Logic and Artificial Intelligence

Lecture 6

Eric Pacuit

Currently Visiting the Center for Formal Epistemology, CMU

Center for Logic and Philosophy of Science
Tilburg University

ai.stanford.edu/∼epacuit
e.j.pacuit@uvt.nl

September 20, 2011

Example (1)

Suppose there are two friends Ann and Bob are on a bus separated by a crowd.

Example (1)

Suppose there are two friends Ann and Bob are on a bus separated by a crowd. Before the bus comes to the next stop a mutual friend from outside the bus yells "get off at the next stop to get a drink?".

Example (1)

Suppose there are two friends Ann and Bob are on a bus separated by a crowd. Before the bus comes to the next stop a mutual friend from outside the bus yells "get off at the next stop to get a drink?".

Say Ann is standing near the front door and Bob near the back door.

Example (1)

Suppose there are two friends Ann and Bob are on a bus separated by a crowd. Before the bus comes to the next stop a mutual friend from outside the bus yells "get off at the next stop to get a drink?".

Say Ann is standing near the front door and Bob near the back door. When the bus comes to a stop, will they get off?

Example (1)

Suppose there are two friends Ann and Bob are on a bus separated by a crowd. Before the bus comes to the next stop a mutual friend from outside the bus yells "get off at the next stop to get a drink?".

Say Ann is standing near the front door and Bob near the back door. When the bus comes to a stop, will they get off?

D. Lewis. *Convention*. 1969.

M. Chwe. *Rational Ritual*. 2001.

Example (2)

A and B are players in the same football team. A has the ball, but an opposing player is converging on him.

Example (2)

A and B are players in the same football team. A has the ball, but an opposing player is converging on him. He can pass the ball to B, who has a chance to shoot.

Example (2)

A and B are players in the same football team. A has the ball, but an opposing player is converging on him. He can pass the ball to B, who has a chance to shoot. There are two directions in which A can move the ball, *left* and *right*, and correspondingly, two directions in which B can run to intercept the pass.

Example (2)

A and B are players in the same football team. A has the ball, but an opposing player is converging on him. He can pass the ball to B, who has a chance to shoot. There are two directions in which A can move the ball, *left* and *right*, and correspondingly, two directions in which B can run to intercept the pass. If both choose *left* there is a 10% chance that a goal will be scored.

Example (2)

A and B are players in the same football team. A has the ball, but an opposing player is converging on him. He can pass the ball to B, who has a chance to shoot. There are two directions in which A can move the ball, *left* and *right*, and correspondingly, two directions in which B can run to intercept the pass. If both choose *left* there is a 10% chance that a goal will be scored. If they both choose *right*, there is a 11% change.

Example (2)

A and B are players in the same football team. A has the ball, but an opposing player is converging on him. He can pass the ball to B, who has a chance to shoot. There are two directions in which A can move the ball, *left* and *right*, and correspondingly, two directions in which B can run to intercept the pass. If both choose *left* there is a 10% chance that a goal will be scored. If they both choose *right*, there is a 11% change. Otherwise, the chance is zero.

Example (2)

A and B are players in the same football team. A has the ball, but an opposing player is converging on him. He can pass the ball to B, who has a chance to shoot. There are two directions in which A can move the ball, *left* and *right*, and correspondingly, two directions in which B can run to intercept the pass. If both choose *left* there is a 10% chance that a goal will be scored. If they both choose *right*, there is a 11% change. Otherwise, the chance is zero. There is no time for communication; the two players must act simultaneously.

Example (2)

A and *B* are players in the same football team. *A* has the ball, but an opposing player is converging on him. He can pass the ball to *B*, who has a chance to shoot. There are two directions in which *A* can move the ball, *left* and *right*, and correspondingly, two directions in which *B* can run to intercept the pass. If both choose *left* there is a 10% chance that a goal will be scored. If they both choose *right*, there is a 11% change. Otherwise, the chance is zero. There is no time for communication; the two players must act simultaneously.

What should they do?

R. Sugden. *The Logic of Team Reasoning*. Philosophical Explorations (6)3, pgs. 165 - 181 (2003).

Example (2)

		B	
		l	r
A	l		
	r		

Example (2)

			B	
		l		r
A	l	10,10		0,0
	r	0,0		11,11

A: What should I do?

Example (2)

			B l	B r
A	l		10,10	0,0
A	r		0,0	11,11

A: What should I do? r if the probability of B choosing r is $> \frac{10}{21}$ and l if the probability of B choosing l is $> \frac{11}{21}$
(symmetric reasoning for B)

Example (2)

			B	
			l	r
A		l	10,10	0,0
		r	0,0	11,11

A: What should I do? r if the probability of B choosing r is $> \frac{10}{21}$ and l if the probability of B choosing l is $> \frac{11}{21}$
(symmetric reasoning for B)

Example (2)

		B: l	B: r
A	l	10,10	0,0
	r	0,0	11,11

A: What should *we* do?

Example (2)

		B	
		l	r
A	l	10,10	0,0
	r	0,0	11,11

A: What should *we* do? **Team Reasoning**: why should this "mode of reasoning" be endorsed?

Example (2)

		B	
		l	r
A	l	10,10	0,0
	r	0,0	11,11

A: What should *we* do? **Team Reasoning**: why should this "mode of reasoning" be endorsed?

"*Common Knowledge*" is informally described as what any fool would know, given a certain situation: It encompasses what is relevant, agreed upon, established by precedent, assumed, being attended to, salient, or in the conversational record.

"*Common Knowledge*" is informally described as what any fool would know, given a certain situation: It encompasses what is relevant, agreed upon, established by precedent, assumed, being attended to, salient, or in the conversational record.

It is not Common Knowledge who "defined" Common Knowledge!

The first formal definition of common knowledge?

M. Friedell. *On the Structure of Shared Awareness*. Behavioral Science (1969).

R. Aumann. *Agreeing to Disagree*. Annals of Statistics (1976).

The first formal definition of common knowledge?

M. Friedell. *On the Structure of Shared Awareness*. Behavioral Science (1969).

R. Aumann. *Agreeing to Disagree*. Annals of Statistics (1976).

The first rigorous analysis of common knowledge

D. Lewis. *Convention, A Philosophical Study*. 1969.

The first formal definition of common knowledge?

M. Friedell. *On the Structure of Shared Awareness*. Behavioral Science (1969).

R. Aumann. *Agreeing to Disagree*. Annals of Statistics (1976).

The first rigorous analysis of common knowledge

D. Lewis. *Convention, A Philosophical Study*. 1969.

Fixed-point definition: $\gamma :=$ *i* and *j* know that (φ and γ)

G. Harman. *Review of* Linguistic Behavior. Language (1977).

J. Barwise. *Three views of Common Knowledge*. TARK (1987).

The first formal definition of common knowledge?

M. Friedell. *On the Structure of Shared Awareness*. Behavioral Science (1969).

R. Aumann. *Agreeing to Disagree*. Annals of Statistics (1976).

The first rigorous analysis of common knowledge

D. Lewis. *Convention, A Philosophical Study*. 1969.

Fixed-point definition: $\gamma :=$ i and j know that (φ and γ)

G. Harman. *Review of* Linguistic Behavior. Language (1977).

J. Barwise. *Three views of Common Knowledge*. TARK (1987).

Shared situation: There is a *shared situation s* such that (1) s entails φ, (2) s entails everyone knows φ, plus other conditions

H. Clark and C. Marshall. *Definite Reference and Mutual Knowledge*. 1981.

M. Gilbert. *On Social Facts*. Princeton University Press (1989).

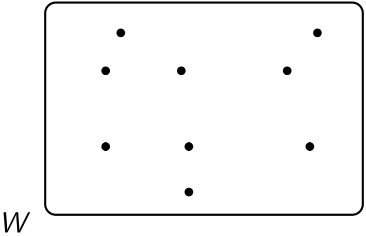
P. Vanderschraaf and G. Sillari. *"Common Knowledge", The Stanford Encyclopedia of Philosophy (2009)*.
http://plato.stanford.edu/entries/common-knowledge/.

The "Standard" Account

R. Aumann. *Agreeing to Disagree*. Annals of Statistics (1976).

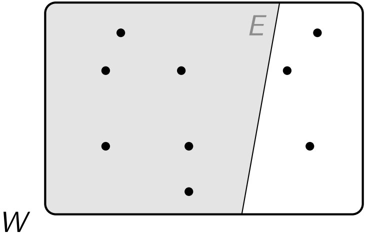
R. Fagin, J. Halpern, Y. Moses and M. Vardi. *Reasoning about Knowledge*. MIT Press, 1995.

The "Standard" Account



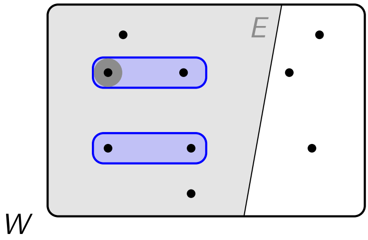
W is a set of **states** or **worlds**.

The "Standard" Account



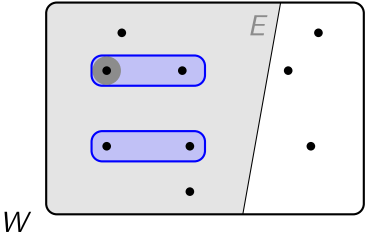
An **event/proposition** is any (definable) subset $E \subseteq W$

The "Standard" Account



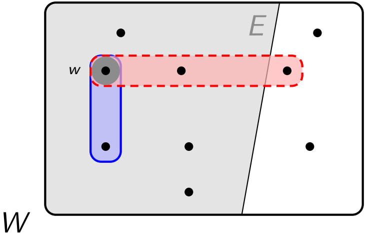
At each state, agents are assigned a set of states they *consider possible* (according to their information).
The information may be (in)correct, partitional, ...

The "Standard" Account



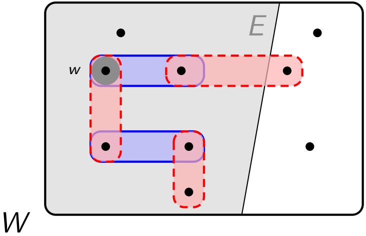
Knowledge Function: $K_i : \wp(W) \to \wp(W)$ where $K_i(E) = \{w \mid R_i(w) \subseteq E\}$

The “Standard” Account



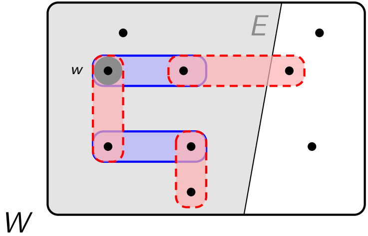
$w \in K_A(E)$ and $w \notin K_B(E)$

The "Standard" Account



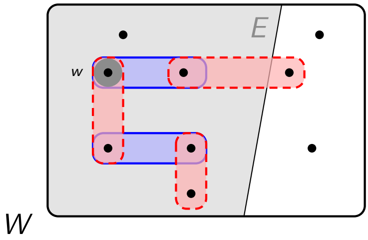
The model also describes the agents' **higher-order knowledge/beliefs**

The "Standard" Account



Everyone Knows: $K(E) = \bigcap_{i \in \mathcal{A}} K_i(E)$, $K^0(E) = E$, $K^m(E) = K(K^{m-1}(E))$

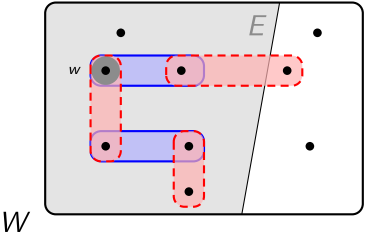
The “Standard” Account



Common Knowledge: $C : \wp(W) \to \wp(W)$ with

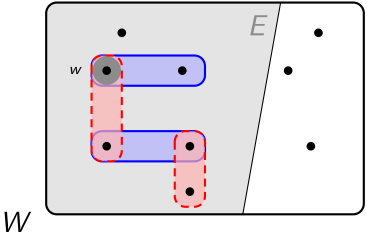
$$C(E) = \bigcap_{m \geq 0} K^m(E)$$

The “Standard” Account



$w \in K(E) \qquad w \notin C(E)$

The "Standard" Account



$w \in C(E)$

Fact. For all $i \in \mathcal{A}$ and $E \subseteq W$, $K_i C(E) = C(E)$.

Fact. For all $i \in \mathcal{A}$ and $E \subseteq W$, $K_i C(E) = C(E)$.

Suppose you are told "Ann and Bob are going together,"' and respond "sure, that's common knowledge." What you mean is not only that everyone knows this, but also that the announcement is pointless, occasions no surprise, reveals nothing new; in effect, that the situation after the announcement does not differ from that before. ...the event "Ann and Bob are going together" — call it E — is common knowledge if and only if some event — call it F — happened that entails E and also entails all players' knowing F (like all players met Ann and Bob at an intimate party). *(Aumann, pg. 271, footnote 8)*

Fact. For all $i \in \mathcal{A}$ and $E \subseteq W$, $K_i C(E) = C(E)$.

An event F is **self-evident** if $K_i(F) = F$ for all $i \in \mathcal{A}$.

Fact. An event E is commonly known iff some self-evident event that entails E obtains.

Fact. For all $i \in \mathcal{A}$ and $E \subseteq W$, $K_i C(E) = C(E)$.

An event F is **self-evident** if $K_i(F) = F$ for all $i \in \mathcal{A}$.

Fact. An event E is commonly known iff some self-evident event that entails E obtains.

Fact. $w \in C(E)$ if every finite path starting at w ends in a state in E

The following axiomatize common knowledge:

- $C(\varphi \to \psi) \to (C\varphi \to C\psi)$
- $C\varphi \to (\varphi \wedge EC\varphi)$ (Fixed-Point)
- $C(\varphi \to E\varphi) \to (\varphi \to C\varphi)$ (Induction)

An Example

Two players Ann and Bob are told that the following will happen. Some positive integer n will be chosen and *one* of n, $n+1$ will be written on Ann's forehead, the other on Bob's. Each will be able to see the other's forehead, but not his/her own.

An Example

Two players Ann and Bob are told that the following will happen. Some positive integer n will be chosen and *one* of n, $n + 1$ will be written on Ann's forehead, the other on Bob's. Each will be able to see the other's forehead, but not his/her own.

Suppose the number are (2,3).

An Example

Two players Ann and Bob are told that the following will happen. Some positive integer n will be chosen and *one* of n, $n+1$ will be written on Ann's forehead, the other on Bob's. Each will be able to see the other's forehead, but not his/her own.

Suppose the number are (2,3).

Do the agents know there numbers are less than 1000?

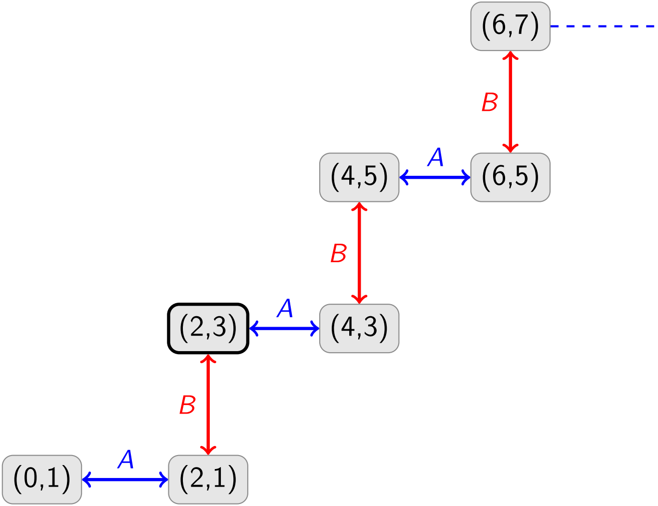
An Example

Two players Ann and Bob are told that the following will happen. Some positive integer n will be chosen and *one* of n, $n + 1$ will be written on Ann's forehead, the other on Bob's. Each will be able to see the other's forehead, but not his/her own.

Suppose the number are (2,3).

Do the agents know there numbers are less than 1000?

Is it common knowledge that their numbers are less than 1000?


(6,7)
B
(4,5)
A
(6,5)
B
(2,3)
A
(4,3)
B
(0,1)
A
(2,1)

The Fixed-Point Definition

The Fixed-Point Definition

$$f_E(X) = K(E \cap X) = \bigcap_{i \in \mathcal{A}} K_i(E \cap X)$$

The Fixed-Point Definition

$f_E(X) = K(E \cap X) = \bigcap_{i \in \mathcal{A}} K_i(E \cap X)$

- $C(E)$ is a fixed point of f_E: $f_E(C(E))$

The Fixed-Point Definition

$f_E(X) = K(E \cap X) = \bigcap_{i \in \mathcal{A}} K_i(E \cap X)$

- $C(E)$ is a fixed point of f_E: $f_E(C(E)) = K(E \cap C(E))$

The Fixed-Point Definition

$f_E(X) = K(E \cap X) = \bigcap_{i \in \mathcal{A}} K_i(E \cap X)$

- $C(E)$ is a fixed point of f_E: $f_E(C(E)) = K(E \cap C(E)) = K(C(E)) = \bigcap_{i \in \mathcal{A}} K_i(C(E))$

The Fixed-Point Definition

$$f_E(X) = K(E \cap X) = \bigcap_{i \in \mathcal{A}} K_i(E \cap X)$$

- $C(E)$ is a fixed point of f_E: $f_E(C(E)) = K(E \cap C(E)) = K(C(E)) = \bigcap_{i \in \mathcal{A}} K_i(C(E)) = \bigcap_{i \in \mathcal{A}} C(E) = C(E)$

The Fixed-Point Definition

$f_E(X) = K(E \cap X) = \bigcap_{i \in \mathcal{A}} K_i(E \cap X)$

- $C(E)$ is a fixed point of f_E: $f_E(C(E)) = K(E \cap C(E)) = K(C(E)) = \bigcap_{i \in \mathcal{A}} K_i(C(E)) = \bigcap_{i \in \mathcal{A}} C(E) = C(E)$
- The are other fixed points of f_E: $f_E(\bot) = \bot$

The Fixed-Point Definition

$f_E(X) = K(E \cap X) = \bigcap_{i \in \mathcal{A}} K_i(E \cap X)$

- $C(E)$ is a fixed point of f_E: $f_E(C(E)) = K(E \cap C(E)) = K(C(E)) = \bigcap_{i \in \mathcal{A}} K_i(C(E)) = \bigcap_{i \in \mathcal{A}} C(E) = C(E)$
- The are other fixed points of f_E: $f_E(\bot) = \bot$
- f_E is monotonic:

The Fixed-Point Definition

$f_E(X) = K(E \cap X) = \bigcap_{i \in \mathcal{A}} K_i(E \cap X)$

- $C(E)$ is a fixed point of f_E: $f_E(C(E)) = K(E \cap C(E)) = K(C(E)) = \bigcap_{i \in \mathcal{A}} K_i(C(E)) = \bigcap_{i \in \mathcal{A}} C(E) = C(E)$
- The are other fixed points of f_E: $f_E(\bot) = \bot$
- f_E is monotonic: $A \subseteq B$ implies $E \cap A \subseteq E \cap B$.

The Fixed-Point Definition

$f_E(X) = K(E \cap X) = \bigcap_{i \in \mathcal{A}} K_i(E \cap X)$

- $C(E)$ is a fixed point of f_E: $f_E(C(E)) = K(E \cap C(E)) = K(C(E)) = \bigcap_{i \in \mathcal{A}} K_i(C(E)) = \bigcap_{i \in \mathcal{A}} C(E) = C(E)$
- The are other fixed points of f_E: $f_E(\bot) = \bot$
- f_E is monotonic: $A \subseteq B$ implies $E \cap A \subseteq E \cap B$. Then $f_E(E \cap A) = K(E \cap A) \subseteq K(E \cap B) = f_E(E \cap B)$

The Fixed-Point Definition

$$f_E(X) = K(E \cap X) = \bigcap_{i \in \mathcal{A}} K_i(E \cap X)$$

- $C(E)$ is a fixed point of f_E: $f_E(C(E)) = K(E \cap C(E)) = K(C(E)) = \bigcap_{i \in \mathcal{A}} K_i(C(E)) = \bigcap_{i \in \mathcal{A}} C(E) = C(E)$
- The are other fixed points of f_E: $f_E(\bot) = \bot$
- f_E is monotonic: $A \subseteq B$ implies $E \cap A \subseteq E \cap B$. Then $f_E(E \cap A) = K(E \cap A) \subseteq K(E \cap B) = f_E(E \cap B)$
- (Tarski) Every monotone operator has a greatest (and least) fixed point

The Fixed-Point Definition

$f_E(X) = K(E \cap X) = \bigcap_{i \in \mathcal{A}} K_i(E \cap X)$

- $C(E)$ is a fixed point of f_E: $f_E(C(E)) = K(E \cap C(E)) = K(C(E)) = \bigcap_{i \in \mathcal{A}} K_i(C(E)) = \bigcap_{i \in \mathcal{A}} C(E) = C(E)$
- The are other fixed points of f_E: $f_E(\bot) = \bot$
- f_E is monotonic: $A \subseteq B$ implies $E \cap A \subseteq E \cap B$. Then $f_E(E \cap A) = K(E \cap A) \subseteq K(E \cap B) = f_E(E \cap B)$
- (Tarski) Every monotone operator has a greatest (and least) fixed point
- Let $K^*(E)$ be the greatest fixed point of f_E.

The Fixed-Point Definition

$f_E(X) = K(E \cap X) = \bigcap_{i \in \mathcal{A}} K_i(E \cap X)$

- $C(E)$ is a fixed point of f_E: $f_E(C(E)) = K(E \cap C(E)) = K(C(E)) = \bigcap_{i \in \mathcal{A}} K_i(C(E)) = \bigcap_{i \in \mathcal{A}} C(E) = C(E)$
- The are other fixed points of f_E: $f_E(\bot) = \bot$
- f_E is monotonic: $A \subseteq B$ implies $E \cap A \subseteq E \cap B$. Then $f_E(E \cap A) = K(E \cap A) \subseteq K(E \cap B) = f_E(E \cap B)$
- (Tarski) Every monotone operator has a greatest (and least) fixed point
- Let $K^*(E)$ be the greatest fixed point of f_E.
- **Fact**. $K^*(E) = C(E)$.

The Fixed-Point Definition

Separating the fixed-point/iteration definition of common knowledge/belief:

J. Barwise. *Three views of Common Knowledge*. TARK (1987).

J. van Benthem and D. Saraenac. *The Geometry of Knowledge*. Aspects of Universal Logic (2004).

A. Heifetz. *Iterative and Fixed Point Common Belief*. Journal of Philosophical Logic (1999).

Some Issues

Some Issues

- What *does* a group know/believe/accept? vs. what *can* a group (come to) know/believe/accept?

Some Issues

- What *does* a group know/believe/accept? vs. what *can* a group (come to) know/believe/accept?

C. List. *Group knowledge and group rationality: a judgment aggregation perspective*. Episteme (2008).

Some Issues

- What *does* a group know/believe/accept? vs. what *can* a group (come to) know/believe/accept?

C. List. *Group knowledge and group rationality: a judgment aggregation perspective*. Episteme (2008).

- Other "group informational attitudes": distributed knowledge, common belief, ...

Some Issues

- What *does* a group know/believe/accept? vs. what *can* a group (come to) know/believe/accept?

C. List. *Group knowledge and group rationality: a judgment aggregation perspective*. Episteme (2008).

- Other "group informational attitudes": distributed knowledge, common belief, ...

- Where does common knowledge come from?

Some Issues

✓ What *does* a group know/believe/accept? vs. what *can* a group (come to) know/believe/accept?

C. List. *Group knowledge and group rationality: a judgment aggregation perspective*. Episteme (2008).

- Other "group informational attitudes": distributed knowledge, common belief, . . .
- Where does common knowledge come from?

Distributed Knowledge

$$D_G(E) = \{w \mid \left(\bigcap_{i \in G} R_i(w)\right) \subseteq E\}$$

Distributed Knowledge

$$D_G(E) = \{w \mid \left(\bigcap_{i \in G} R_i(w)\right) \subseteq E\}$$

- $K_A(p) \wedge K_B(p \to q) \to D_{A,B}(q)$
- $D_G(\varphi) \to \bigwedge_{i \in G} K_i \varphi$

Distributed Knowledge

$$D_G(E) = \{w \mid \left(\bigcap_{i \in G} R_i(w)\right) \subseteq E\}$$

- $K_A(p) \wedge K_B(p \to q) \to D_{A,B}(q)$
- $D_G(\varphi) \to \bigwedge_{i \in G} K_i \varphi$

F. Roelofsen. *Distributed Knowledge*. Journal of Applied Nonclassical Logic (2006).

Distributed Knowledge

$$D_G(E) = \{w \mid \left(\bigcap_{i \in G} R_i(w)\right) \subseteq E\}$$

- $K_A(p) \wedge K_B(p \to q) \to D_{A,B}(q)$
- $D_G(\varphi) \to \bigwedge_{i \in G} K_i \varphi$

F. Roelofsen. *Distributed Knowledge*. Journal of Applied Nonclassical Logic (2006).

$$w \in K_G(E) \text{ iff } R_G(w) \subseteq E \quad (\text{without necessarily } R_G(w) = \bigcap_{i \in G} R_i(w))$$

A. Baltag and S. Smets. *Correlated Knowledge: an Epistemic-Logic view on Quantum Entanglement*. Int. Journal of Theoretical Physics (2010).

Common p-belief

The typical example of an event that creates common knowledge is a **public announcement**.

Common p-belief

The typical example of an event that creates common knowledge is a **public announcement**.

Shouldn't one always allow for some small probability that a participant was absentminded, not listening, sending a text, checking facebook, proving a theorem, asleep, ...

Common p-belief

The typical example of an event that creates common knowledge is a **public announcement**.

Shouldn't one always allow for some small probability that a participant was absentminded, not listening, sending a text, checking facebook, proving a theorem, asleep, ...

"We show that the weaker concept of "common belief" can function successfully as a substitute for common knowledge in the theory of equilibrium of Bayesian games."

D. Monderer and D. Samet. *Approximating Common Knowledge with Common Beliefs*. Games and Economic Behavior (1989).

Common p-belief: definition

Let $(W, \{R_i\}_{i \in \mathcal{A}}, p, V)$ be a Bayesian model.

Common p-belief: definition

Let $(W, \{R_i\}_{i \in \mathcal{A}}, p, V)$ be a Bayesian model.

$$B_i^p(E) = \{w \mid p(E \mid R_i(w)) \geq p\}$$

Common p-belief: definition

Let $(W, \{R_i\}_{i \in \mathcal{A}}, p, V)$ be a Bayesian model.

$B_i^p(E) = \{w \mid p(E \mid R_i(w)) \geq p\}$

An event E is **evident p-belief** if for each $i \in \mathcal{A}$, $E \subseteq B_i^p(E)$

Common p-belief: definition

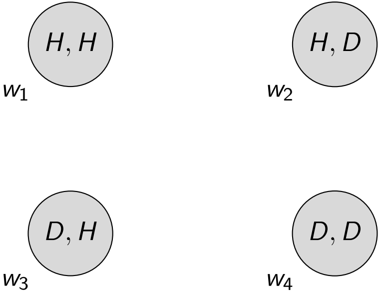
Let $(W, \{R_i\}_{i\in\mathcal{A}}, p, V)$ be a Bayesian model.

$B_i^p(E) = \{w \mid p(E \mid R_i(w)) \geq p\}$

An event E is **evident p-belief** if for each $i \in \mathcal{A}$, $E \subseteq B_i^p(E)$

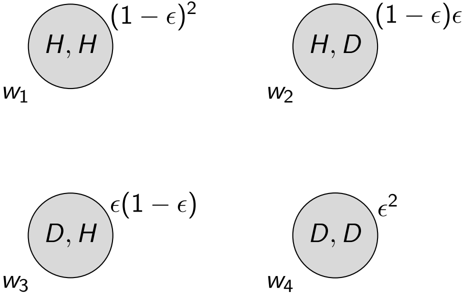
An event F is **common p-belief** at w if there exists and evident p-belief event E such that $w \in E$ and for all $i \in \mathcal{A}$, $E \subseteq B_i^p(F)$

Common p-belief: example



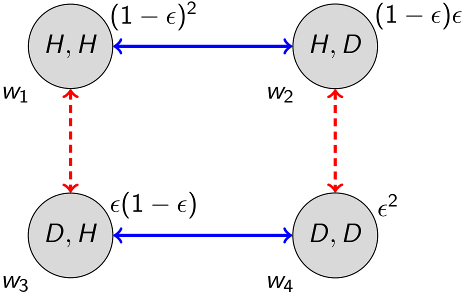
Two agents either hear (H) or don't hear (D) the announcement.

Common p-belief: example



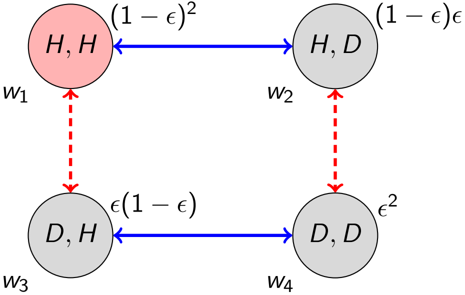
The probability that an agent hears is $1 - \epsilon$.

Common p-belief: example



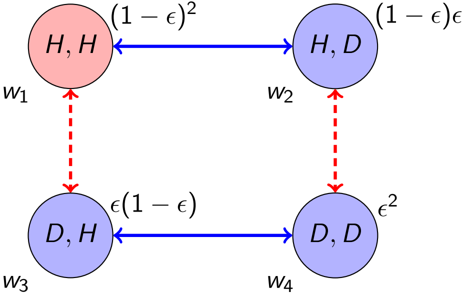
The agents *know* their “type”.

Common p-belief: example



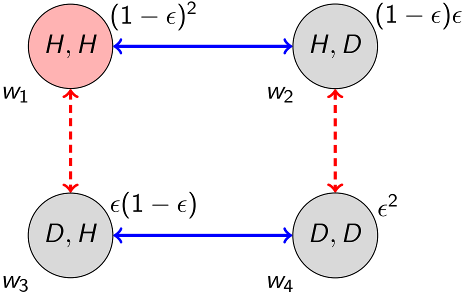
The event "everyone hears" ($E = \{w_1\}$)

Common p-belief: example



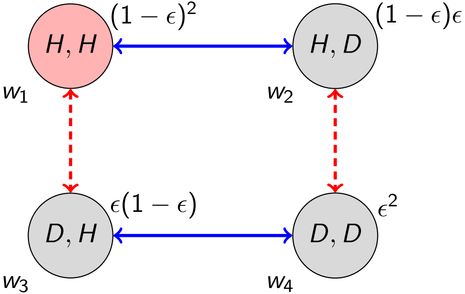
The event "everyone hears" ($E = \{w_1\}$) is **not** common knowledge

Common p-belief: example



The event "everyone hears" ($E = \{w_1\}$) is **not** common knowledge, but it is common $(1-\epsilon)$-belief

Common p-belief: example



The event "everyone hears" ($E = \{w_1\}$) is **not** common knowledge, but it is common $(1-\epsilon)$-belief:
$B_i^{(1-\epsilon)}(E) = \{w \mid p(E \mid \Pi_i(w)) \geq 1-\epsilon\} = \{w_1\} = E$, for $i = 1, 2$

Back

Some Issues

✓ What *does* a group know/believe/accept? vs. what *can* a group (come to) know/believe/accept?

C. List. *Group knowledge and group rationality: a judgment aggregation perspective*. Episteme (2008).

✓ Other "group informational attitudes": distributed knowledge, common belief, . . .

- Where does common knowledge come from?

Digression: Lewis' common knowledge

R. Cubitt and R. Sugden. *Common Knowledge, Salience and Convention: A Reconstruction of David Lewis' Game Theory*. Economics and Philosophy, 19, pgs. 175-210 , 2003..

Reason to Believe

$B_i\varphi$: "i believes φ"

Reason to Believe

$B_i\varphi$: "i believes φ" vs. $R_i(\varphi)$: "i has a *reason to believe* φ"

Reason to Believe

$B_i\varphi$: "i believes φ" vs. $R_i(\varphi)$: "i has a *reason to believe* φ"

- "Although it is an essential part of Lewis' theory that human beings are *to some degree* rational, he does not want to make the strong rationality assumptions of conventional decision theory or game theory." (CS, pg. 184).

Reason to Believe

$B_i\varphi$: "i believes φ" vs. $R_i(\varphi)$: "i has a *reason to believe* φ"

- "Although it is an essential part of Lewis' theory that human beings are *to some degree* rational, he does not want to make the strong rationality assumptions of conventional decision theory or game theory." (CS, pg. 184).
- Anyone who accept the rules of arithmetic has a reason to believe $618 \times 377 = 232,986$, but most of us do not hold have firm beliefs about this.

Reason to Believe

$B_i\varphi$: "i believes φ" vs. $R_i(\varphi)$: "i has a *reason to believe* φ"

- "Although it is an essential part of Lewis' theory that human beings are *to some degree* rational, he does not want to make the strong rationality assumptions of conventional decision theory or game theory." (CS, pg. 184).
- Anyone who accept the rules of arithmetic has a reason to believe $618 \times 377 = 232,986$, but most of us do not hold have firm beliefs about this.
- Definition: $R_i(\varphi)$ means φ is true within some logic of reasoning that is *endorsed* by (that is, accepted as a normative standard by) person i...φ must be either regarded as *self-evident* or derivable by rules of inference (deductive or inductive)

A indicates to i that φ

A is a "state of affairs"

$A\ ind_i\ \varphi$: i's reason to believe that A holds *provides* i's reason for believing that φ is true.

(*A*1)For all i, for all A, for all φ: $[R_i(A\ holds) \wedge (A\ ind_i\ \varphi)] \Rightarrow R_i(\varphi)$

Some Properties

Some Properties

- $[(A \textit{ holds}) \textit{ entails } (A' \textit{ holds})] \Rightarrow A \; ind_i(A' \textit{ holds})$

Some Properties

- $[(A \ holds) \ entails \ (A' \ holds)] \Rightarrow A \ ind_i(A' \ holds)$
- $[(A \ ind_i \ \varphi) \wedge \ (A \ ind_i \ \ \psi)] \Rightarrow A \ ind_i(\varphi \wedge \psi)$

Some Properties

- $[(A\ \mathit{holds})\ \mathit{entails}\ (A'\ \mathit{holds})] \Rightarrow A\ \mathit{ind}_i(A'\ \mathit{holds})$
- $[(A\ \mathit{ind}_i\ \varphi) \wedge (A\ \mathit{ind}_i\ \psi)] \Rightarrow A\ \mathit{ind}_i(\varphi \wedge \psi)$
- $[(A\ \mathit{ind}_i\ [A'\ \mathit{holds}]) \wedge (A'\ \mathit{ind}_i\ \varphi)] \Rightarrow A\ \mathit{ind}_i\varphi$

Some Properties

- $[(A\ holds)\ entails\ (A'\ holds)] \Rightarrow A\ ind_i(A'\ holds)$
- $[(A\ ind_i\ \varphi) \wedge\ (A\ ind_i\ \ \psi)] \Rightarrow A\ ind_i(\varphi \wedge \psi)$
- $[(A\ ind_i\ [A'\ holds]) \wedge (A'\ ind_i\ \ \varphi)] \Rightarrow A\ ind_i\varphi$
- $[(A\ ind_i\varphi) \wedge (\varphi\ entails\ \psi)] \Rightarrow A\ ind_i\psi$

Some Properties

- $[(A\ holds)\ entails\ (A'\ holds)] \Rightarrow A\ ind_i(A'\ holds)$
- $[(A\ ind_i\ \varphi) \wedge\ (A\ ind_i\ \ \psi)] \Rightarrow A\ ind_i(\varphi \wedge \psi)$
- $[(A\ ind_i\ [A'\ holds]) \wedge (A'\ ind_i\ \ \varphi)] \Rightarrow A\ ind_i \varphi$
- $[(A\ ind_i \varphi) \wedge (\varphi\ entails\ \psi)] \Rightarrow A\ ind_i \psi$
- $[(A\ ind_i\ R_j[A'\ holds]) \wedge R_i(A'\ ind_j \varphi)] \Rightarrow A\ ind_i R_j(\varphi)$

Reflexive Common Indicator

Reflexive Common Indicator

- $A\ holds \Rightarrow R_i(A\ holds)$

Reflexive Common Indicator

- $A \ holds \Rightarrow R_i(A \ holds)$
- $A \ ind_i \ R_j(A \ holds)$

Reflexive Common Indicator

- $A \text{ holds} \Rightarrow R_i(A \text{ holds})$
- $A \ ind_i \ R_j(A \text{ holds})$
- $A \ ind_i \ \varphi$

Reflexive Common Indicator

- $A\ holds \Rightarrow R_i(A\ holds)$
- $A\ ind_i\ R_j(A\ holds)$
- $A\ ind_i\ \varphi$
- $(A\ ind_i\ \psi) \Rightarrow R_i[A\ ind_j\ \psi]$

Let $R^G(\varphi)$: $R_i\varphi, R_j\varphi, \ldots$, $R_i(R_j\varphi)$, $R_j(R_i(\varphi))$, $\ldots$
iterated reason to believe φ.

Let $R^G(\varphi)$: $R_i\varphi, R_j\varphi, \ldots$, $R_i(R_j\varphi)$, $R_j(R_i(\varphi))$, $\ldots$
iterated reason to believe φ.

Theorem. (Lewis) For all states of affairs A, for all propositions φ, and for all groups G: if A holds, and if A is a reflexive common indicator in G that φ, then $R^G(\varphi)$ is true.

Lewis and Aumann

Lewis common knowledge that φ *implies* the iterated definition of common knowledge ("Aumann common knowledge")

Lewis and Aumann

Lewis common knowledge that φ *implies* the iterated definition of common knowledge ("Aumann common knowledge"), but the converse is not generally true....

Lewis and Aumann

Lewis common knowledge that φ *implies* the iterated definition of common knowledge ("Aumann common knowledge"), but the converse is not generally true....

Example. Suppose there is an agent $i \notin G$ that is *authoritative* for each member of G.

Lewis and Aumann

Lewis common knowledge that φ *implies* the iterated definition of common knowledge ("Aumann common knowledge"), but the converse is not generally true....

Example. Suppose there is an agent $i \notin G$ that is *authoritative* for each member of G. So, for $j \in G$, "i states to j that φ is true" *indicates to* j *that* φ.

Lewis and Aumann

Lewis common knowledge that φ *implies* the iterated definition of common knowledge ("Aumann common knowledge"), but the converse is not generally true....

Example. Suppose there is an agent $i \notin G$ that is *authoritative* for each member of G. So, for $j \in G$, "i states to j that φ is true" *indicates to j that φ*. Suppose that *separately and privately* to each member of G, i states that φ and $R^G(\varphi)$ are true.

Lewis and Aumann

Lewis common knowledge that φ *implies* the iterated definition of common knowledge ("Aumann common knowledge"), but the converse is not generally true....

Example. Suppose there is an agent $i \notin G$ that is *authoritative* for each member of G. So, for $j \in G$, "i states to j that φ is true" *indicates to j that φ*. Suppose that *separately and privately* to each member of G, i states that φ and $R^G(\varphi)$ are true. Then, we have $R^i\varphi$ and $R_i(R^G(\varphi))$ for each $i \in G$.

Lewis and Aumann

Lewis common knowledge that φ *implies* the iterated definition of common knowledge ("Aumann common knowledge"), but the converse is not generally true....

Example. Suppose there is an agent $i \notin G$ that is *authoritative* for each member of G. So, for $j \in G$, "i states to j that φ is true" *indicates to j that φ*. Suppose that *separately and privately* to each member of G, i states that φ and $R^G(\varphi)$ are true. Then, we have $R^i\varphi$ and $R_i(R^G(\varphi))$ for each $i \in G$. *But there is no common indicator that φ is true.*

Lewis and Aumann

Lewis common knowledge that φ *implies* the iterated definition of common knowledge ("Aumann common knowledge"), but the converse is not generally true....

Example. Suppose there is an agent $i \notin G$ that is *authoritative* for each member of G. So, for $j \in G$, "i states to j that φ is true" *indicates to j that φ*. Suppose that *separately and privately* to each member of G, i states that φ and $R^G(\varphi)$ are true. Then, we have $R^i\varphi$ and $R_i(R^G(\varphi))$ for each $i \in G$. *But there is no common indicator that φ is true.* The agents $j \in G$ may have no reason to believe that everyone heard the statement from i or that all agents in G treat i as authoritative.