaparse los ojos, cada una de ellas puede ver los colores de los sombreros de las otras dos, pero no el propio.

DOI: https://doi.org/10.29092/uacm.v20i53.1036



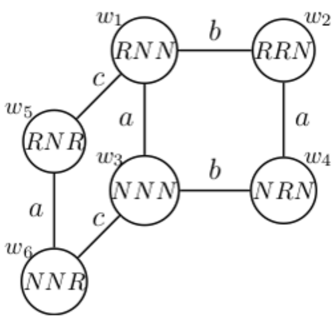
La figura de arriba muestra un modelo epistémico que representa el conocimiento que Ana (*a*), Belén (*b*) y Carmen (*c*) tienen tras descubrir sus ojos. Como no sabemos cuál es la distribución de sombreros, no podemos marcar un estado como distinguido. Las letras *R* y *N* representan los colores de los sombreros; así, *NRN* que vemos en w_4 representa que Ana tiene un sombrero negro (la *N* inicial), Belén tiene un sombrero rojo (la *R* central) y Carmen un sombrero negro (la *N* final). Formalmente, podemos usar tres proposiciones *p*, *q* y *r* para indicar los colores de los sombreros de Ana, Belén y Carmen, respectivamente, considerando que a cada valor de verdad corresponde un color. De momento, no necesitamos formalizar tanto. La accesibilidad viene indicada por las líneas que unen distintos estados. Consideramos que las relaciones de accesibilidad son relaciones de equivalencia, pero por aligerar el diagrama no representamos las líneas que pueden ser presupuestas.

Como vemos, cada agente no puede distinguir entre dos estados en que el único color que cambie sea el de su propio sombrero. Así, Ana no puede distinguir entre w_2 (*RRN*) y w_4 (*NRN*) porque los colores de los sombreros de Belén y Carmen son iguales en ambos estados (rojo y negro, respectivamente) y ella no sabe si el suyo es rojo (y el estado real sería w_2) o es negro (el estado real sería w_4). Además, si Ana estuviera dudando entre w_2 y w_4, sabe que Belén no conoce el color de su sombrero, porque en ambos casos hay un estado diferente accesible para Belén (w_1 desde w_2 y w_3 desde w_4). Sin embargo, en tal caso de dudar entre w_2 y w_4, Ana no sabe si Carmen conoce

DOI: https://doi.org/10.29092/uacm.v20i53.1036

o no el color de su sombrero, dado que en w_2 Carmen sí conocería que su sombrero es negro (Carmen observaría dos sombreros rojos de Ana y Belén, y no hay un tercer sombrero rojo), pero en w_4, siendo también negro el sombrero de Carmen, no puede saberlo porque le resulta indistinguible de w_7, en que su sombrero es rojo. Estos comentarios solo valen desde el punto de vista de Ana en caso de que el estado real fuera w_2 o w_4, pero el modelo ofrece muchas más posibilidades. Como solo hay dos sombreros rojos, cuando una de las tres amigas observa que los sombreros de sus compañeras son rojos, sabe que el suyo es negro. Eso ocurre en w_2 (para Carmen), w_5 (para Belén) y w_7 (para Ana). Por ello, estos estados solo tienen líneas que indican accesibilidad a estados diferentes para dos de las tres agentes.

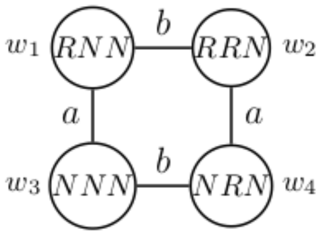
En esta situación, Ana, como en la versión original del acertijo, dice "yo no sé el color de mi sombrero". Con ello está realizando un anuncio público, es decir, ofrece una información que las tres agentes pueden tomar como verdadera. Si el color del sombrero de Ana viniera determinado por la proposición p, su anuncio se correspondería con la fórmula $\neg K_a p \wedge \neg K_a \neg p$. El efecto de este anuncio público es modificar el modelo epistémico.



Como vemos, la diferencia del nuevo modelo con respecto al original es que se ha eliminado el estado w_7, único en que Ana conocía el color de su sombrero (negro). Se ha eliminado porque (bajo la suposición, no trivial, de que Ana no miente en su anuncio) la información que da Ana solo es verdadera en el resto de estados. Como podemos observar, la eliminación de w_7 ofrece a Belén y Carmen nuevas opciones de aprender el color de sus sombreros. Al descartar la opción *NRR*, en el estado w_4 ahora Carmen aprendería el

DOI: https://doi.org/10.29092/uacm.v20i53.1036

color de su sombrero tras el anuncio de Ana. Lo mismo ocurre a Belén en w_6. Sin embargo, Belén, como en el acertijo original, anuncia que ella tampoco conoce el color de su sombrero. Esto tiene el efecto de eliminar los dos estados, w_5 y w_6, en que ahora Belén conocería el color de su sombrero, y vuelve a reducirse el modelo:



Finalmente, Carmen dice "ya sé el color de mi sombrero". En el acertijo original se trataba de descubrir cuál era el color de su sombrero y cómo era posible que Carmen lo hubiera aprendido tras los anuncios de Ana y Belén. En el modelo epistémico resultante podemos ver que en los cuatro estados resultantes el sombrero de Carmen es negro y que en todos ellos lo conoce.

Podemos introducir una variante del acertijo, si Carmen dijera que aprendió el color de su sombrero tras el anuncio de Belén, ¿qué más sabríamos? ¿Y si lo aprendió tras el anuncio de Ana?

Formalmente, la lógica de anuncios públicos (Baltag *et al.*, 1999) introduce en el lenguaje un operador $[\varphi]\psi$ que se lee "tras el anuncio público de φ se verifica ψ". En el acertijo, tenemos que, dado el modelo inicial, $[\varphi_a][\varphi_b]\psi$, donde φ_a es el anuncio que hace Ana (por ejemplo $\neg K_a p \wedge \neg K_a \neg p$, si p representa el color del sombrero de Ana), φ_b el anuncio de Belén ($\neg K_b q \wedge \neg K_b \neg q$) y ψ la fórmula que expresa que Carmen conoce el color de su sombrero ($K_c r \vee K_c \neg r$, por ejemplo).

Para evaluar una fórmula como $[\varphi]\psi$ necesitamos especificar formalmente cómo queda un modelo epistémico $M=\langle S,(R_e),V\rangle$ tras el anuncio público de φ. Formalmente $M_{|\varphi}=\langle S',(R'_e),V'\rangle$ es el resultado de anunciar φ en M, y se compone de:

- $S'=\{s\in S \mid M,s\models\varphi\}$. Es decir, los estados de $M_{|\varphi}$ son los estados de M en que φ es verdadera. Por tanto, el modelo se reduce eliminando los estados en los que φ era falsa.

DOI: https://doi.org/10.29092/uacm.v20i53.1036

- Para cada $e \in A$, $R'_e = R_e \cap (S' \times S')$. Dado que se reducen los estados, la accesibilidad R_e de cada agente e queda limitada a los estados de $\mathcal{M}_{|\varphi}$.
- Para cada $\lambda \in P$, $V'(\lambda) = V(\lambda) \cap S'$. Es decir, en los estados de $\mathcal{M}_{|\varphi}$ siguen siendo verdaderas las mismas proposiciones que en $\mathcal{M}$.

Ya tenemos los elementos para evaluar una fórmula como $[\varphi]\psi$. Dado un modelo epistémico $\mathcal{M}$ y un estado w del mismo, tenemos que $\mathcal{M},w \vDash [\varphi]\psi$ si y solo si

$$\mathcal{M},w \vDash \varphi \quad \text{implica} \quad \mathcal{M}_{|\varphi},w \vDash \psi.$$

De modo que una fórmula como $[\varphi]\psi$ (tras anunciar φ se verifica ψ) es verdadera en w si y solo si, en caso de que φ sea verdadera en w (lo que se anuncia es verdad), tras obtener $\mathcal{M}_{|\varphi}$ (básicamente eliminar de $\mathcal{M}$ los estados que no hacen verdadera φ), se verifica ψ. Hay varias cosas que vale la pena observar. Lo primero es qué pasa si φ no es verdadera en w. Entonces, por definición $[\varphi]\psi$. Eso significa que si φ es algo falso, tras anunciarlo se verifica cualquier cosa. Ello es así porque un anuncio φ falso en w no puede ser evaluado en w porque el modo de construir $\mathcal{M}_{|\varphi}$ elimina w del conjunto de estados. Existen otras lógicas que sí permiten anunciar mentiras (van Ditmarsch, 2014), produciendo modelos en los que algunos agentes (que saben que se ha anunciado una mentira) obtendrán una información distinta de otros (que creen que lo anunciado es verdad). Pero a nivel de conocimiento, no podemos considerar anuncios públicos falsos. Para no trivializar el anuncio público, existe el operador dual $\langle\varphi\rangle\psi$, que equivale a $\neg[\varphi]\neg\psi$, lo que significa que $\mathcal{M},w \vDash \langle\varphi\rangle\psi$ si y solo si

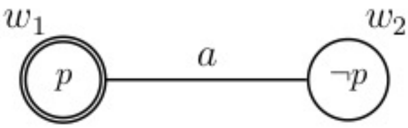
$$\mathcal{M},w \vDash \varphi \quad \text{y} \quad \mathcal{M}_{|\varphi},w \vDash \psi.$$

Por tanto, $\mathcal{M},w \vDash \langle\varphi\rangle\psi$ requiere que φ sea verdadera en ($\mathcal{M},w$) y que tras su anuncio se verifique ψ en w.

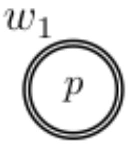
Otra observación importante es que podríamos intuir que tras anunciar φ, obtenemos $K_c\,\varphi$ para cada agente c, incluso que φ se vuelve conocimiento común. Sin embargo, esto solo es cierto si φ es una fórmula puramente proposicional, sin operadores epistémicos. En el caso de fórmulas epistémicas,

DOI: https://doi.org/10.29092/uacm.v20i53.1036

es posible que tras anunciarlas se vuelvan falsas, y por tanto no se conviertan en conocimiento. Este fenómeno es una variante de lo que se conoce como paradoja de la incognoscibilidad (Salerno, 2008): hay verdades que no pueden ser conocidas. El caso más simple se observa en el siguiente modelo:



Tenemos que en w_1 resulta verdadera la fórmula $p\wedge\neg K_a p$ (es verdad p pero Ana no lo sabe). Sin embargo, Ana no podrá conocer $p\wedge\neg K_a p$, pues los axiomas que hemos presentado hacen de $K_a (p\wedge\neg K_a p)$ una inconsistencia, dado que de ahí se sigue $K_a p$ y $\neg K_a p$, que es una contradicción. Pero como $p\wedge\neg K_a p$ es verdadera en w_1 sí que se puede anunciar, y obtenemos el modelo siguiente:



Ahora, Ana conoce p, por lo que se ha vuelto falsa $p\wedge\neg K_a p$ justo tras su anuncio. Es como si al llegar Ana a casa de Carmen, Belén dijera "yo traigo un regalo para Carmen y no lo sabes". Eso era verdad en nuestro ejemplo, pero si Belén lo anuncia se vuelve falso, simplemente porque ahora Ana sí sabrá que Belén trae un regalo para Carmen.

La lógica de anuncios públicos que hemos presentado se conoce como *PAL*, y si añadimos conocimiento común se llama *PAC* y resulta axiomatixable. Además de los axiomas de *S5C* que presentamos más arriba, se incluyen (van Ditmarsch *et al.*, 2007):

$[\varphi]p\leftrightarrow(\varphi\rightarrow p)$	*Permanencia atómica*
$[\varphi]\neg\psi\leftrightarrow(\varphi\rightarrow\neg[\varphi]\psi)$	Anuncios y negación
$[\varphi](\psi\wedge\chi)\leftrightarrow([\varphi]\psi\wedge[\varphi]\chi)$	Anuncios y conjunción

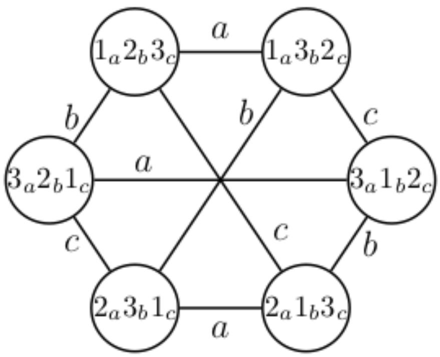
DOI: https://doi.org/10.29092/uacm.v20i53.1036

$[\varphi]\, K\_c\, \psi \leftrightarrow (\varphi \rightarrow K\_a\, [\varphi]\psi)$	Anuncios y conocimiento
$[\varphi][\psi]\chi \leftrightarrow [\varphi \wedge [\varphi]\psi]\chi$	Composición de anuncios
Si φ, *entonces* $[\psi]\varphi$	Necesitación de [ ]
Si $\chi \rightarrow [\varphi]\psi$ *y* $(\chi \wedge \varphi) \rightarrow E\_B\, \chi$, *entonces* $\chi \rightarrow [\varphi]\, C\_B\, \psi$	Anuncios y conocimiento común

Modelos de acción

Además de anuncios públicos, son posibles otros tipos de acciones epistémicas que no suponen la distribución de la misma información a todos los agentes de un grupo. Por ejemplo, si jugando a las cartas, Ana enseña a Belén sus cartas a la vista de Carmen, Belén aprende las cartas de Ana, pero Carmen no. Sin embargo, Carmen aprende que cualesquiera que sean las cartas de Ana, Belén las conoce. El modo en que se puede modelar este tipo de acciones epistémicas es mediante modelos de acción (Baltag, 1999). Vamos a ver un pequeño ejemplo que muestra qué es un modelo de acción y cómo funciona.

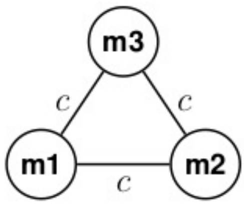
Supongamos que Ana (a), Belén (b) y Carmen (c) juegan un sencillo juego de cartas. De un mazo mezclado de tres cartas (1, 2 y 3), cada una toma una carta en secreto, y la mira sin mostrarla a las demás. El modelo epistémico que representa el conocimiento que cada una obtiene al ver su carta, es el siguiente:



DOI: https://doi.org/10.29092/uacm.v20i53.1036

Las etiquetas que encontramos en cada estado indican el reparto resultante. Así, $1_a 2_b 3_c$ indica que Ana tiene la carta 1, Belén la 2 y Carmen 3. Podríamos usar variables proposicionales para representar estos estados. La accesibilidad indica, como sabemos, la incertidumbre de cada agente. Así, $1_a 2_b 3_c$ es indistinguible para Ana de $1_a 3_b 2_c$, dado que en ambos repartos Ana tiene la carta 1 y no distingue cuál de las cartas 2 y 3 tiene cada una de sus amigas.

Vamos a ver cómo podemos modelar la acción de que Ana le enseña su carta a Belén. La herramienta que permite hacerlo es un modelo de acción. Parece un modelo epistémico, pero los estados representan acciones, a las que hay asociadas precondiciones (que vienen dadas por fórmulas que deben verificarse para que la acción pueda realizarse) y relaciones de accesibilidad. Veamos:



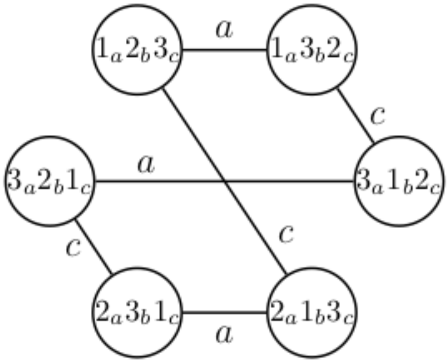
$$pre(\mathbf{m1}) = 1_a$$
$$pre(\mathbf{m2}) = 2_a$$
$$pre(\mathbf{m3}) = 3_a$$

El diagrama superior es un modelo de acción. Tiene tres acciones distintas, que Ana muestre a Belén el 1 (acción m1), que muestre el 2 (m2) o que muestre el 3 (m3). Se trata de las tres posibilidades en que se materializa el hecho de que Ana enseñe a Belén su carta. A cada acción está asociada una precondición, que es una fórmula que debe verificarse para que la acción se pueda ejecutar. Así, pre(m1)=1_a, es decir, la precondición de que Ana muestre el 1 a Belén es que Ana tenga el 1 (proposición 1_a). Como vemos, hay accesibilidad epistémica entre las diferentes acciones, en este caso c no puede distinguir entre ellas, porque no conoce la carta de Ana ni puede verla cuando Ana la muestra a Belén. Sin embargo, Ana y Belén sí que distinguen entre las diferentes acciones.

La operación principal en la lógica de los modelos de acción es ejecutar una acción (modelo de acción) sobre un modelo epistémico. Lo que resulta es un nuevo modelo epistémico cuyos estados son aquellos pares (w,m) del producto cartesiano de estados del modelo epistémico inicial por el conjun-

DOI: https://doi.org/10.29092/uacm.v20i53.1036

to de acciones del modelo epistémico, tales que la precondición de la acción (*pre(m)*) es verdadera en el estado del modelo original ($M,w \vDash pre(m)$). En nuestro caso, en cada estado del modelo epistémico solo será aplicable una acción, dado que la precondición de cada acción es la carta de Ana. La accesibilidad en el modelo epistémico resultante es tal que (w,m) es indistinguible de (w',m') para el agente e si y solo sí w era indistinguible de w' para e, así como m y m'. El modelo epistémico que resulta en nuestro ejemplo es:



Como vemos, Belén ha aprendido cuál es el reparto de cartas, cualquiera que este sea. ¿Y no ha cambiado el conocimiento de Ana y Carmen? Por supuesto que sí. Su conocimiento proposicional no ha cambiado, conocen las mismas fórmulas de la lógica proposicional. Pero ahora saben que Belén conoce cuál es el reparto de cartas, y eso es conocimiento común, dado que

$$C_{\{a,b,c\}}((K_b 1_a \vee K_b 2_a \vee K_b 3_a) \wedge (K_b 1_b \vee K_b 2_b \vee K_b 3_b) \wedge (K_b 1_c \vee K_b 2_c \vee K_b 3_c))$$

es una fórmula válida en el modelo resultante.

Existen variantes de los modelos de acción que permiten modificar las variables proposicionales tras la ejecución de una acción. Eso permite, por ejemplo, que Ana y Belén cambien sus cartas a la vista de Carmen.

Conocimiento y creencia en modelos de plausibilidad

Los modelos de plausibilidad (Baltag, Smets, 2008) permiten, de modo elegante, combinar conocimiento y creencias. Esto resulta útil, por ejem-

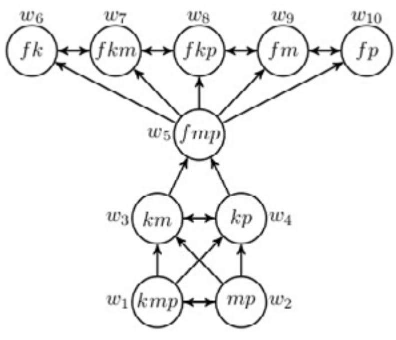
DOI: https://doi.org/10.29092/uacm.v20i53.1036

plo, cuando queremos distinguir el conocimiento que tiene un agente de las conjeturas que realiza por ejemplo en contextos de razonamiento no deductivo (Nepomuceno *et al.*, 2022). El conocimiento se evalúa tal como hemos visto en los modelos epistémicos anteriores: es conocimiento de un agente aquello que es verdadero en todos los estados accesibles, pero existe un orden subjetivo (cada agente tiene uno propio) entre estados que determina las creencias. Cada agente creerá aquello que es verdadero en los estados más plausibles.

Para ilustrar esta lógica, supongamos que Ana y Belén preparan un pastel de fruta sorpresa para Carmen por su cumpleaños. La dan la siguiente información a Carmen:

- Han podido usar cuatro tipos de fruta: fresa (f), kiwi (k), manzana (m) y plátano (p).
- Han hecho una tarta usando al menos dos tipos de fruta, y a lo sumo tres.
- Con esta información, Carmen considera diez posibles combinaciones de frutas que Ana y Belén pueden haber empleado en la tarta. Pero además, Carmen cree que:
- Usarán fresa, porque es su fruta favorita.
- Si usan manzana, no usarán plátano, porque Ana piensa que no combinan bien, aunque Belén no está muy de acuerdo.

El conocimiento y la creencia de Carmen acerca de las frutas que sus amigas han usado para elaborar el pastel se puede representar por el siguiente modelo de plausibilidad:



DOI: https://doi.org/10.29092/uacm.v20i53.1036

En este modelo observamos diez estados que representan las combinaciones de frutas que Ana y Belén han podido utilizar. En este caso, representamos en cada estado solo las proposiciones verdaderas. Así, en w_4, por ejemplo, se usa kiwi y plátano pero no fresa ni manzana. Vemos que entre los estados hay flechas, que indican un cierto orden entre ellos. Suponemos transitividad en el orden (si hay una flecha de w_i *a* w_j y otra de w_j a w_k, también existe entre w_i y w_k, aunque no aparezca explícitamente representada), así como reflexividad (cada estado está conectado consigo mismo). Arriba del todo tenemos los estados w_6–w_{10} en que se cumplen las dos creencias de Carmen (todas las opciones incluyen fresa y si tienen manzana, no tienen plátano). Un poco por debajo se encuentra w_5, en que se cumple solo la creencia más fuerte de Carmen, que el pastel tenga fresa. Más abajo están w_3 y w_4, en que se cumple solo la creencia menos segura (si el pastel lleva manzana, no lleva plátano), y finalmente, los estados menos plausibles son w_1 y w_2, que no verifican ninguna de las creencias de Carmen. Por supuesto, hay combinaciones de frutas (aquellas con menos de dos tipos o más de tres) que no aparecen en el modelo, porque Carmen las descarta, dado que por la información que sus amigas le han dado sabe que no son posibles.

Simplificando a un solo agente, el lenguaje de la lógica de plausibilidad se define como,

$$\varphi ::= \lambda \mid \neg\varphi \mid \varphi\wedge\varphi \mid \varphi\vee\varphi \mid \varphi\rightarrow\varphi \mid \langle\leq\rangle\varphi \mid \langle\sim\rangle\varphi$$

para cualquier variable proposicional $\lambda\in P$. Las fórmulas de tipo $\langle\leq\rangle\varphi$ se leen “hay un estado al menos tan plausible como el actual donde se verifica φ”, y las fórmulas como $\langle\sim\rangle\varphi$ se leen “hay un estado epistémicamente indistinguible del actual donde se verifica φ”.

Las dos modalidades $\langle\leq\rangle$ *y* $\langle\sim\rangle$, permiten definir las nociones de creencia y conocimiento, que como comentamos dependen del orden de plausibilidad que el agente establece entre los estados del modelo. Formalmente, un modelo de plausibilidad es una estructura $M=\langle S,\leq,V\rangle$ donde

- S es un conjunto no vacío de estados.
- $\leq\subseteq(S\times S)$ es la relación de plausibilidad, y representa el orden de credibilidad que el agente establece entre los estados. Leemos $w\leq u$ como “u es al menos tan plausible como w”. Entre las propiedades

DOI: https://doi.org/10.29092/uacm.v20i53.1036

de ≤ está ser una relación reflexiva y transitiva. Además, es localmente conexa, lo que implica que siempre que dos elementos sean comparables a un tercero, esos dos elementos sean comparables entre sí, es decir: para todo $w,w_1,w_2 \in S$, si se dan Rww_1 o $Rw_1 w$ y también Rww_2 o $Rw_2 w$, entonces tienen que darse $Rw_1 w_2$ o $Rw_2 w_1$. Además, se requiere que no exista ninguna cadena R'-ascendente de longitud infinita, siendo R' la versión estricta de R, es decir $R'wu$ si y solo si Rwu y no Ruw.
- V es una función que asigna a cada variable proposicional el conjunto de estados en que es verdadera.

Representamos mediante (M,w) un modelo de plausibilidad M con un *estado distinguido* $w \in W$.

A partir de la relación de plausibilidad ≤ podemos definir las creencias del agente como aquellas fórmulas que son verdaderas en los estados más plausibles. Para definir el conocimiento, necesitamos una relación de accesibilidad ~ que divide el conjunto S en clases de equivalencia de estados *indistinguibles*, tal como es habitual en lógica epistémica. Así, el agente conocerá φ en el estado w si y solo si φ es verdadera en todos los estados indistinguibles de w. Dadas las propiedades de ≤, podemos definir la *relación epistémica de indistinguibilidad* ~ como la unión de ≤ y su inversa, es decir, $\sim = \leq \cup \geq$. Así, aunque un estado u sea más plausible que w, $w \leq u$, ambos son epistémicamente posibles para el agente, y por ello la preferencia de u no excluye la posibilidad de w, por lo que a efectos de conocimiento debemos considerar por igual w y u.

Ahora podemos ver cómo se evalúa una fórmula en un modelo de plausibilidad. Sea (M,w) un modelo $M=\langle S,\leq,V \rangle$ con el estado distinguido $w \in S$. Indicamos mediante $(M,w) \vDash \psi$ que la fórmula es verdadera en el estado w de M. Las fórmulas proposicionales se evalúan de la forma habitual. En cuanto a las nuevas conectivas:

- $(M,w) \vDash \langle \leq \rangle \varphi$ si y solo si existe un $u \in S$ tal que $w \leq u$ y $(M,u) \vDash \varphi$. Es decir, hay un estado al menos tan plausible como w en que φ es verdadera.
- $(M,w) \vDash \langle \sim \rangle \varphi$ si y solo si existe un $u \in W$ tal que $w \sim u$ y $(M,u) \vDash \varphi$. Es decir, hay un estado epistémicamente indistinguible de w (ya sea más o menos plausible) en que φ es verdadera.

DOI: https://doi.org/10.29092/uacm.v20i53.1036

Los operadores $[\leq]$ y $[\sim]$ se definen como los duales de $\langle\leq\rangle$ y $\langle\sim\rangle$, respectivamente. Esto es, $[\leq]\varphi\equiv\neg\langle\leq\rangle\neg\varphi$ y $[\sim]\varphi\equiv\neg\langle\sim\rangle\neg\varphi$. Mediante $[\leq]\varphi$ expresamos que en todo estado al menos tan plausible como el actual se verifica φ. De forma análoga, $[\leq]\varphi$ se interpreta como que en todo estado indistinguible del actual se verifica φ.

Los nuevos operadores permiten definir las nociones de conocimiento y creencia. El conocimiento funciona igual que en los modelos epistémicos anteriores. El agente conoce φ ($K\varphi$) cuando se trata de una fórmula verdadera en todos los estados epistémicamente indistinguibles del actual. Formalmente, $K\varphi\equiv[\sim]\varphi$. En cuanto a la creencia de φ (que se representa como $B\varphi$), se trata de que en todos los estados más plausibles para el agente se verifique φ. Entre el estado distinguido y los estados más plausibles puede haber estados en que φ sea falsa, pero a partir de cierto momento solo encontraremos estados que hagan φ verdadera. Formalmente, $B\varphi\equiv\langle\leq\rangle[\leq]\varphi$.

En el modelo de plausibilidad de nuestro ejemplo, Carmen cree que el pastel tiene fresa, lo que podemos expresar como Bf, que equivale a que f es verdadera en todos los estados de la capa superior del modelo. Igualmente cree que $m\rightarrow\neg p$. Pero aunque crea estas proposiciones, no las conoce, porque hay estados que las hacen falsas, pero son estados menos plausibles que los que las verifican. Sí que conoce que sus amigas no han usado las cuatro frutas $K\neg(f\wedge k\wedge m\wedge p)$ y que han usado al menos una $(K(f\vee k\vee m\vee p))$, incluso dos.

Conocimiento y creencia se pueden modificar mediante diferentes acciones. La primera que vamos a presentar, la observación, modifica el conocimiento. Funciona como un anuncio público, pues consiste en eliminar todos los estados donde la fórmula observada no se satisface, de modo que se reduce el dominio del modelo. Formalmente, sea $M=\langle S,\leq,V\rangle$ un modelo de plausibilidad. La operación de observación de ψ produce el modelo $M_{\psi!}=\langle S',\leq',V'\rangle$ donde $W'=\{w\in S|(M,w)\vDash\psi\}$, $\leq'=\leq\cap(S'\times S')$ y, para cada proposición λ, $V'(\lambda)=V(w)\cap S'$. En resumen, la operación elimina estados de W, dejando solo aquellos que satisfacen la ψ observada. La relación de plausibilidad queda restringida a los estados que sobreviven.

Otra operación que podemos hacer es modificar la relación de plausibilidad. Esto se puede hacer de diversas formas. La operación que llamamos conjetura es conocida como radical upgrade en la literatura. Dado un modelo de plausibilidad $M=\langle S,\leq,V\rangle$, la conjetura de ψ produce el modelo $M_{\psi\Uparrow}=\langle S,\leq',V\rangle$,

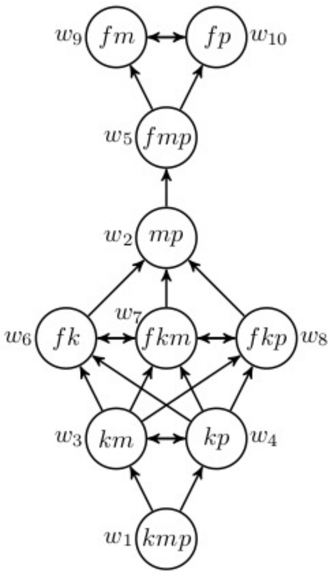
DOI: https://doi.org/10.29092/uacm.v20i53.1036

que solo difiere de M en la relación de plausibilidad, que ahora es:

$$
\begin{aligned}
\leq' = \;& \{(w,u) \mid w \leq u \ \text{"y"}\ (M,u) \vDash \psi\} \cup \\
& \{(w,u) \mid w \leq u \ \text{"y"}\ (M,w) \vDash \neg\psi\} \cup \\
& \{(w,u) \mid w \sim u \ \text{"y"}\ (M,w) \vDash \neg\psi \ y \ (M,u) \vDash \psi\}
\end{aligned}
$$

La nueva relación de plausibilidad indica que después de conjeturar ψ, todos los ψ-estados devienen más plausibles que todos los $\neg\psi$-estados. El orden que hubiera previamente dentro de los ψ-estados o dentro de los $\neg\psi$-estados no cambia (van Benthem, 2007). Esta operación preserva las propiedades de la relación de plausibilidad, tal como se muestra en (Velázquez-Quesada, 2010).

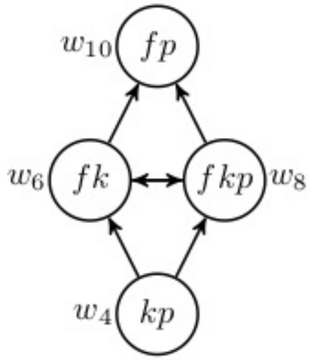
Para ilustrar las operaciones en modelos de plausibilidad, supongamos que Carmen oye una conversación entre Ana y Belén por la que cree que va a venir Diana a la fiesta, a quien no le gusta el kiwi. Entonces, al conjeturar $\neg k$ se modifica la relación de plausibilidad, favoreciendo los estados en que el pastel no tiene kiwi. El modelo resulta:



Ahora, Carmen cree que la tarta tiene manzana o plátano ($B(m \vee p)$), cosa que originalmente no creía. Pero supongamos que sus amigas le dan una

DOI: https://doi.org/10.29092/uacm.v20i53.1036

pista, diciéndole que la tarta no tiene manzana. Lo modelamos como la observación de $\neg m$, y el efecto que tiene es eliminar los estados en que está presente la manzana, resultando el modelo:



En este último modelo, Carmen cree que la tarta tiene fresa y plátano, $Bf \wedge Bp$, pero aún no es conocimiento, $\neg Kf \wedge \neg Kp$. Sin embargo, ha aprendido que la tarta tiene fresa o plátano, $K(f \vee p)$, algo que anteriormente no sabía.

Fuentes consultadas

Alchourrón, C., Gärdenfors, P. y Makinson, D. (1985). On the Logic of Theory Change: Partial Meet Contraction and Revision Functions. En *The Journal of Symbolic Logic*. Vol. 50. Núm. 2. 510-530. DOI:10.2307/2274239

Baltag, A. (1999). A Logic of Epistemic Actions. En *Proceedings of the Workshop on Foundations and Applications of Collective Agent Based Systems (ESSLLI 99)*. Utrecht University.

Baltag, A., Boddy, R. y Smets, S. (2018). Group Knowledge in Interrogative Epistemology. En H. van Ditmarsch, G. Sandu. *Jaakko Hintikka on Knowledge and Game-Theoretical Semantics*. pp. 131-164. Springer. DOI: 10.1007/978-3-319-62864-6_5

Baltag, A. y Smets, S. (2008). A Qualitative Theory of Dynamic Interactive Belief Revision. En G. Bonanno, W. van der Hoek, M.

DOI: https://doi.org/10.29092/uacm.v20i53.1036

Wooldridge. En *Logic and the Foundations of Game and Decision Theory (LOFT7)*. pp. 13-60. Amsterdam: Amsterdam University Press.

Baltag, A., Moss, L. y Solecki, S. (1999). The Logic of Public Announcements, Common Knowledge and Private Suspicious. En *Technical Report SEN-R9922*. Amsterdam: CWI.

van Benthem, J. (2011). *Logical Dynamics of Information and Interaction*. Cambridge University Press. DOI: 10.1017/CBO9780511974533

van Benthem, J. (2007). Dynamic Logic for Belief Revision. En *Journal of Applied Non-Classical Logics*. Vol. 17. Núm. 2. pp. 129-155. DOI: 10.3166/jancl.17.129-155

Cordón-Franco, A., van Ditmarsch, H., Fernández-Duque, D. y Soler-Toscano, F. (2013). A Colouring Protocol for the Generalized Russian Cards Problem. En *Theoretical Computer Science*. Núm. 495. pp. 81-95. DOI: 10.1016/j.tcs.2013.05.010

van Ditmarsch, H. (2014). Dynamics of Lying. En *Synthese*. Vol. 191. Núm. 5. pp. 745-777. DOI: 10.1007/s11229-013-0275-3

van Ditmarsch, H. y Kooi, B. (2015). *One Hundred Prisoners and a Light Bulb*. Springer. DOI: 10.1007/978-3-319-16694-0

van Ditmarsch, H., van der Hoek, W. y Kooi, B. (2007). *Dynamic Epistemic Logic.* Springer. DOI: 10.1007/978-1-4020-5839-4

Nepomuceno, A., Soler-Toscano, F. y Velázquez-Quesada, F. (2023). Abduction from a Dynamic Epistemic Perspective: Non-omniscient Agents and Multiagent Settings. En L. Magnani. *Handbook of Abductive Cognition*. pp. 251-279. Springer. DOI: 10.1007/978-3-031-10135-9_25

Salerno, J. (2008). *New Essays on the Knowability Paradox*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199285495.001.0001

Soler-Toscano, F. y Velázquez-Quesada, F. (2014). Generation and Selection of Abductive Explanations for Non-Omniscient Agents. En *Journal of Logic, Language and Information*. Vol. 23. Núm. 2. pp. 141-168. DOI: 10.1007/s10849-014-9192-1

DOI: https://doi.org/10.29092/uacm.v20i53.1036

Velázquez-Quesada, F. (2010). Dynamic Epistemic Logic for Implicit and Explicit Beliefs. En O. Boissier, A. Seghrouchni, S. Hassas, N. Maudet. *MALLOW*. CEUR Workshop Proceedings, 627.

Fecha de recepción: 29 de abril de 2023
Fecha de aceptación: 6 de agosto de 2023

DOI: https://doi.org/10.29092/uacm.v20i53.1036