

Sens et Interprétation

Connaissance commune et croyance commune

P. Égré (CNRS, IJN)
paul.egre@ens.fr

Séance 2 - 25 octobre 2010

1 Convention et connaissance commune

Le concept de connaissance commune est introduit par Lewis dans Convention et sert en réalité à définir la notion de convention.

- (1) Conduire à droite:
- (2)
 - a. Je m'attends à ce que chacun roule à droite
 - b. Je m'attends à ce que x s'attende à ce que je m'attende à ce que chacun roule à droite.
 - c. etc.

X est connaissance commune parmi les membres du groupe P ssi il existe un état de choses A tel que:

- (3)
 - a. Chacun dans la population P a une raison de croire que A
 - b. A indique à chacun dans P que chacun dans P a une raison de croire que A
 - c. A indique à chacun dans P que X .

Une régularité R dans le comportement des membres de P dans une situation récurrente S est une convention ssi il est connaissance commune dans toute instance de S parmi les membres de P que:

- (4)
 - a. Chacun se conforme à R
 - b. Chacun s'attend à ce que chacun se conforme à R
 - c. Chacun préfère se conformer à R à la condition que les autres le fassent, sachant que S est un problème de coordination et que la conformité uniforme à R est un équilibre de coordination en S .

Exemple: R = rouler à droite.

2 Aumann sur la connaissance commune

Cf. Aumann 1976 et Osborne et Rubinstein 1994

2.1 Incertitude et possibilité

- Notion de correspondance de possibilité:

(5) i sait que A en w : $P_i(w) \subseteq A$

- Le modèle de partition de la connaissance

$w \in P_i(w)$ (véridicité de la connaissance)

si $w' \in P_i(w)$ alors $P_i(w') = P_i(w)$ (introspection positive et négative)

2.2 La définition par événement auto-évident de la connaissance commune

- Soit W un ensemble de mondes possibles. Supposons qu'on ait deux agents:

(6) $A \subseteq W$ est auto-évident entre a et b ssi pour tout $w \in A$: $P_i(w) \subseteq A$

- Représentez vous graphiquement la notion d'événement auto-évident entre deux agents. C'est une proposition (ensemble de mondes possibles) telle qu'en toute situation qui la rend vraie, chaque agent sait que cette proposition est vraie.

(7) $W = \{1, 2, 3, 4, 5, 6\}$

$P_a = \{\{1, 2\}, \{3, 4, 5\}, \{6\}\}$

$P_b = \{\{1\}, \{2, 3, 4\}, \{5\}, \{6\}\}$ Soit $A = \{1, 2, 3, 4\}$

Montrer que en 3:

- (8) a. a ne sait pas que A
b. b sait que A

A est auto-évident pour b , mais pas pour a , donc pas entre a et b .

Théorème 1

A est auto-évident entre a et b ssi A est l'union de membres des partitions des P_i ($i = a, b$)

Preuve: Supposons que A est auto-évident entre a et b : pour tout $w \in A$, $P_a(w) \subseteq A$ et $P_b(w) \subseteq A$, ie $\cup_{w \in A} P_a(w) \subseteq A$, et $\cup_{w \in A} P_b(w) \subseteq A$. Donc A est une union de membres de chaque partition. Réciproquement, supposons que A soit exactement l'union de membres de chaque partition. Alors pour chaque agent, pour tout $w \in A$, $P_i(w) \subseteq A$.

Autre formulation: A est auto-évident entre a et b ssi on peut trouver un grossissement commun des partitions de a et b tel que A coïncide avec l'une de celles de la nouvelle partition.

Retour à l'exemple: Montrer que $E = \{1, 2, 3, 4, 5\}$ and $F = \{6\}$ sont des événements auto-évidents.

2.3 La définition itérative de la connaissance commune

Posons: $K_i(A) = \{w \in W; P_i(w) \subseteq A\}$: ensemble des mondes dans lesquels l'agent sait que A

- (9) on dira que A est connaissance commune entre des agents a et b en w ssi w appartient à chacun des ensembles de la suite: $K_a(A)$, $K_b(A)$, $K_a K_b(A)$, $K_b K_a(A)$, etc.
- (10) Considérer le cas de $A = \{1, 2, 3, 4\}$

2.4 La définition par point fixe de la connaissance commune

- (11) A est un point fixe de K_i ssi $K_i(A) = A$
- (12) On dira qu'une proposition A est connaissance commune entre a et b en w ssi w appartient à une proposition qui est un point fixe des K_i et qui implique A .
- (13) Vérifier que $\{1, 2, 3, 4, 5\}$ and $\{6\}$ sont des points fixes.

Noter que si A est un point fixe, alors: $K_i(A) = K_j(A) = K_i K_j(A) = \dots$ Un point fixe garantit la définition itérative de la connaissance commune.

2.5 Coïncidence des définitions

Théorème 2

Soit W un ensemble fini de mondes. On suppose que P_a et P_b sont des fonctions d'information partitionnelles. E est connaissance commune au sens itératif entre 1 et 2 ssi E est connaissance commune au sens de l'existence d'un événement auto-évident A qui implique E ssi E est impliquée par un point fixe pour a et b .

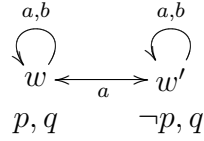
Preuve: La preuve fait essentiellement usage de l'équivalence des trois conditions suivantes, sous l'hypothèse que a et b ont des incertitudes partitionnelles:

- (14)
 - a. $K_i(A) = A$ ($i = a, b$)
 - b. A est auto-évident pour a et b
 - c. A est une union de membres des partitions P_a et P_b

Preuve: (a) implique clairement (b). Nous avons déjà vu que (b) implique (c). Enfin, supposons (c): $K_i(A) \subseteq A$ par véridicité; par ailleurs, pour tout $w \in A$, $P_i(w) \subseteq A$, donc $A \subseteq K_i(A)$.

3 Éléments de logique épistémique

Langage: $p \mid \neg\phi \mid \phi \wedge \psi \mid K_i\phi$



$K_i\phi$: i croit/sait que ϕ

$M, w \models p$ iff $w \in V(p)$

$M, w \models (\phi \wedge \psi)$ iff $M, w \models \phi$ and $M, w \models \psi$

$M, w \models \neg\phi$ iff $M, w \not\models \phi$

$M, w \models K_a\phi$ iff for every w' such that wR_aw' , $M, w' \models \phi$.

- Axiomes standards pour la connaissance/croyance

K	$K_a(p \rightarrow q) \rightarrow (K_ap \rightarrow K_aq)$	Closure
T	$K_ap \rightarrow p$	Knowledge, Veridicality
D	$K_a\neg p \rightarrow \neg K_ap$	Consistency
4	$K_ap \rightarrow K_aK_ap$	Positive Introspection
5	$\neg K_ap \rightarrow K_a\neg K_ap$	Negative Introspection

- Propriétés des cadres en sémantique de Kripke

K	—	All frames
T	$\forall x(xR_ax)$	Reflexive frames
D	$\forall x\exists y(xR_ay)$	Serial frames
4	$\forall xyz(xR_ay \wedge xR_az \rightarrow xR_az)$	Transitive frames
5	$\forall xyz(xR_ay \wedge xR_az \rightarrow yR_az)$	Euclidean frames

- Version Aumann des propriétés des cadres

T	$w \in P_a(w)$
D	$P_a(w) \neq \emptyset$
4	$w' \in P_a(w) \Rightarrow P_a(w') \subseteq P_a(w)$
5	$w' \in P_a(w) \Rightarrow P_a(w) \subseteq P_a(w')$

- Notion de validité: ϕ est une formule valide relativement à une classe de cadres (W, R) ssi pour tout V et pour tout w de W , $(W, R, V), w \models \phi$

(15) $Kp \rightarrow p$ valide pour les modèles réflexifs.

Supposez que (W, R) est un cadre réflexif. Soit une valuation V quelconque, telle que $M, w \models Kp$. Comme wRw , $M, w \models p$, donc $M, w \models Kp \rightarrow p$. Inversement, supposons que (W, R) n'est pas un cadre réflexif. Il existe un monde tel que: non wRw . Soit la valuation V telle que $V(p) = \{w'; wRw'\}$. Clairement, $M, w \models Kp$, mais $M, w \not\models p$.

- En abandonnant la réflexivité de la relation de possibilité épistémique, on peut représenter le fait qu'un agent ait des croyances fausses.

(16) Représenter les croyances d'un agent qui croit faussement qu'il pleut.

4 La connaissance commune

Connaissance partagée (shared knowledge), connaissance commune (common knowledge)

$M, w \models E_G \phi$ ssi pour tout w' tel que $w R_{E_G} w'$, $M, w' \models \phi$, avec R_{E_G} = union des relations d'accessibilité des membres de G .

$M, w \models C_G \phi$ ssi pour tout w' tel que $w R_{C_G} w'$, $M, w' \models \phi$, avec R_G = clôture (réflexive) transitive de l'union des R_i , ou de façon équivalente, ϕ est vrai en tout point accessible à partir de w par un chemin constitué de transitions étiquetées par les membres du groupe G

- Pour comprendre la définition des conditions de satisfaction de la connaissance commune, il faut se rapporter à la définition itérative:

$w \models K_a \phi$ ssi ϕ est vrai dans tous les mondes accessibles à 1 a -pas

$w \models K_a K_a \phi$ ssi ϕ est vrai dans tous les mondes accessibles à 2 a -pas

$w \models K_b K_b \phi$ ssi ϕ est vrai dans tous les mondes accessibles à 2 b -pas

$w \models K_b K_a \phi$ ssi ϕ est vrai dans tous les mondes accessibles à 1 a -pas de tous les mondes accessibles à 1 b -pas

$w \models K_a K_b \phi$ ssi ϕ est vrai dans tous les mondes accessibles à 1 b -pas de tous les mondes accessibles à 1 a -pas

Donc: $E_{a,b} \phi$ (chacun sait que chacun sait) est vrai dans w ssi ϕ est accessible dans tous les mondes accessibles par un chemin le long de l'union des a - et b -relations.

$$(17) \quad E_{a,b} \phi = K_a \phi \wedge K_b \phi$$

$$(18) \quad E_{a,b} E_{a,b} \phi = K_a (K_a \phi \wedge K_b \phi) \wedge K_b (K_a \phi \wedge K_b \phi)$$

$$(19) \quad C_{a,b} \phi = \bigwedge_{n \geq 1} (E_{a,b})^n \phi$$

- Axiomes pour la connaissance commune

(Point fixe) $C_G \phi \leftrightarrow E_G(\phi \wedge C_G \phi)$

(Récurrence) from $\phi \rightarrow E_G(\phi \wedge \psi)$, infer $\phi \rightarrow C_G \psi$.

Preuve:

(Point fixe) Supposons que $M, w \models C_G \phi$. Pour tout point G -accessible de w , ϕ est vrai en ce point. En particulier, pour tout point accessible à partir d'une transition, ϕ est vrai en ce point et ϕ est accessible à partir de tout point accessible par un chemin à partir de ce point. Donc, $M, w \models E_G(\phi \wedge C_G \phi)$. Inversement, supposons que $M, w \models E_G(\phi \wedge C_G \phi)$. Soit w' un point G -accessible à partir de w , et soit t le premier point sur ce chemin accessible

de w . Par hypothèse, $M, t \models \phi \wedge C_G \phi$, donc ϕ est vrai en t , mais ϕ est vrai en tout point G -accessible de t , donc est vrai en w' . CQFD.

- Quelques propriétés de la connaissance/croyance commune:

- Noter que $C_G \phi \rightarrow C_G C_G \phi$.

• Supposons que $\neg C_G \phi$. Alors il existe un modèle M et un monde w , et un G -chemin à partir de w jusqu'à w' tel que $M, w' \not\models \phi$. Supposons que chacun des agents est de type S5 (connaissance réflexive, symétrique et transitive). Soit n'importe quel point t accessible à partir de w par un G -chemin. On veut montrer qu'en ce point, $M, t \models \neg C_G \phi$. Par symétrie des relations, il existe un G -chemin de t à w' passant par w . Donc, pour tout G -chemin accessible à partir de w , il existe un G -chemin qui conduit jusqu'à un point non ϕ .

• Noter en revanche que si les possibilités épistémiques des agents ne sont pas symétriques et réflexives, l'absence de croyance commune n'est pas croyance commune.

5 Comment atteindre la connaissance commune?

5.1 L'exemple du jeu électronique (Rubinstein 1989)

g_1	A	B	g_2	A	B
A	10,10	0, -5	A	0,0	0, -5
B	-5,0	0,0	B	-5,0	10,10

$\overset{a,b}{\curvearrowright}$	$\overset{a,b}{\curvearrowright}$	$\overset{a,b}{\curvearrowright}$	$\overset{a,b}{\curvearrowright}$	$\overset{a,b}{\curvearrowright}$	
$(0,0)$	$\xrightarrow{b}$	$(1,0)$	$\xrightarrow{a}$	$(1,1)$	$\xrightarrow{b}$
g_1		g_2		g_2	

	$\overset{a,b}{\curvearrowright}$	$\overset{a,b}{\curvearrowright}$	$\overset{a,b}{\curvearrowright}$	$\overset{a,b}{\curvearrowright}$	
	$(1,1)$	$\xrightarrow{b}$	$(2,1)$	$\xrightarrow{a}$	$(2,2)$
	g_2		g_2		g_2

.....

Théorème: le jeu a un unique équilibre de Nash dans lequel les agents jouent (A,A)

5.2 Perspectives sur l'atteignabilité de CK

La notion itérative de connaissance commune a été largement critiquée, notamment pour être psychologiquement implausible:

- Sperber et Wilson 1986:

“Something can be manifest without being known.... Can something be manifest without being actually assumed? The answer must again be yes... The notion of a mutually manifest assumption is clearly weaker than that of mutual assumption (and a fortiori than that of mutual knowledge)”

- Clark and Marshall 1981 distinguent deux heuristiques:

1) Truncation heuristics:

”In principle, truncation heuristics seem capable of doing the job. They may even allow for felicitous definite reference. We suspect that they may be used on at least some occasions.... But are they plausible as the way people normally assess mutual knowledge in making definite reference? We believe not.”

“recursive reasoning event two levels deep is not easy for children under age twelve. The trouble is that children much younger – six to eight years of age (Maratsos 1976, Warden 1976) – appear to use definite reference felicitously, at least much of the time. And even younger children sometimes spontaneously repair definite references to take account of what their listeners know (E. Clark and Andersen 1979)”

2) Copresence heuristics: essentiellement la solution de Lewis, celle d’un événement qui indique à chacun des participants une information, et qui indique à chacun que chacun a l’information.

Le cas des descriptions définies:

(20) Anne à Bob: As-tu déjà vu le film qui passe au Gaumont ce soir?

Problème discuté par Clark et Marshall: que doit supposer Anne pour que son acte de référence par l’usage de la description définie soit réussi ?

“she must be certain that once she has made her reference he and she can establish certain shared knowledge about the identity of that referent”

6 Références

Manuels d’introduction à la logique épistémique:

R. Fagin, J. Halpern, Y. Moses, M. Vardi (1995), Reasoning about Knowledge, MIT Press.
C. Osborne et A. Rubinstein (1994). A Course in Game Theory (chap. 5).
H. van Ditmarsch, B. Kooi, W. van der Hoek (2007). Dynamic Epistemic Logic. Springer.

Articles

R. Aumann (1976). Agree to Disagree. Annals of Statistics.
H. Clark and C. Marshall (1981). Definite Reference and Mutual Knowledge. Repr. in H. Clark, Arenas of Language Use.

Approfondissement

D. Bonnay and P. Égré (2009). Inexact Knowledge with Introspection (partie 4) sur la connaissance quasi-commune

A lire pour la fois prochaine:

R. Stalnaker. chap. 5 in Context and Content: “On the Representation of Context”
R. Stalnaker. “Common Ground”