EL CONOCIMIENTO EN GRUPOS DE AGENTES DESDE UN PUNTO DE VISTA LÓGICO*

José Rafael Herrera González
rahego@ull.es

RESUMEN

En este trabajo se presentan algunos desarrollos de la lógica epistémica que resultan particularmente interesantes para la teoría de la acción colectiva. Concretamente, nos centramos en el tratamiento formal de las nociones de *conocimiento conjunto*, *conocimiento común* y *conocimiento implícito o distribuido en un grupo de agentes*. Estas nociones epistémicas nos ayudan a comprender el estado del conocimiento en situaciones en las que están implicados múltiples agentes, lo cual es fundamental para el estudio de la acción colectiva, pues parece claro que los cursos de acción emprendidos van a depender, en gran medida, del tipo de conocimiento del que dispongan los agentes. Finalmente, se sugieren algunas líneas de investigación que pueden contribuir al desarrollo de una adecuada lógica de la acción colectiva.

ABSTRACT

«Group Knowledge since a logical point of view». In this paper we present some developments of Epistemic Logic which are especially interesting in Theory of Collective Action. In particular, we describe the formal treatment of three important notions of group knowledge, namely, «*everyone knows*», *Common Knowledge* and *Implicit or Distributed Knowledge in a group*. These epistemic notions can help us to understand the epistemic states in those situations in which multiple agents are concerned. This seems very important in the study of Collective Action, because the agents' knowledge will determine the diferent courses of action. Finally, some lines for future research are suggested to achieve a suitable Logic of Collective Action.

No has de desear conocer la acción;
has de desear conocer al que actúa.
Kaushitaki UPANISHAD

1. LA LÓGICA DEL CONOCIMIENTO

El estudio formal de las nociones de conocimiento y creencia, tal y como éste se lleva a cabo en la lógica epistémica, contribuye fructíferamente al análisis conceptual de importantes nociones de la epistemología, por lo que su interés filosófico se hizo patente ya desde sus orígenes, allá por los años cincuenta y sesenta del pasado siglo. Asimismo, los resultados alcanzados en este campo han encontrado aplicación en muy diversos ámbitos, entre ellos la teoría de la acción colectiva. En este trabajo nos proponemos presentar algunas de las aportaciones fundamentales de la lógica epistémica al análisis formal de diversas nociones de conocimiento que son centrales para el estudio de contextos en los que están implicados múltiples agentes.

Comenzaremos introduciendo uno de los sistemas axiomáticos más conocidos y estudiados de la lógica epistémica, el *Sistema S5*, que nos servirá de base para la consideración de otros desarrollos formales relacionados con contextos en los que intervienen múltiples agentes. A partir de ahí, podremos presentar el tratamiento formal que desde la lógica epistémica se lleva a cabo respecto a las nociones de *conocimiento conjunto*, *conocimiento común* y *conocimiento implícito*. Estos tres conceptos epistémicos resultan de enorme interés para el análisis de acciones en las que está involucrado un grupo de agentes. Y, como veremos, los elementos formales presentados permiten establecer con claridad los distintos matices que caracterizan dichas nociones epistémicas, lo cual no siempre se consigue desde teorizaciones no formales en las que intervienen los conceptos aquí estudiados.

A lo largo de este trabajo trataremos el conocimiento como un concepto modal, por lo cual nos ceñiremos al marco de la lógica modal epistémica. Y es que tal concepción modal de las nociones epistémicas ha acarreado enormes ventajas teóricas desde los trabajos pioneros de Von Wright y Hintikka hasta nuestros días[1].

Para configurar el lenguaje formal de la lógica epistémica proposicional (LEP), añadimos al lenguaje de la lógica clásica de proposiciones los operadores monarios K y M, de tal forma que si la variable metalingüística φ representa una fórmula bien formada (fbf), también serán fbf las siguientes:

$K_i\varphi$ («el agente i sabe que φ es el caso»)
$M_i\varphi$ («el agente i considera posible que φ es el caso»)

Por otra parte, puesto que nos circunscribimos al ámbito de la lógica *modal* epistémica, para la semántica de nuestra LEP tomamos como base la *semántica kripkeana de mundos posibles*. Así pues, definimos un modelo de la lógica epistémica como una estructura $M=<M, R_1,...,R_h, v>$, donde

1. $M\neq\emptyset$, es un conjunto de estados (o mundos) epistémicamente posibles.
2. $R_i\subseteq M^2$ (para $i= 1, ..., h$), son relaciones de accesibilidad entre mundos o estados epistémicos, una para cada agente. Las propiedades que deben cumplir las relaciones R_i variarán en función del sistema axiomático de la LEP de que se trate[2].

[*] Este artículo ha sido elaborado en el marco del Proyecto de Investigación «La realidad sin velos» (PI 2003/099), financiado por la Consejería de Educación del Gobierno Autónomo de Canarias, de enero de 2004 a junio de 2005, del que Manuel Liz Gutiérrez es su investigador principal.

[1] G.H. Von Wright, *An Essay in Modal Logic*, Amsterdam, North-Holland, 1951; J. Hintikka, *Knowledge and Belief*, Cornell, Cornell University Press, 1962.

[2] A lo largo de este trabajo vamos a indicar que entre dos estados m y n se establece una relación de accesibilidad R_i de una de las siguientes formas: mR_in o $(m,n)\in R_i$. Por otra parte, en el

3. v es una función de evaluación que asigna el valor verdadero o falso a cada fbf en un mundo o estado epistémico (siendo F el conjunto de fbfs, v: Fx$M \longrightarrow \{1,0\}$). La función de evaluación cumple las siguientes condiciones, para cualesquiera $m,n \in M$, $\varphi,\psi \in F$ y variable proposicional p:

(i) $v(p,m)=1$ o $v(p,m)=0$
(ii) $v(\neg\varphi,m)=1$ syss $v(\varphi,m)=0$
(iii) $v(\varphi\longrightarrow\psi,m)=1$ syss $v(\varphi,m)=0$ o $v(\psi,m)=1$
(iv) $v(K_i\varphi,m)=1$ syss $\forall n\in M$: mR_in, $v(\varphi,n)=1$

Esta última cláusula pone de manifiesto que un agente i sabe que φ *es el caso* ($K_i\varphi$), si φ es verdadero en todos los estados epistémicos (mundos posibles) que son compatibles con lo que i sabe.

Por último, decimos que una fórmula φ es *válida* (y lo representamos como $\vdash\varphi$) si, para todo modelo y todo $m\in M$, $v(\varphi,m)=1$.

El *Sistema S5* de la lógica epistémica, que será el que aquí tomaremos como base, consta de los siguientes axiomas (o esquemas de axiomas):

Ax1. Un conjunto suficiente de axiomas para derivar todas las tautologías de la LCP.
Ax2. $[K_ip \wedge K_i(p\longrightarrow q)] \longrightarrow K_iq$ *(Para i=1, ..., h)*
Ax3. $K_ip\longrightarrow p$ *(Para i=1, ..., h)*
Ax4. $K_ip\longrightarrow K_iK_ip$ *(Para i=1, ..., h)*
Ax5. $\neg K_ip\longrightarrow K_i\neg K_ip$ *(Para i=1, ..., h)*

Como reglas de derivación contamos con el *Modus Ponens* (MP), la *Regla de Sustitución* y la *Generalización de K* (*Si* $\vdash\varphi$, *entonces* $\vdash K_i\varphi$, *para i=1, ...,h*). Además, disponemos de la siguiente definición metalingüística: $M_i\varphi =_{df} \neg K_i\neg\varphi$.

Finalmente, el *Sistema S5* es consistente y completo en relación a la semántica definida anteriormente[3].

2. CONOCIMIENTO CONJUNTO Y CONOCIMIENTO COMÚN

Decimos que un grupo de h agentes tiene *conocimiento conjunto* de un hecho φ, si todos y cada uno de los agentes del grupo saben que φ. A su vez, φ es *conocimiento común* en dicho grupo si todos los miembros del mismo saben que φ,

Sistema S5 de la LEP, que será nuestro sistema axiomático de partida, las relaciones R_i son *relaciones de equivalencia*, esto es, *reflexivas*, *simétricas* y *transitivas*.

[3] Para una exposición más detallada de todos estos elementos formales, cf. J.-J. Ch. MEYER & W. VAN DER HOEK, *Epistemic Logic for AI and Computer Science*, Cambridge, Cambridge University

todos saben que todos lo saben, todos saben que todos saben que todos lo saben... y así, *ad infinitum*.

Las nociones de conocimiento conjunto y conocimiento común resultan extraordinariamente útiles para analizar el conocimiento de un grupo de agentes. Quien primero estudió formalmente estos conceptos fue D. Lewis en el contexto del análisis de las convenciones[4]. Así, Lewis señala que para que algo funcione como una convención en un grupo de agentes, no basta con que ese algo forme parte del conocimiento conjunto de dicho grupo, sino que además ha de ser conocimiento común. Un sencillo ejemplo puede aclararnos esto: si el agente *i* cruza el paso de cebra cuando el semáforo está en verde, mientras que *j* detiene su automóvil cuando ve el semáforo en rojo, no es sólo porque *i* y *j* conozcan la convención que rige el funcionamiento de los semáforos, sino que ambos saben que ambos lo saben, y también saben que ambos saben que ambos lo saben, y así indefinidamente. De no ser así, ¿se atrevería *i* a cruzar la calle alguna vez?[5]. Estos conceptos también han tenido aplicación en estudios realizados en el ámbito de la economía, y asimismo intervienen en el entendimiento mismo de discursos donde hay suposiciones implícitas. Lewis sostiene que para que un oyente interprete correctamente lo que se propone comunicarle un hablante, es condición necesaria que toda la información contextual que se precisa para la interpretación sea no sólo conocimiento conjunto entre ambos agentes, sino que debe ser, además, conocimiento común. Respecto a este punto, Halpern plantea el siguiente ejemplo: Supongamos que un agente *i* le pregunta a otro agente *j*: «¿Has visto la película que ponen en el Cine Roxy esta noche?». Pues bien, para que esta pregunta sea apropiadamente interpretada, no basta con que *i* y *j* sepan qué película ponen esta noche en el Cine Roxy, sino que *i* debe saber que *j* lo sabe, *j* debe saber que *i* sabe que *j* lo sabe, etc.[6]. Sobre estos importantes aspectos volveremos con más detalle un poco más adelante, cuando analicemos algunos problemas teóricos muy conocidos, como «el enigma de los chicos embarrados» o «el problema del ataque coordinado», en función de las nociones epistémicas aquí presentadas. Ahora vamos a ocuparnos de los aspectos más formales de la lógica del conocimiento conjunto y del conocimiento común.

Press, 1995, pp. 1-28. Interesantes y breves exposiciones de algunos de los desarrollos fundamentales de la lógica epistémica pueden encontrarse en: J.-J. Ch. Meyer, «Epistemic Logic», in Lou Goble (ed.), *Philosophical Logic*, Blackwell Publishers, 2001, pp. 183-202; y también en W. Lenzen, «Epistemic Logic», in I. Niiniluoto, M. Sintonen & J. Wolenski (eds.), *Handbook of Epistemology*, Kluwer Academic Publishers, 2004, pp. 963-983.

4 D.K. Lewis, *Convention*, Cambridge, MA, Harvard University Press, 1969.

5 A nivel teórico y formal esta concatenación de conocimiento conjunto, en función de la cual se define el conocimiento común, es infinita. En la práctica está claro que el proceso no es infinito, pues actuamos sin necesidad de tener una seguridad total en que una convención de este tipo sea conocimiento común; en caso contrario, nadie habría cruzado un paso de cebra jamás, por mucho que un semáforo indicara que podía hacerlo.

6 J.Y. Halpern, «Reasoning about Knowledge: An Overview», *Proceedings of the 1986 Conference Theoretical Aspects of Reasoning about Knowledge*, Los Altos California, Morgan Kaufmann Publishers, 1986, p. 9.

Necesitamos extender nuestro lenguaje para la LEP, presentado en el apartado anterior, con los operadores monarios E (*«conocimiento conjunto»*) y C (*«conocimiento común»*), de tal modo que si φ es una fbf, también lo son $E\varphi$ *(«todos saben que* φ*»)* y $C\varphi$ (*«φ es conocimiento común»*).

De acuerdo con lo que hemos venido diciendo, tendremos, pues, las siguientes definiciones de nuestros nuevos operadores epistémicos:

$E\varphi =_{df} K_1\varphi \wedge \ldots \wedge K_h\varphi$ (siendo h el número de agentes del grupo).

$C\varphi =_{df} \varphi \wedge E\varphi \wedge EE\varphi \wedge EEE\varphi \wedge \ldots = \forall i \geq 0\ E^i\varphi$ (aquí $E^i\varphi$ es definido como $EEE\ldots\varphi$ *-i* operadores E-. Esta expresión sólo puede entenderse intuitivamente, en tanto que no podemos considerar infinitas conjunciones).

Para la definición semántica de E y C tomamos como base un modelo M como el presentado en el apartado 1, al que sólo necesitamos añadir las dos cláusulas siguientes para la función de evaluación v:

$v(E\varphi,m)=1$ *syss* $\forall n \in M$: mR_1n, $v(\varphi,n)=1$ *y ... y* $\forall n \in M$: mR_hn, $v(\varphi,n)=1$[7]
$v(C\varphi,m)=1$ *syss* $\forall n \in M$: $m\beta n$, $v(\varphi,n)=1$ *(también* $m\beta m$*)*

Si consideramos un único agente (es decir, si $h=1$), el conocimiento conjunto coincide con el conocimiento de dicho agente ($E\varphi \equiv K\varphi$).

Respecto a los nuevos operadores epistémicos E y C, podemos considerar los siguientes axiomas (o esquemas de axiomas)[8]:

Ax6. $Ep \leftrightarrow K_1p \wedge \ldots \wedge K_hp$
Ax7. $Cp \rightarrow p$
Ax8. $Cp \rightarrow ECp$
Ax9. $[Cp \wedge C(p \rightarrow q)] \rightarrow Cq$
Ax10. $C(p \rightarrow Ep) \rightarrow (p \rightarrow Cp)$

El *Ax6* expresa la definición del operador E que hemos visto anteriormente. El *Ax7* nos indica que el conocimiento común establecido en un grupo de agentes versa siempre sobre hechos verdaderos. A su vez, los axiomas *Ax8*, *Ax9* (cierre bajo la implicación del operador C) y *Ax10* recogen la interacción formal entre los operadores E y C.

Además, introducimos la *Regla de Generalización de C*: *Si* $\vdash\varphi$, *entonces* $\vdash C\varphi$.

Como sistema axiomático para las nuevas nociones epistémicas merece destacar el *Sistema S5EC*, que consta del *Sistema S5* (recogido en el apartado 1), más

[7] También podemos expresar la definición semántica de E de las dos formas siguientes:
$v(E\varphi,m)=1$ *syss* $\forall n \in M$: $(m,n) \in R_1 \cup \ldots \cup R_h$, $v(\varphi,n)=1$
$v(E\varphi,m)=1$ *syss* $v(K_1\varphi \wedge \ldots \wedge K_h\varphi,m)=1$

[8] Cf. J.-J. Ch. Meyer & W. van der Hoek, *op. cit.*, p. 46.

los axiomas 6, 7, 8, 9 y 10 anteriores y la *Regla de Generalización de C*. Este nuevo sistema es, asimismo, consistente y completo[9].

Podría pensarse que la noción de conocimiento común es baladí si contamos con la noción de conocimiento conjunto, pues cabría suponer que la primera no añade nada sustancial respecto a la segunda, a no ser una especie de superconciencia colectiva respecto al conocimiento conjunto de un hecho, que nada importante aportaría en la práctica. Pero ya hemos visto cómo Lewis pone de manifiesto que para que algo funcione efectivamente como una convención para un conjunto de agentes, no basta con que dicho grupo tenga conocimiento conjunto de ese hecho, sino que se requiere, además, que sea conocimiento común. El ejemplo del semáforo que comentamos más arriba pretendía ilustrar esta idea. También podemos comprobar esto mediante el análisis de algunos conocidos problemas teóricos, valiéndonos de las nociones epistémicas que acabamos de presentar. Dos de estos problemas son los conocidos como *el enigma de los chicos embarrados* y *el problema del ataque coordinado*, cuyos análisis encontramos en muchos trabajos que estudian el conocimiento característico de grupos de agentes.

2.1. El enigma de los chicos embarrados

El enigma de los chicos embarrados puede plantearse del siguiente modo[10]. Supongamos un número h de chicos que forman un círculo alrededor de su padre. Hay un número k (con $1 \leq \kappa \leq h$) de chicos que tienen la cabeza manchada de barro. Además, cada uno de ellos puede ver a todos los demás, pero no su propia cabeza, por lo cual cada chico sabe quiénes de los otros tienen barro en la cabeza, pero no si él mismo lo tiene. Se supone que los chicos no pueden comunicarse entre ellos, son completamente honestos (al igual que el padre) y bastante inteligentes (al menos por lo que al manejo de nociones epistémicas se refiere). Así las cosas, el padre dice en voz alta: «Hay al menos un chico que tiene barro en la cabeza. ¿Pueden dar un paso al frente todos los que sepan que su cabeza está manchada de barro?».

Si $\kappa>1$ ningún chico da un paso al frente. El padre formula la pregunta nuevamente (diciendo ahora «hay al menos dos chicos ...»). Si $\kappa>2$ tampoco hay respuesta por parte de los chicos. Este proceso se repite hasta que, cuando el padre ha formulado su pregunta κ veces, sorprendentemente todos los chicos que tienen la cabeza embarrada dan un paso al frente.

Las nociones de conocimiento conjunto y conocimiento común nos permiten analizar este problema y comprobar el porqué de su misterioso desenlace. Para ello vamos a describir cada situación o estado mediante una tupla de unos y

---

[9] Cf. J.-J. Ch. Meyer & W. van der Hoek, ibídem, pp. 47-48.

[10] En esta parte de nuestra exposición seguimos, en su mayor parte, el análisis de R. Fagin *et al.*, *Reasoning about Knowledge*, MIT Press, 1995, pp. 3-7 & 24-30; así como de J.-J. Ch. Meyer & W. van der Hoek, *op. cit.*, pp. 56-59.

ceros de la forma $<x_1, ..., x_h>$, donde $x_i=1$ si el chico i tiene la cabeza manchada y $x_i=0$ si no la tiene. Así, por ejemplo, si $h=3$, una tupla de la forma $<1,0,1>$ expresaría que los chicos 1 y 3 tienen la cabeza manchada, mientras que el 2 no la tiene. Supongamos que la tupla anterior describe la situación real. Entonces, antes de que el padre hable, el chico 1 consideraría posibles dos situaciones, descritas por las tuplas $<1,0,1>$ (la situación real) y $<0,0,1>$, puesto que no puede saber si su propia cabeza está o no manchada de barro. De igual modo, el chico 2 considera posibles dos situaciones: $<1,0,1>$ y $<1,1,1>$, y el chico 3 otras dos: $<1,0,1>$ y $<1,0,0>$.

Para nuestro análisis del problema vamos a necesitar un lenguaje formal que nos permita expresar si un chico del grupo tiene o no la cabeza manchada de barro. Por ello, tomamos un conjunto de variables proposicionales $\Phi=\{p_1,...,p_h, q_z, r_z\}$, donde, intuitivamente, p_i expresa que «el chico i tiene la cabeza embarrada», q_z, con $1\leq z\leq \kappa$, expresa que «al menos z chicos tienen la cabeza embarrada» y r_z que «hay exactamente z chicos con la cabeza manchada de barro». Vamos también a considerar una estructura $M=<M, R_i, v>$, donde

1. M es un conjunto de estados o situaciones (cada una de las cuales está descrita por una tupla de la forma que hemos visto), siendo $|M| =2^h$

2. $R_i\subseteq M^2$, son relaciones de accesibilidad entre situaciones, una para cada chico. Si consideramos dos situaciones $m,n \in M$, tales que el chico i considera posible el estado n desde el estado m, es decir, tales que mR_in, significa que las tuplas que representan ambas situaciones concuerdan en todo excepto posiblemente en el componente x_i. Esto supone que las relaciones R_i son relaciones de equivalencia.

3. $v: \Phi x M \longrightarrow \{1,0\}$, es una función de asignación de valores de verdad a las variables proposicionales en cada situación. En nuestro caso, v cumple las tres condiciones siguientes:

(i) $v(p_i,<x_1,...,x_h>)=1$ syss $x_i=1$
(ii) $v(q_z,<x_1,...,x_h>)=1$ *syss* $z\leq \kappa$
(iii) $v(r_z,<x_1,...,x_h>)=1$ *syss* $z=\kappa$

Una vez que el padre enuncia en voz alta q_z, tenemos que Cq_z (y, por tanto Eq_z). Como cada chico embarrado sólo puede ver κ-1 chicos embarrados, si $z<\kappa$ ningún chico con la cabeza manchada de barro será capaz de concluir que él mismo es uno de los que tiene la cabeza manchada y, por consiguiente, ninguno dará un paso al frente; con lo cual tendremos para esa ronda: $C\neg r_z$ (y $\neg Er_z$), y pasaríamos a la ronda siguiente, en la cual el padre enunciaría q_{z+1}. Por lo mismo, si $z=\kappa$ todo chico embarrado sabrá que él lo está y, por tanto, dará un paso al frente.

Así pues, por inducción sobre z (con $1\leq z\leq \kappa$) podemos probar que, tras la ronda número z de enunciados del padre, se obtiene Cq_z (y, por tanto, Eq_z). Entonces, tras la ronda número κ, se da Eq_k, y así todos los chicos saben que al menos κ chicos tienen la cabeza embarrada. Y como cada chico embarrado sólo puede ver κ-1 chicos embarrados, todo el que tenga la cabeza manchada sabrá que la tiene y, así, dará un paso al frente.

Si, como hemos venido suponiendo, la situación real estuviera descrita por la tupla $<1,0,1>$, tendríamos, antes de que el padre hablara, que $v(K_1\neg p_2,<1,0,1>)=1$, puesto que en esta situación el chico 2 no tiene la cabeza manchada de barro en ninguna de las dos situaciones que el chico 1 considera posibles. También se daría que $v(K_1p_3,<1,0,1>)=1$ y $v(\neg K_1p_1,<1,0,1>)=1$. De hecho, es conocimiento común que cada chico sabe si la cabeza de los demás está o no manchada de barro; serían válidas, pues, las dos expresiones siguientes: $C(p_i \rightarrow K_jp_i)$ y $C(\neg p_i \rightarrow K_j\neg p_i)$. Una vez que el padre dice en voz alta q_1, se da Cq_1 y Eq_1, y ninguno de los chicos es capaz de saber si su cabeza está manchada o no, pues 1 y 3 ven un chico manchado y 2 ve dos. En la siguiente ronda, cuando el padre dice q_2, se da Cq_2 y Eq_2. El chico 2 no da un paso al frente, pues ya sabía q_2, mientras que 1 y 3 sí lo dan, pues cada uno de ellos sólo ve un chico manchado de barro y, por tanto, deducen que ellos mismos han de estar manchados. En suma, cuando el padre enuncia q_κ se cumple que K_ir_κ para todo chico *i* que tiene la cabeza manchada de barro, con lo cual dará un paso al frente.

En las situaciones en las que hay dos o más chicos que tienen la cabeza manchada de barro, como ocurre con $<1,0,1>$, se cumple que *Ep* antes incluso de que el padre hable por primera vez. No obstante, después de que el padre diga que q_1, el estado del conocimiento del grupo varía, aun cuando todos los chicos supieran ya que q_1. Así, en la situación $<1,0,1>$ el chico 1 considera posible la situación $<0,0,1>$, y en esta última el chico tres considera posible $<0,0,0>$. Por tanto, en el estado $<1,0,1>$, antes de que el padre hable, aunque todos saben que al menos uno de los chicos tiene barro en la cabeza, el chico 1 considera posible que el chico 3 considera posible que ninguno de ellos tiene la cabeza manchada de barro, lo cual deja de ocurrir una vez que el padre dice q_1.

Por tanto, para que en el enigma de los chicos embarrados se dé el «misterioso» desenlace que hemos comentado, las aserciones del padre han de ser *conocimiento común* entre los miembros del grupo, no bastando con que sean simplemente conocimiento conjunto. Esto significa que las cosas no funcionarían, por ejemplo, si el padre se limitara a decir sus aserciones en el oído de cada chico sin que los demás supieran qué les decía, pues de esta forma nunca se alcanzaría tal conocimiento común.

Por otra parte, aunque la definición formal del conocimiento común encierra una conjunción infinita de operadores *E*, hemos de incidir en que esto no significa que en la práctica se requiera un tiempo infinito para que los chicos adquieran tal conocimiento común, sino que lo adquieren de una sola vez en cada ocasión que el padre hace su anuncio público. Como dijimos más arriba, la concatenación de operadores *E* es infinita a nivel teórico y formal, no así en el funcionamiento práctico del conocimiento común de un grupo de agentes.

2.2. El problema del ataque coordinado

Las nociones epistémicas que venimos presentando, en particular la de conocimiento común, también pueden ser aplicadas al estudio de sistemas distriduidos. El comportamiento de este tipo de sistemas puede ilustrarse mediante el muy discutido *problema del ataque coordinado*. Además, el análisis de este problema nos permite sacar a relucir los aspectos formales-epistémicos involucrados en la coordinación de ciertas acciones simultáneas.

Consideramos ahora una situación en la que hay dos regimientos de un mismo ejército, situados en dos colinas enclavadas a ambos lados de un valle que ocupa el ejército enemigo[11]. Si ambos regimientos atacan simultáneamente, vencerán al enemigo, si no es así, será el enemigo quien venza. Por ello, ninguno de los generales de los regimientos (llamémosles A y B) se decidirá a atacar a menos que esté completamente seguro de que el otro regimiento atacará simultáneamente. El general A pretende coordinar un ataque con el general B. Para conseguir esto, sólo puede ponerse en contacto con él a través de un mensajero que lleve el mensaje p («atacaremos el día D a la hora H»). Pero el mensajero ha de atravesar las líneas enemigas para cumplir su cometido, y el éxito de su empresa no está en absoluto garantizado. Por ello, para estar seguros de que el mensaje p ha sido recibido por el general B, este último ha de enviar un acuse de recibo ($K_B p$) al general A. Como este nuevo mensaje ha de ser enviado también por un mensajero, cuyo éxito tampoco es seguro, para que el ataque coordinado sea posible es necesario que A envíe a B un nuevo acuse de recibo expresando $K_A K_B p$. Pero A necesita recibir ahora una confirmación de que B recibió este último mensaje, es decir, que $K_B K_A K_B p$, pues de lo contrario no habría una garantía total de que los dos regimientos fueran a atacar simultáneamente. Este proceso tendría que repetirse infinitamente, pues siempre cabría la duda de si el último acuse de recibo fue o no recibido. Es decir, nos encontramos con que, a pesar de que es necesario el conocimiento común de p por parte de los generales A y B (para estar seguros de que ambos atacarán simultáneamente), no es posible obtener tal conocimiento común en un tiempo finito.

Fagin *et al.*[12] muestran que el conocimiento común de p (Cp) es necesario para solventar el problema del ataque coordinado, y sin embargo Cp no es obtenible en sistemas distribuidos en los que el acceso a la información pertinente por parte de los distintos componentes del sistema no está garantizado. Es decir, que el problema del ataque coordinado es irresoluble.

Ahora bien, como señala Halpern, para coordinar acciones en la práctica es suficiente con variantes del conocimiento común menos exigentes que la que venimos presentando, para cuyo establecimiento no sería necesaria una certeza absoluta respecto a que el flujo de información sea exitoso, sino que basta con que se tenga una confianza suficiente en que tal cosa ocurre[13]. Algo parecido acontecería en el ejemplo del semáforo que comentábamos más arriba.

---

[11] Para el planteamiento del problema seguimos, en líneas generales, la exposición de éste que encontramos en J.-J. Ch. MEYER & W. VAN DER HOEK, ibídem, pp. 64-65.

[12] *Op. cit.*, p. 140 (teorema 4.5.4) y pp. 177-178.

[13] J.Y. HALPERN, *op. cit.*, p. 11.

3. CONOCIMIENTO IMPLÍCITO O DISTRIBUIDO EN UN GRUPO DE AGENTES

Entre las nociones epistémicas relativas a grupos de agentes destaca la de *conocimiento implícito o distribuido*. Se dice que un grupo de agentes tiene conocimiento implícito o distribuido de un hecho j, si del conocimiento «combinado» de los miembros de dicho grupo se sigue que j[14]. En general, para combinar el conocimiento de los agentes de un grupo eliminamos todos los mundos considerados epistémicamente imposibles por algún miembro del mismo. Esto lo hacemos formalmente como sigue.

Extendemos nuestro lenguaje formal para la LEP con el operador monario I («*conocimiento implícito*»), resultando entonces que si φ es una fbf, también lo es $I\varphi$ («*φ es conocimiento implícito*»). También extendemos el modelo semántico del apartado 1 mediante la siguiente definición de nuestro nuevo operador I:

$$v(I\varphi,m)=1 \text{ syss } \forall n\in M\text{: } (m,n) \in R_1\cap\ldots\cap R_h,\ v(\varphi,n)=1$$

Meyer y van der Hoek[15] recogen el siguiente axioma relativo al conocimiento implícito:

Ax11. $K_i\varphi\longrightarrow I\varphi$ *(Para* $i=1,\ldots, h$*)*

Claro está que este axioma sólo tiene sentido respecto a aquellos grupos de agentes en los que efectivamente se dé un conocimiento implícito, o, dicho de otro modo, respecto a aquellos grupos de agentes cuyos miembros comparten al menos una alternativa epistémica. Además, el *Ax11* se corresponde con la propiedad $R_I\subseteq R_i$, para $1\leq i\leq h$, donde R_I representa la relación de accesibilidad asociada con el operador modal I.

Respecto al conocimiento implícito cabe destacar el sistema axiomático *S5I*, que consta del *Sistema S5* más el *Ax11* y el *Sistema* $S5_I$[16]. El *Sistema S5I* es consistente y completo[17].

La comparación de las definiciones semánticas de los operadores I («*conocimiento implícito*») y E («*conocimiento conjunto*») nos muestra la diferencia existente entre ambas nociones epistémicas. Si afirmamos, respecto a un conjunto de h agentes, que se da $E\varphi$, significa que todos y cada uno de los agentes del grupo, desde 1 hasta h, saben que φ. En cambio, $I\varphi$ indica que tal conocimiento está implícito o

---

14 Cf. Fagin *et al.*, *op. cit.*, p. 24.

15 *Op. cit.*, p. 66.

16 El *Sistema* $S5_I$ resulta de sustituir el operador K por el operador I en cada axioma del *Sistema S5*.

17 Cf. J.-J- Ch. Meyer & W. van der Hoek, *op. cit.*, p. 67.

distribuido en el grupo de agentes, pudiendo darse que ninguno de los agentes del grupo sepa por sí mismo que *j* (es decir: $\neg K_i\varphi$, para $i= 1, ..., h$).

Como señalan Meyer y van der Hoek[18], la intuición última, tras la noción de conocimiento implícito o distribuido, es que en un sistema cerrado un grupo de agentes cooperativos únicamente puede adquirir los conocimientos que están ya implícitos en dicho grupo. Y si los agentes de un grupo tal pudieran mancomunar su conocimiento, sólo podrían considerar como posibles aquellos mundos que fueran alternativas epistémicas para todos y cada uno de los agentes por separado.

Se suele expresar la intuición subyacente a *Ij* como «*el hombre sabio conoce que* φ *es el caso*». Esta idea puede parecer, a primera vista, antiintuitiva, pues de acuerdo con ella cuanto menor sea el número de alternativas epistémicas que considere posibles un agente, más sabio será éste. Y así ha de ser, efectivamente, si tenemos en cuenta la interpretación modal del conocimiento que se mantiene en la LEP, según la cual si un agente *i* sabe que φ, entonces φ es verdadero en todos los mundos que *i* considera epistémicamente posibles. Así pues, si un agente conoce un hecho φ, entonces sólo considera una posibilidad respecto a tal hecho, esto es, aquella en la que se da φ. En cambio, si el agente alberga dudas acerca de tal hecho, contemplará al respecto dos alternativas epistémicas, una en la que se da φ y otra en la que se da $\neg\varphi$, por lo que el número de sus alternativas epistémicas aumenta.

Para ilustrar algunas de las ideas que venimos exponiendo, consideremos un grupo formado por los agentes 1, 2 y 3. Tendremos que contemplar entonces tres relaciones de accesibilidad, una por cada agente, que denotamos por R_1, R_2 y R_3. Como las R_i son relaciones de equivalencia, y, por tanto, reflexivas, el «mundo real» (o actual) será siempre una de las alternativas epistémicas que consideren posibles los agentes 1, 2 y 3. Se da, pues, que $\models(p\wedge q)\longrightarrow M_i(p\wedge q)$ (i= 1, 2, 3); aunque muy bien puede ocurrir que cada agente considere también posible la alternativa $\neg p$ y, por tanto se daría que $\neg K_1(p\wedge q)\wedge\neg K_2(p\wedge q)\wedge\neg K_3(p\wedge q)$. Veamos qué ocurre en cada caso:

(a) Supongamos que los agentes comparten alguna otra alternativa epistémica además de la correspondiente al mundo real. Por ejemplo, el agente 1 considera posible $p\wedge q$, $p\wedge\neg q$ y $\neg p\wedge\neg q$; el agente 2 tiene como alternativas epistémicas $p\wedge q$ y $p\wedge\neg q$; y el agente 3 considera posible $p\wedge q$, $p\wedge\neg q$ y $\neg p\wedge q$; dándose en el mundo real que $p\wedge q$. En función de esto tenemos que: $I[(p\wedge q)\vee(p\wedge\neg q)]$. Y, de acuerdo con lo dicho, podríamos afirmar que «el hombre sabio sabe que $(p\wedge q)\vee(p\wedge\neg q)$». Aquí vemos claramente que «el hombre sabio» (que en este caso coincide con el agente 2) es sabio no porque conozca lo que de hecho ocurre en el mundo real, sino porque sabe más que los demás miembros del grupo. Es decir, que tiene la suficiente información como para no admitir como posible ninguna alternativa epistémica

[18] Ibídem, p. 65.

que sea considerada imposible por alguno de los miembros del grupo. O, dicho de otro modo: tiene más información que los demás, y por eso puede eliminar alternativas no reales que, sin embargo, son consideradas posibles por otros miembros del grupo menos informados.

(b) El otro caso posible sería que la única alternativa epistémica que comparten los agentes del grupo sea la correspondiente al mundo real (en nuestro ejemplo $p \wedge q$). Entonces tendríamos que $I(p \wedge q)$. En este caso, «el hombre sabio» sabe más que los otros miembros del grupo por las mismas razones que antes, pero, además, es más sabio que «el hombre sabio» del apartado anterior, pues su información es tan completa que no considera posibles alternativas epistémicas distintas al mundo real en relación a los hechos *p* y *q*.

4. LÓGICA EPISTÉMICA Y ACCIÓN COLECTIVA. CONSIDERACIONES FINALES

Nociones como las de conocimiento común o conocimiento implícito, que acabamos de analizar, resultan muy útiles a la hora de estudiar los estados epistémicos de un grupo de agentes. De hecho, ambas nociones cuentan entre las más utilizadas para el estudio del conocimiento y la acción en contextos en los que intervienen múltiples agentes[19].

Al teorizar la acción colectiva, el conocimiento de los agentes constituye una de las variables fundamentales a tener en cuenta, pues dependiendo de tal conocimiento pueden seguirse unos u otros cursos de acción. En este trabajo hemos tratado de poner de manifiesto cómo las nociones epistémicas relativas a grupos de agentes pueden verse ampliamente clarificadas desde el paradigma de la lógica epistémica. Claro está que para el estudio de la acción colectiva, especialmente de sus vertientes formales, no basta con el desarrollo de una lógica del conocimiento, sino que también hay que contemplar la lógica de la creencia, la lógica temporal (que nos permita temporalizar estados epistémicos y estados mentales en general) y la lógica de los objetivos y las intenciones. En particular, es posible desarrollar sistemas doxásticos en los que se formaliza las nociones de *creencia*, *creencia conjunta* y *creencia común* a partir de los elementos que hemos presentado para la lógica epistémica[20]. Asimismo, han sido propuestos desarrollos formales que tratan de configurar sistemas lógicos para los objetivos y las intenciones desde una perspectiva modal[21].

---

[19] Meyer y van der Hoek presentan un interesante sistema lógico en el que se incluye el conocimiento conjunto, el conocimiento común y el conocimiento implícito: Cf. J.-J. Ch. MEYER & W. VAN DER HOEK, *A Complete Epistemic Logic for Múltiple Agents: Combining Distributed and Common Knowledge*, Technical Report UU-CS-1996-52, Utrecht University, 1996.

[20] Cf., p.e: S. KRAUS & D. LEHMANN, «Knowledge, Belief and Time», *Theoretical Computer Science*, 58, 1988, pp. 155-174.

[21] Cf., p.e., la propuesta contenida en: J.X. ARRAZOLA, *Acción Colectiva: Bases conceptuales y lógicas*, Tesis Doctoral, Servicio Editorial de la Universidad del País Vasco, 1998, pp. 133-139.

También la lógica temporal nos proporciona interesantes elementos formales, articulándose sistemas que recogen distintas ontologías temporales (para tiempo lineal, ramificado hacia el futuro, etc.). Falta ahora desarrollar sistemas que combinen todos estos elementos en una lógica de la acción colectiva, de tal modo que pueda reflejarse el comportamiento formal de los mismos en situaciones que involucren a grupos de agentes.

En cualquier caso, el análisis lógico del conocimiento en grupos de agentes pone de relieve que hay tipos de conocimiento de grupo (como el conocimiento común y el conocimiento implícito) que no son sólo cuantitativa, sino también cualitativamente diferentes al conocimiento que alberga cada uno de los agentes del grupo por separado. Está claro que la noción epistémica representada por el operador K es muy diferente a la que denotan los operadores C o I, y ello no sólo debido al número de sujetos epistémicos implicados, sino, sobre todo, por la definición misma y el comportamiento lógico-formal de cada una de dichas nociones epistémicas.

Es por ello que la lógica epistémica puede aportar interesantes análisis e intuiciones a la lógica de la acción colectiva para, de este modo, sacar a relucir de forma más precisa y rigurosa las relaciones que se establecen entre conocimiento, creencias, intenciones, decisiones y acciones, todas ellas nociones centrales para la filosofía analítica de la acción.