Epistemic Game Theory

Lecture 3

Eric Pacuit

University of Maryland, College Park
pacuit.org
epacuit@umd.edu

February 10, 2014

The “Axiom” of Game Theory

Common Knowledge of Rationality

The "Axiom" of Game Theory

Common Knowledge of Rationality

"Choose *optimally* given the players' opinions about what the opponents might do (Bayesian Decision Theory)"

The "Axiom" of Game Theory

Common Knowledge of Rationality

believes, strongly/robustly
believes, knows...

"Choose *optimally* given the players' opinions about what the opponents might do (Bayesian Decision Theory)"

The "Axiom" of Game Theory

Common Knowledge of Rationality

believes, strongly/robustly believes, knows...

"it is completely *transparent* to the players that everyone..."

"Choose *optimally* given the players' opinions about what the opponents might do"

"*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. Nozick. *The Normative Theory of Individual Choice*. PhD dissertation, 1963.

The first formal definition of common knowledge?

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

R. Nozick. *The Normative Theory of Individual Choice*. PhD dissertation, 1963.

The first rigorous analysis of common knowledge (iterated definition)

D. Lewis. *Convention, A Philosophical Study*. 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. Nozick. *The Normative Theory of Individual Choice*. PhD dissertation, 1963.

The first rigorous analysis of common knowledge

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

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

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. Nozick. *The Normative Theory of Individual Choice*. PhD dissertation, 1963.

The first rigorous analysis of common knowledge

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

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

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).

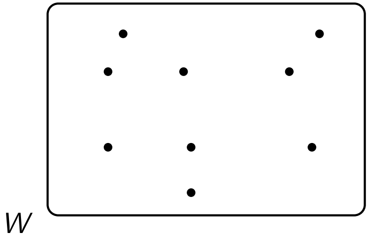
Shared situation: There is a *shared situation* s such that (1) s entails φ, (2) s entails everyone knows φ, plus other conditions

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

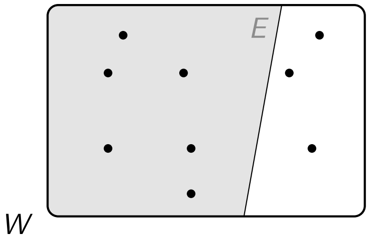
M. Gilbert. *On Social Facts*. Princeton University Press (1989).

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

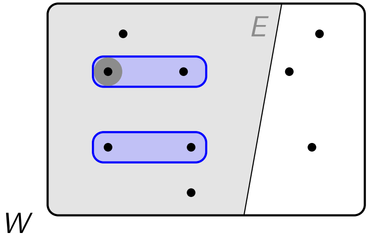
The “standard” definition of common knowledge.



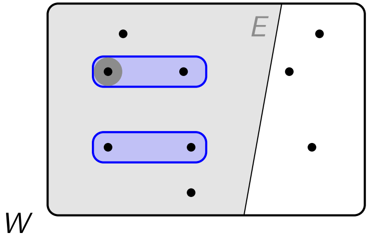
W is a set of **states** or **worlds**.



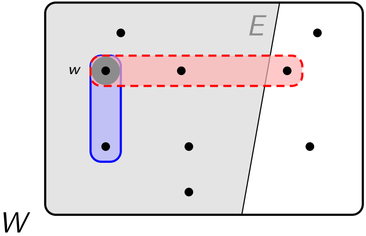
An **event**/**proposition** is any (definable) subset $E \subseteq W$



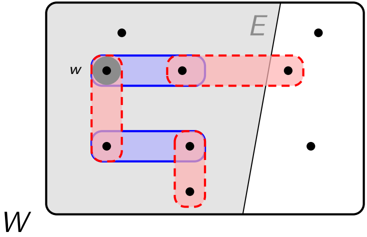
The agents receive signals in each state. States are considered equivalent for the agent if they receive the same signal in both states.



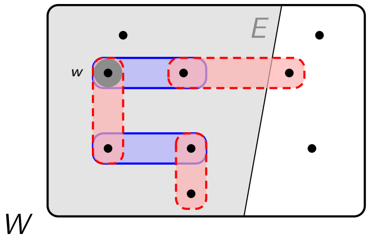
Knowledge Function: $K_i : \wp(W) \to \wp(W)$ where $K_i(E) = \{w \mid R_i(w) \subseteq E\}$



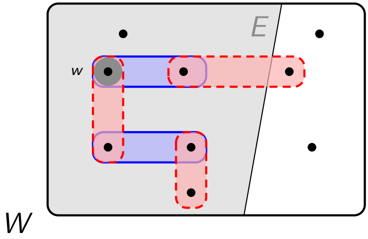
$w \in K_A(E)$ and $w \notin K_B(E)$



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

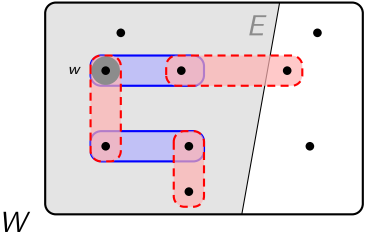


Everyone Knows: $K(E) = \bigcap_{i\in\mathcal{A}} K_i(E)$, $K^0(E) = E$, $K^m(E) = K(K^{m-1}(E))$

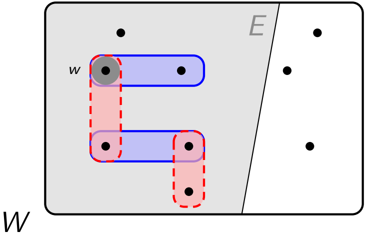


Common Knowledge: $C : \wp(W) \rightarrow \wp(W)$ with

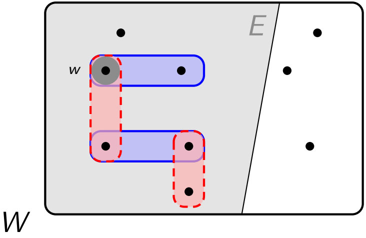
$$C(E) = \bigcap_{m \geq 0} K^m(E)$$



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



$w \in C(E)$



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

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?

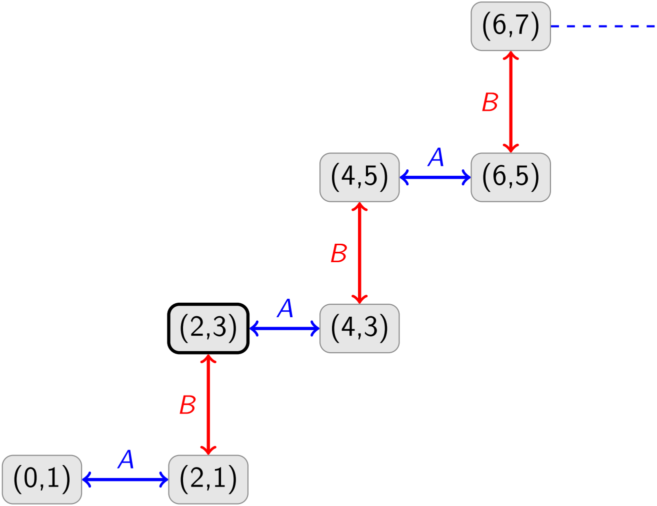
An Example

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)

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)

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: $f_E(\perp) = \perp$
- 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: $f_E(\perp) = \perp$
- 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).

Common Belief

Let $R_1, \ldots, R_n$ be relations on a set of state W. (Typically, each R_i is serial, transitive and Euclidean, but that is not crucial)

Common Belief

Let $R_1, \ldots, R_n$ be relations on a set of state W. (Typically, each R_i is serial, transitive and Euclidean, but that is not crucial)

$R_G = (\bigcup_{i \in G} R_i)^+$, where R^+ is the transitive closure of R.

$\mathcal{M}, w \models B\varphi$ iff for all $v \in W$, if $wR_G v$, then $\mathcal{M}, v \models \varphi$

Alternative Approaches

- Common *p*-belief
- Lewisian common belief

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."

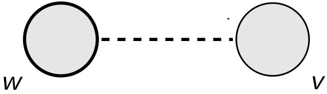
D. Monderer and D. Samet. *Approximating Common Knowledge with Common Beliefs*. Games and Economic Behavior (1989).

Representing Uncertainty

Finitely additive probability measures, upper and lower probability measures, Dempster-Shafer belief functions, imprecise probability measures (interval valued probabilities, sets of probability measures), possibility measures, plasuibility measures.

J. Halpern. *Reasoning about Uncertainty*. The MIT Press, 2003.

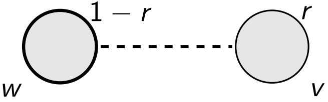
Models of Hard and Soft Information



$\mathcal{M} = \langle W, \{\Pi_i\}_{i \in \mathcal{A}} \rangle$
Π_i is agent i's partition with $\Pi_i(w)$ the partition cell containing w.

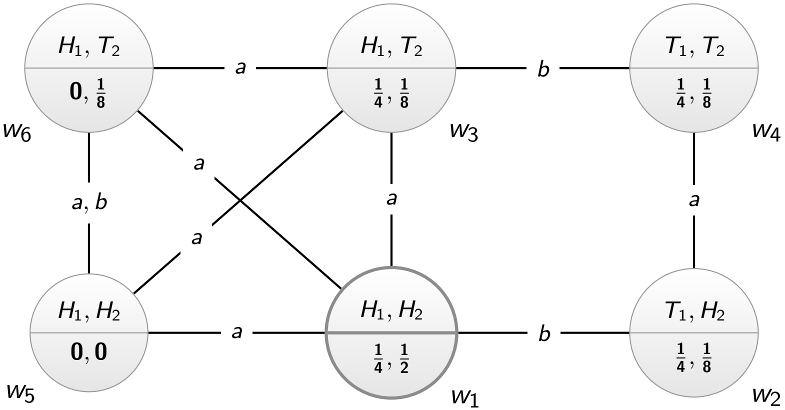
$K_i(E) = \{w \mid \Pi_i(w) \subseteq E\}$

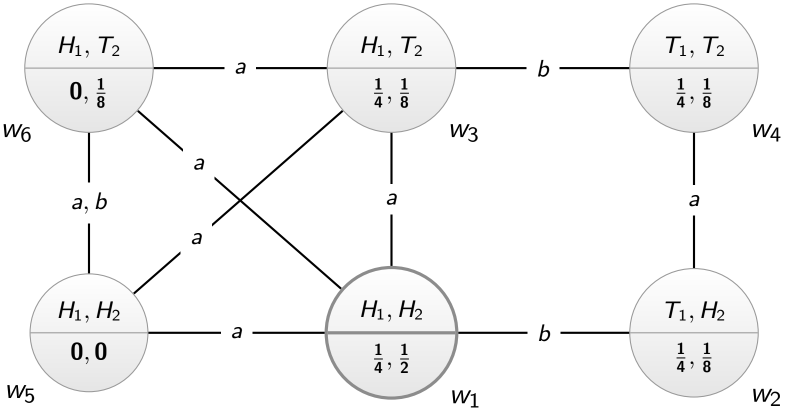
Models of Hard and Soft Information



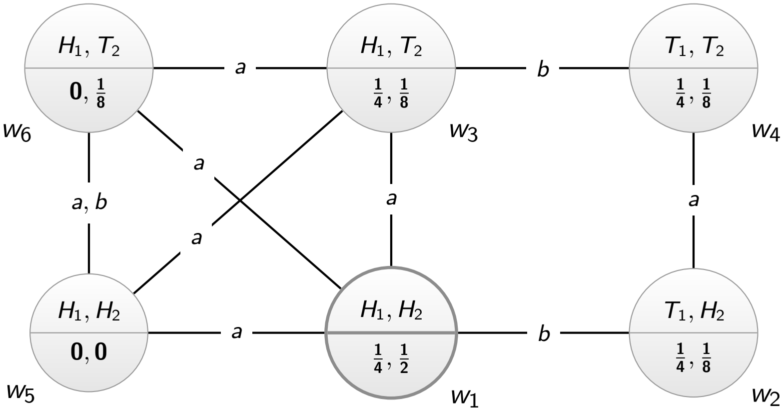
$\mathcal{M} = \langle W, \{\Pi_i\}_{i\in\mathcal{A}}, \{\pi_i\}_{i\in\mathcal{A}}\rangle$
for each i, $\pi_i : W \to [0,1]$ is a probability measure

$$B^p(E) = \{w \mid \pi_i(E \mid \Pi_i(w)) = \frac{\pi_i(E \cap \Pi_i(w))}{\pi_i(\Pi_i(w))} \geq p\}$$

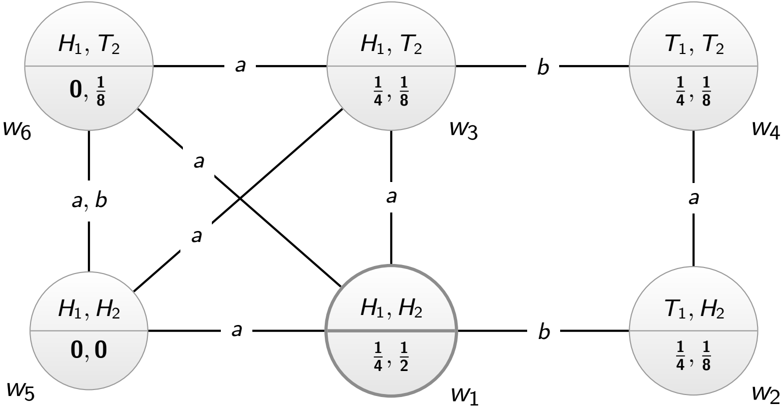

H_1, T_2
$\mathbf{0}, \frac{1}{8}$
w_6
H_1, T_2
$\frac{1}{4}, \frac{1}{8}$
w_3
T_1, T_2
$\frac{1}{4}, \frac{1}{8}$
w_4
H_1, H_2
$\mathbf{0}, \mathbf{0}$
w_5
H_1, H_2
$\frac{1}{4}, \frac{1}{2}$
w_1
T_1, H_2
$\frac{1}{4}, \frac{1}{8}$
w_2
a
b
a, b
a
a
a
a
a
b



- $\mathcal{M}, w_1 \models \neg K_a H_2 \wedge \neg K_a T_2 \wedge B_a^{\frac{1}{2}} H_2 \wedge B_a^{\frac{1}{2}} T_2$



- $\mathcal{M}, w_1 \models \neg K_a H_2 \wedge \neg K_a T_2 \wedge B_a^{\frac{1}{2}} H_2 \wedge B_a^{\frac{1}{2}} T_2$
- $\mathcal{M}, w_1 \models \neg K_b H_1 \wedge \neg K_b T_1 \wedge B_b^{\frac{4}{5}} H_1 \wedge B_b^{\frac{1}{5}} T_1$



- $\mathcal{M}, w_1 \models \neg K_a H_2 \wedge \neg K_a T_2 \wedge B_a^{\frac{1}{2}} H_2 \wedge B_a^{\frac{1}{2}} T_2$
- $\mathcal{M}, w_1 \models \neg K_b H_1 \wedge \neg K_b T_1 \wedge B_b^{\frac{4}{5}} H_1 \wedge B_b^{\frac{1}{5}} T_1$
- $\mathcal{M}, w_1 \models \neg K_a(K_b H_2 \vee K_b T_2) \wedge B_a^1(K_b H_2 \vee K_b T_2)$

1. $B_i^p(B_i^p(E)) = B_i^p(E)$

2. If $E \subseteq F$ then $B_i^p(E) \subseteq B_i^p(F)$

3. $\pi(E \mid B_i^p(E)) \geq p$

Common p-belief: definition

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

Common p-belief: definition

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

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

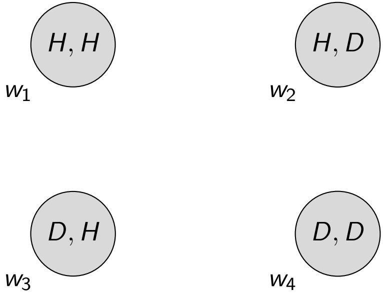
Common p-belief: definition

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

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

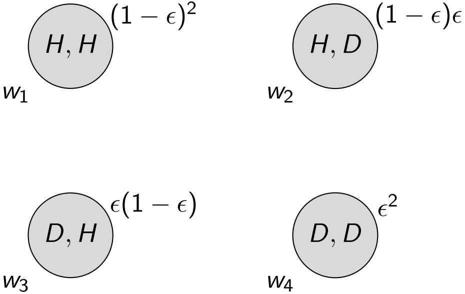
An event F is **common p-belief** at w if there exists and evident p-belief event E such that $w \in E$ and for all $i \in \mathcal{A}$, $E \subseteq B_i^p(F)$

Common p-belief: example



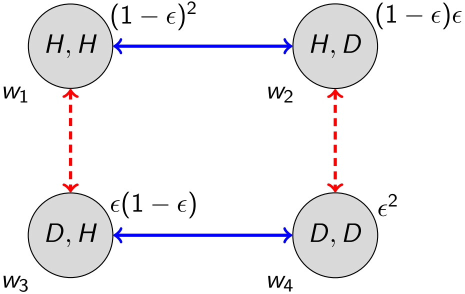
Two agents either hear (H) or don't hear (D) the announcement.

Common p-belief: example



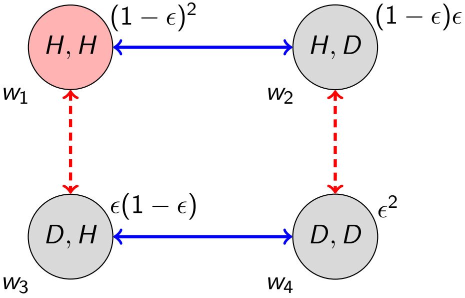
The probability that an agent hears is $1 - \epsilon$.

Common p-belief: example



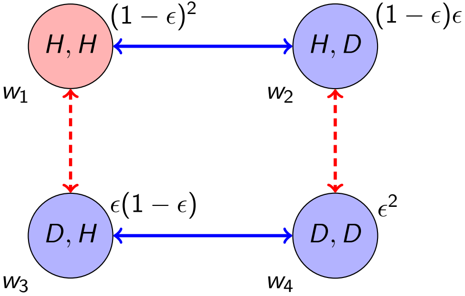
The agents *know* their "type".

Common p-belief: example



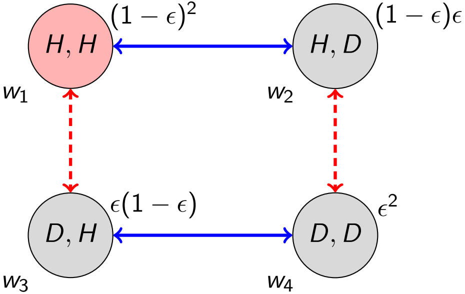
The event "everyone hears" ($E = \{w_1\}$)

Common p-belief: example



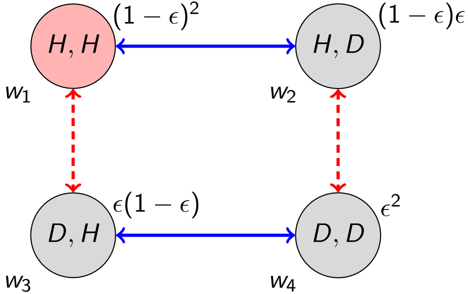
The event "everyone hears" ($E = \{w_1\}$) is **not** common knowledge

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 $i = 1, 2$

Agreeing to Disagree

"A group of agents cannot agree to disagree"

Agreeing to Disagree

"A group of agents cannot agree to disagree"

Theorem. Suppose that n agents share a common prior and have different private information. If there is common knowledge in the group of the posterior probabilities, then the posteriors must be equal.

Robert Aumann. *Agreeing to Disagree*. Annals of Statistics **4** (1976).

Agreeing to Disagree, generalized

Theorem. If the posteriors of an event X are common p-belief at some state w, then any two posteriors can differ by at most $2(1-p)$.

D. Samet and D. Monderer. *Approximating Common Knowledge with Common Beliefs*. Games and Economic Behavior, Vol. 1, No. 2, 1989.

Lewisian Common Belief

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)

State of Affairs

States of affairs are alternative specifications of how the world, as seen by the modeler, really might be.

These are primitives in Lewis's framework.

Given a state of affairs A, the proposition that A is in fact the case is denoted "A holds"

A indicates to i that φ

A is a "state of affairs"

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

(A1) 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

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

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

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

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 for φ

Reflexive Common Indicator for φ

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

Reflexive Common Indicator for φ

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

Reflexive Common Indicator for φ

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

Reflexive Common Indicator for φ

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

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 $i \not\in G$ 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 $i \not\in G$ 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.

Lewisian Common Belief in Game Theory

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

Lewisian Common Belief in Game Theory

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.

Lewisian Common Belief in Game Theory

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.

Lewisian Common Belief in Game Theory

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.

Lewisian Common Belief in Game Theory

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.

Lewisian Common Belief in Game Theory

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.

Lewisian Common Belief in Game Theory

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.

Lewisian Common Belief in Game Theory

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

		B	
		l	r
A	l		
	r		

Example

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

A: What should I do?

Example

			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

		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

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

A: What should *we* do?

Example

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

A: What should *we* do? **Team Reasoning**: an escape from the infinite regress? why should this "mode of reasoning" be endorsed?

Example

			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?

Reason to Believe Logic

$R_i(\varphi)$: "agent i has reason to believe φ"

Reason to Believe Logic

$R_i(\varphi)$: "agent i has reason to believe φ" this is interpreted as φ follows from rules (deductive, inductive, norm of practical reason) *endorsed by agent i*.

Reason to Believe Logic

$R_i(\varphi)$: "agent i has reason to believe φ" this is interpreted as φ follows from rules (deductive, inductive, norm of practical reason) *endorsed by agent i*.

Inference rules associated with the Reason-to-believe logic:
$inf(R) : \varphi, \psi \rightarrow \chi$

Reason to Believe Logic

$R_i(\varphi)$: "agent i has reason to believe φ" this is interpreted as φ follows from rules (deductive, inductive, norm of practical reason) *endorsed by agent i*.

Inference rules associated with the Reason-to-believe logic:
$inf(R) : \varphi, \psi \to \chi$

Assume each person's logic at least contains propositional logic:
$inf(R) : \varphi_1, \ldots \varphi_n, \neg(\varphi_1 \wedge \cdots \wedge \varphi_n \wedge \neg\psi) \to \psi$

Subject of the Proposition

Agent i is the **subject of the proposition** φ_i if φ_i makes an assertion about a current or future act of is will

Subject of the Proposition

Agent i is the **subject of the proposition** φ_i if φ_i makes an assertion about a current or future act of is will:

- a prediction about what i will choose in a future decision problem;
- a deontic statement about what i ought to choose;
- assert that i endorses some inference rule; or
- assert that i has reason to believe some proposition

Subject of the Proposition

Agent i is the **subject of the proposition** φ_i if φ_i makes an assertion about a current or future act of is will:

- a prediction about what i will choose in a future decision problem;
- a deontic statement about what i ought to choose;
- assert that i endorses some inference rule; or
- assert that i has reason to believe some proposition

$R_i(\varphi_i)$ vs. $R_j(\varphi_i)$: Suppose i reliable takes a bus every Monday.

Subject of the Proposition

Agent i is the **subject of the proposition** φ_i if φ_i makes an assertion about a current or future act of is will:

- a prediction about what i will choose in a future decision problem;
- a deontic statement about what i ought to choose;
- assert that i endorses some inference rule; or
- assert that i has reason to believe some proposition

$R_i(\varphi_i)$ vs. $R_j(\varphi_i)$: Suppose i reliable takes a bus every Monday. The other commuters may all make the inductive inference that i will take the bus next Monday (M_i).

Subject of the Proposition

Agent i is the **subject of the proposition** φ_i if φ_i makes an assertion about a current or future act of *i*s will:

- a prediction about what i will choose in a future decision problem;
- a deontic statement about what i ought to choose;
- assert that i endorses some inference rule; or
- assert that i has reason to believe some proposition

$R_i(\varphi_i)$ vs. $R_j(\varphi_i)$: Suppose i reliable takes a bus every Monday. The other commuters may all make the inductive inference that i will take the bus next Monday (M_i). In fact, we may assume that this is a *common mode of reasoning*, so everyone reliably makes the inference that i will catch the bus next Monday.

Subject of the Proposition

Agent i is the **subject of the proposition** φ_i if φ_i makes an assertion about a current or future act of is will:

- a prediction about what i will choose in a future decision problem;
- a deontic statement about what i ought to choose;
- assert that i endorses some inference rule; or
- assert that i has reason to believe some proposition

$R_i(\varphi_i)$ vs. $R_j(\varphi_i)$: Suppose i reliable takes a bus every Monday. The other commuters may all make the inductive inference that i will take the bus next Monday (M_i). In fact, we may assume that this is a *common mode of reasoning*, so everyone reliably makes the inference that i will catch the bus next Monday. So, $R_j(M_i)$, $R_iR_j(M_i)$

Subject of the Proposition

Agent i is the **subject of the proposition** φ_i if φ_i makes an assertion about a current or future act of is will:

- a prediction about what i will choose in a future decision problem;
- a deontic statement about what i ought to choose;
- assert that i endorses some inference rule; or
- assert that i has reason to believe some proposition

$R_i(\varphi_i)$ vs. $R_j(\varphi_i)$: Suppose i reliable takes a bus every Monday. The other commuters may all make the inductive inference that i will take the bus next Monday (M_i). In fact, we may assume that this is a *common mode of reasoning*, so everyone reliably makes the inference that i will catch the bus next Monday. So, $R_j(M_i)$, $R_iR_j(M_i)$, but i should still be *free* to choose whether he wants to take the bus on Monday, so $\neg R_i(M_i)$ and $\neg R_j(R_i(M_i))$, etc.

Common Reason to Believe

Awareness of Common Reason: for all $i \in G$ and all propositions φ,

$$R^G(\varphi) \Rightarrow R_i[R^G(\varphi)]$$

Common Reason to Believe

Awareness of Common Reason: for all $i \in G$ and all propositions φ,

$$R^G(\varphi) \Rightarrow R_i[R^G(\varphi)]$$

Authority of Common Reason: for all $i \in G$ and all propositions φ for which i is not the subject

$$inf(R_i) : R^G(\varphi) \rightarrow \varphi$$

Common Reason to Believe

Awareness of Common Reason: for all $i \in G$ and all propositions φ,

$$R^G(\varphi) \Rightarrow R_i[R^G(\varphi)]$$

Authority of Common Reason: for all $i \in G$ and all propositions φ for which i is not the subject

$$inf(R_i) : R^G(\varphi) \to \varphi$$

Common Attribution of Common Reason: for all $i \in G$, for all propositions φ for which i is not the subject

$$inf(R^G) : \varphi \to R_i(\varphi)$$

Common Reason to Believe to Common Belief

Theorem The three previous properties can generate any hierarchy of belief (i has reason to believe that j has reason to believe that... that φ) for any φ with $R^G(\varphi)$.

Team Maximising

$inf(R_i) : R^N[opt(v, N, s^N)]$,
$R^N[$ each $i \in N$ endorses team maximising with respect to N and v $]$,
$R^N[$ each member of N acts on reasons $] \rightarrow ought(i, s_i)$

Team Maximising

$inf(R_i)$: $R^N[opt(v, N, s^N)]$,
R^N[each $i \in N$ endorses team maximising with respect to N and v],
R^N[each member of N acts on reasons] $\rightarrow ought(i, s_i)$

$R_i[ought(i, s_i)]$: i has reason to choose s_i

Team Maximising

$inf(R_i) : R^N[opt(v, N, s^N)]$,
$R^N[$ each $i \in N$ endorses team maximising with respect to N and v $]$,
$R^N[$ each member of N acts on reasons $] \rightarrow ought(i, s_i)$

i acts on reasons if for all s_i, $R_i[ought(i, s_i)] \Rightarrow choice(i, s_i)$

Team Maximising

$inf(R_i) : R^N[opt(v, N, s^N)]$,
$R^N[$ each $i \in N$ endorses team maximising with respect to N and v $]$,
$R^N[$ each member of N acts on reasons $] \rightarrow ought(i, s_i)$

$opt(v, N, s^N)$: s^N is maximal for the group N w.r.t. v

Team Maximising

$inf(R_i) : R^N[opt(v, N, s^N)]$,
$R^N[$ each $i \in N$ endorses team maximising with respect to N and v $]$,
$R^N[$ each member of N acts on reasons $] \rightarrow ought(i, s_i)$

Recursive definition: i's endorsement of the rule depends on i having a reason to believe everyone else endorses the rule...

Many Questions!

Other modes of team reasoning, group identification, frames and team preferences

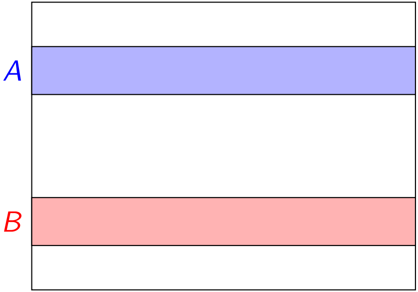
1. Common knowledge of rationality is not an event.

1. Common knowledge of rationality is not an event.
2. Hierarchies of beliefs in game situations.

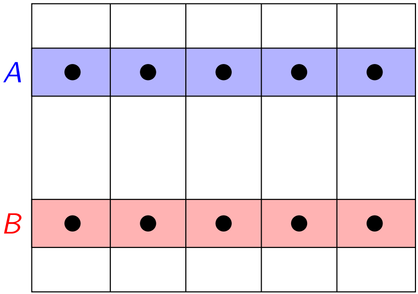
1. Common knowledge of rationality is not an event.
2. Hierarchies of beliefs in game situations.
3. What is the status of the epistemic models?

1. Common knowledge of rationality is not an event.
2. Hierarchies of beliefs in game situations.
3. What is the status of the epistemic models?
4. A paradox of self-reference in game theory

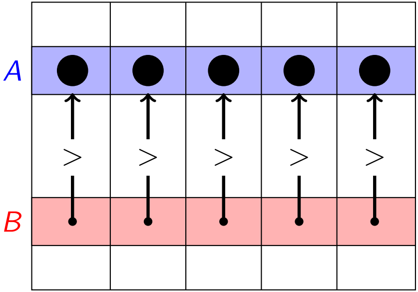
Dominance Reasoning



Dominance Reasoning



Dominance Reasoning



		Bob	
		L	R
Ann	U	2,2	4,1
	D	1,4	3,3

Game 1

		Bob	
		L	R
Ann	U	2,1	1,0
	D	1,0	0,1

Game 2

Bob

Ann	L	R
U	2,2	4,1
D	1,4	3,3

Game 1

Bob

Ann	L	R
U	2,1	1,0
D	1,0	0,1

Game 2

Game 1: U strictly dominates D and L strictly dominates R.

		Bob	
		L	*R*
Ann	*U*	2,2	4,1
	D	1,4	3,3

Game 1

		Bob	
		L	*R*
Ann	*U*	2,1	1,0
	D	1,0	0,1

Game 2

Game 1: U strictly dominates D and L strictly dominates R.

Game 2: U strictly dominates D

		Bob	
		L	R
Ann	U	2,2	4,1
	D	1,4	3,3

Game 1

		Bob	
		L	R
Ann	U	2,1	1,0
	D	1,0	0,1

Game 2

Game 1: U strictly dominates D and L strictly dominates R.

Game 2: U strictly dominates D, and *after removing* D, L strictly dominates R.

Bob

Ann		L	R
	U	2,2	4,1
	D	1,4	3,3

Game 1

Bob

Ann		L	R
	U	2,1	1,0
	D	1,0	0,1

Game 2

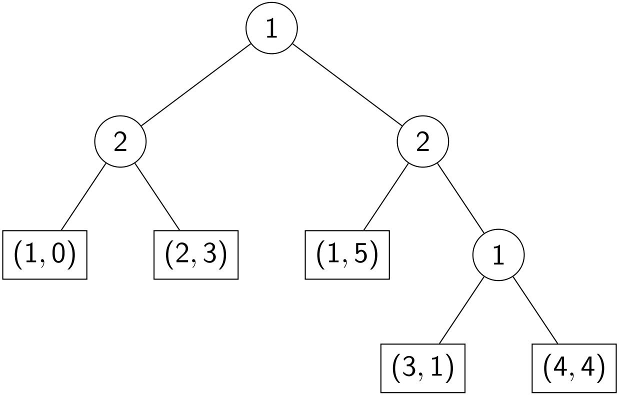
Game 1: U strictly dominates D and L strictly dominates R.

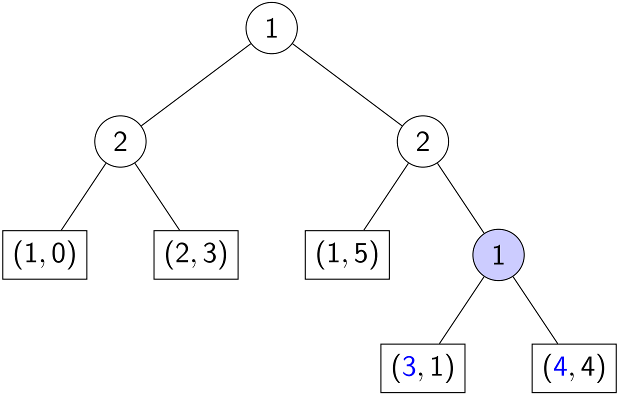
Game 2: U strictly dominates D, and *after removing* D, L strictly dominates R.

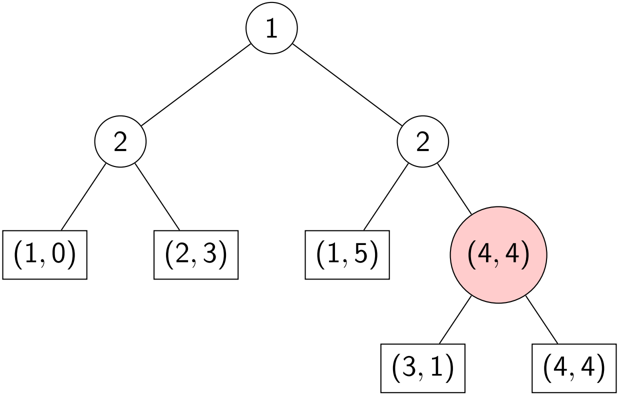
Theorem. In all models where the players are *rational* and there is *common belief of rationality*, the players choose strategies that survive iterative removal of strictly dominated strategies (and, conversely...).

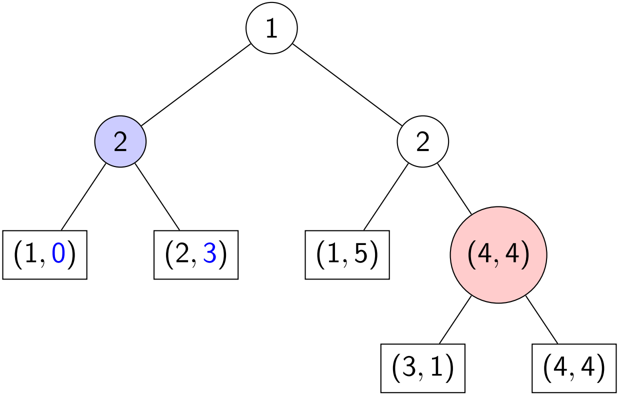
Backwards Induction

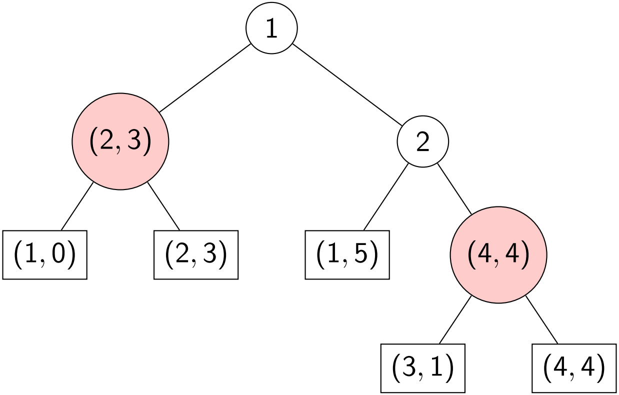
Invented by Zermelo, Backwards Induction is an iterative algorithm for "solving" and extensive game.

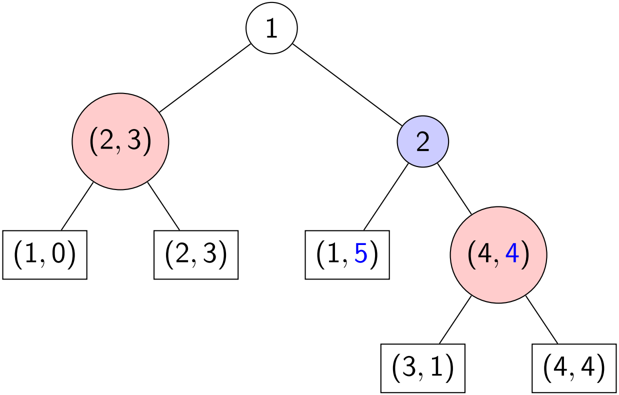

1
2
2
(1, 0)
(2, 3)
(1, 5)
1
(3, 1)
(4, 4)

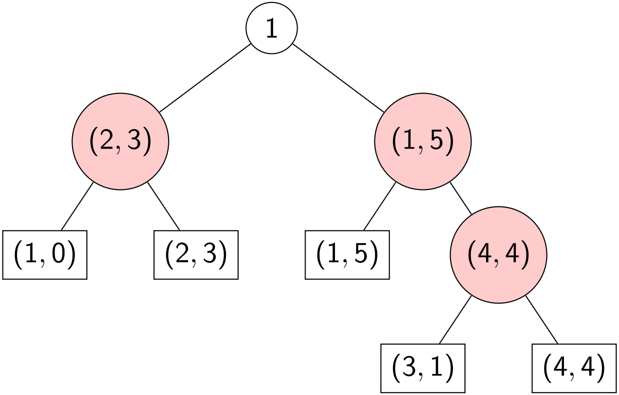

1
2
2
(1, 0)
(2, 3)
(1, 5)
1
(3, 1)
(4, 4)

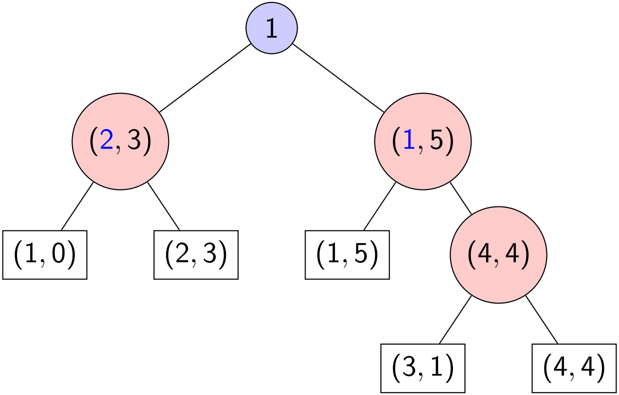

1
2
2
(1, 0)
(2, 3)
(1, 5)
(4, 4)
(3, 1)
(4, 4)

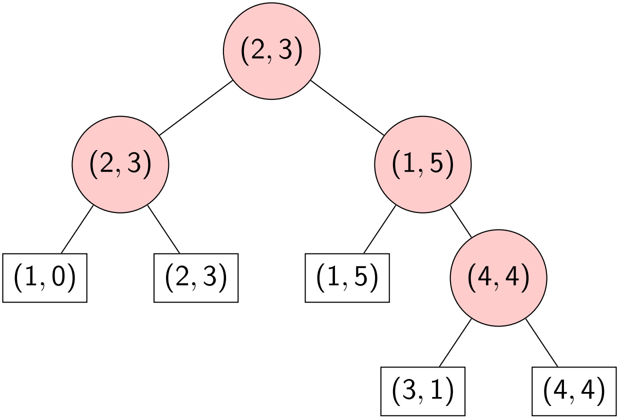

1
2
2
(1, 0)
(2, 3)
(1, 5)
(4, 4)
(3, 1)
(4, 4)

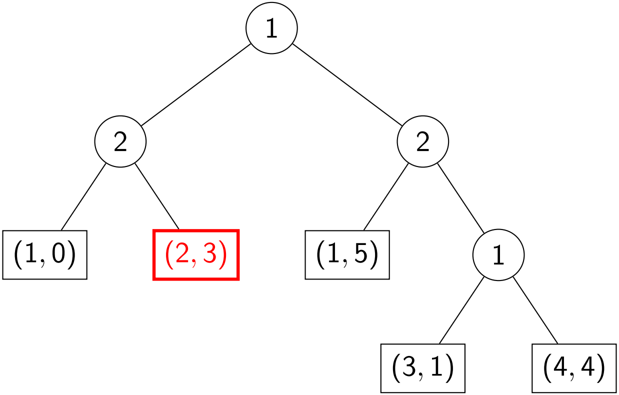

1
(2, 3)
2
(1, 0)
(2, 3)
(1, 5)
(4, 4)
(3, 1)
(4, 4)

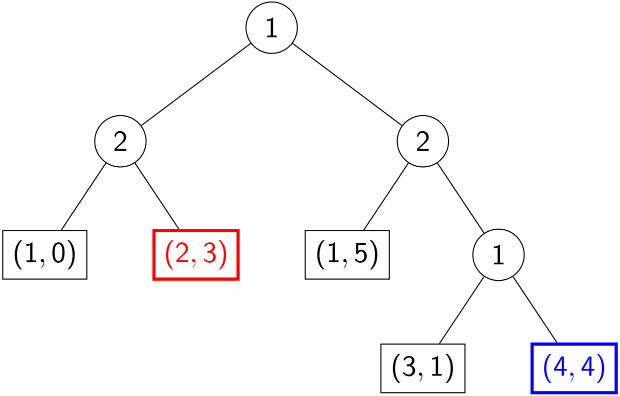

1
(2, 3)
2
(1, 0)
(2, 3)
(1, 5)
(4, 4)
(3, 1)
(4, 4)


1
(2, 3)
(1, 5)
(1, 0)
(2, 3)
(1, 5)
(4, 4)
(3, 1)
(4, 4)


1
(2, 3)
(1, 5)
(1, 0)
(2, 3)
(1, 5)
(4, 4)
(3, 1)
(4, 4)


(2, 3)
(2, 3)
(1, 5)
(1, 0)
(2, 3)
(1, 5)
(4, 4)
(3, 1)
(4, 4)


1
2
2
(1, 0)
(2, 3)
(1, 5)
1
(3, 1)
(4, 4)


1
2
2
(1, 0)
(2, 3)
(1, 5)
1
(3, 1)
(4, 4)

Hierarchies of Beliefs in a Game Situation

" A possible problem with the theory advocated here is the infinite regress. If he thinks I think he'll do x, then he'll do y. If he thinks I think he thinks I think he'll do y, etc.

Hierarchies of Beliefs in a Game Situation

" A possible problem with the theory advocated here is the infinite regress. If he thinks I think he'll do x, then he'll do y. If he thinks I think he thinks I think he'll do y, etc. It is true that a subjectivist Bayesian will have an opinion not only on his opponent's behavior, but also on his opponent's belief about his own behavior, his opponent's belief about his belief about his opponent's behavior, etc. (He also has opinions about the phase of the moon, tomorrow's weather and the winner of the next Superbowl).

Hierarchies of Beliefs in a Game Situation

" A possible problem with the theory advocated here is the infinite regress. If he thinks I think he'll do x, then he'll do y. If he thinks I think he thinks I think he'll do y, etc. It is true that a subjectivist Bayesian will have an opinion not only on his opponent's behavior, but also on his opponent's belief about his own behavior, his opponent's belief about his belief about his opponent's behavior, etc. (He also has opinions about the phase of the moon, tomorrow's weather and the winner of the next Superbowl). *However, in a single-play game, all aspects of his opinion except his opinion about his opponent's behavior are irrelevant, and can be ignored in the analysis by integrating them out of the joint opinion.*" (KL, pg. 239, my emphasis)

Hierarchies of Beliefs in a Game Situation

Belief hierarchies...

Hierarchies of Beliefs in a Game Situation

Belief hierarchies...

- are an explicit description (perhaps overly precise) of the contents of the players thoughts about her opponents

Hierarchies of Beliefs in a Game Situation

Belief hierarchies...

- are an explicit description (perhaps overly precise) of the contents of the players thoughts about her opponents
- represent the *outcome* of a reasoning process: the *reasons* rational players can point to in order to justify their choices

Hierarchies of Beliefs in a Game Situation

Belief hierarchies...

- are an explicit description (perhaps overly precise) of the contents of the players thoughts about her opponents
- represent the *outcome* of a reasoning process: the *reasons* rational players can point to in order to justify their choices
- track the back-and-forth reasoning that players are engaged in as they deliberate about what to do

Iterative Solution Concepts: Two Views

Iterative Solution Concepts: Two Views

Eg., Iterated removal of weakly/strictly dominated strategies

Iterative Solution Concepts: Two Views

Eg., Iterated removal of weakly/strictly dominated strategies

1. iterative procedures narrow down or assist in the search for a equilibria

2. iterative procedures represent a rational deliberation process

Iterative Solution Concepts: Two Views

Eg., Iterated removal of weakly/strictly dominated strategies

1. iterative procedures narrow down or assist in the search for a equilibria

   successive stages of strategy deletion may correspond to different levels of belief

2. iterative procedures represent a rational deliberation process

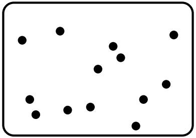
   successive stages of a strategy deletion can be interpreted as tracking successive steps of reasoning that players can perform

✓ Common knowledge of rationality is not an event.

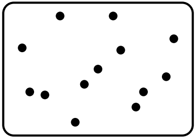
✓ Hierarchies of beliefs in game situations.

1. What is the status of the epistemic models?
2. A paradox of self-reference in game theory

Two key assumptions



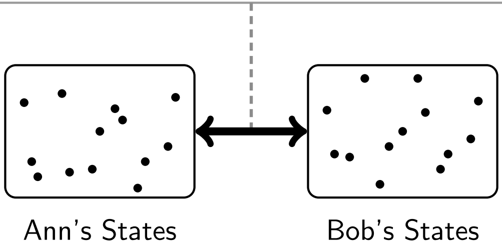
Ann's States



Bob's States

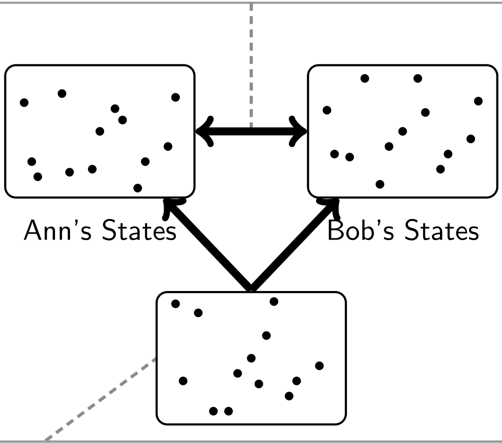
Two key assumptions

1. The players recognize that they are in a game situation



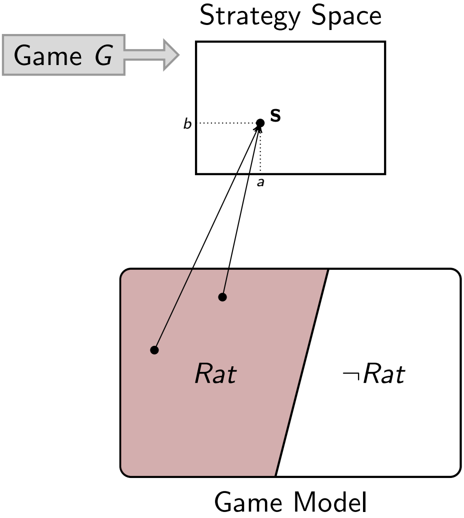
Two key assumptions

1. The players recognize that they are in a game situation



2. The players *agree* on a common initial model

Two key assumptions



- Each state in a game model is associated with a strategy profile *and* a description of the players beliefs.
- *Rat* is event that the players optimize (and there is common belief that they optimize)
- "The viewpoint is *descriptive*. Not 'why,' not 'should,' just *what*. Not that i does a *because* he believes E; simply that he does a and believes E."

What is a *State*?

Possible worlds, or states, are taken as primitive in Kripke structures.
But in many applications, we intuitively understand what a state *is*:

What is a *State*?

Possible worlds, or states, are taken as primitive in Kripke structures.
But in many applications, we intuitively understand what a state *is*:

Dynamic logic: a program state (assignment of values to variables)
Temporal logic: a moment in time
Distributed system: a sequence of local states for each process

What is a *State*?

Possible worlds, or states, are taken as primitive in Kripke structures.
But in many applications, we intuitively understand what a state *is*:

Dynamic logic: a program state (assignment of values to variables)
Temporal logic: a moment in time
Distributed system: a sequence of local states for each process

What about in *game situations*?

What is a *State*?

Possible worlds, or states, are taken as primitive in Kripke structures. But in many applications, we intuitively understand what a state *is*:

Dynamic logic: a program state (assignment of values to variables)
Temporal logic: a moment in time
Distributed system: a sequence of local states for each process

What about in *game situations*?
Answer: a *description* of the first-order and higher-order information of the players

R. Fagin, J. Halpern and M. Vardi. *Model theoretic analysis of knowledge*. Journal of the ACM 91 (1991).

Is an Epistemic Model "Common Knowledge"?

"The implicit assumption that the information partitions...are themselves common knowledge...constitutes no loss of generality... the assertion that each individual 'knows' the knowledge operators of all individual has no real substance; it is part of the framework."

R. Aumann. *Interactive Epistemology I & II*. International Journal of Game Theory (1999).

Is an Epistemic Model "Common Knowledge"?

"The implicit assumption that the information partitions...are themselves common knowledge...constitutes no loss of generality... the assertion that each individual 'knows' the knowledge operators of all individual has no real substance; it is part of the framework."

R. Aumann. *Interactive Epistemology I & II*. International Journal of Game Theory (1999).

"it is an informal but *meaningful* meta-assumption....It is not trivial at all to assume it is "common knowledge" which partition every player has."

A. Heifetz. *How canonical is the canonical model? A comment on Aumann's interactive epistemology*. International Journal of Game Theory (1999).

J. Halpern and W. Kets. *A logic for reasoning about ambiguity*. Artificial Intelligence, to appear.

J. Halpern and W. Kets. *Language and consensus*. working paper, 2013.

✓ Common knowledge of rationality is not an event.

✓ Hierarchies of beliefs in game situations.

✓ What is the status of the epistemic models?

1. A paradox of self-reference in game theory

Doesn't such talk of what Ann believes Bob believes about her, and so on, suggest that some kind of self-reference arises in games, similar to the well-known examples of self-reference in mathematical logic.

A. Brandenburger and H. J. Keisler. *An Impossibility Theorem on Beliefs in Games*. *Studia Logica* (2006).

A Paradox

Ann believes that Bob's strongest belief is
that Ann believes that Bob's strongest belief is false.

Does Ann believe that Bob's strongest belief is false?

$*$ A **strongest belief** is a belief that implies all other beliefs.

A. Brandenburger and H. J. Keisler. *An Impossibility Theorem on Beliefs in Games*. *Studia Logica* (2006).

A Paradox

Ann believes that Bob's strongest belief is
that Ann believes that Bob's strongest belief is false.

Does Ann believe that Bob's strongest belief is false? Suppose Yes.

A Paradox

Ann believes that Bob's strongest belief is that Ann believes that Bob's strongest belief is false.

Does Ann believe that Bob's strongest belief is false? Suppose Yes.

Then, Ann believes that it's not the case that Ann believes that Bob's strongest belief is false.

A Paradox

Ann believes that Bob's strongest belief is
that Ann believes that Bob's strongest belief is false.

Does Ann believe that Bob's strongest belief is false? Suppose Yes.

Then, Ann believes that it's not the case that Ann believes that Bob's strongest belief is false.

So, it's not the case that Ann believes that Bob's strongest belief is false.
$(B\neg B\varphi \to \neg B\varphi)$

A Paradox

Ann believes that Bob's strongest belief is
that Ann believes that Bob's strongest belief is false.

Does Ann believe that Bob's strongest belief is false? Suppose Yes.

Then, Ann believes that it's not the case that Ann believes that Bob's strongest belief is false.

So, it's not the case that Ann believes that Bob's strongest belief is false.
$(B\neg B\varphi \rightarrow \neg B\varphi)$

So, the answer is no.

A Paradox

Ann believes that Bob's strongest belief is
that Ann believes that Bob's strongest belief is false.

Does Ann believe that Bob's strongest belief is false? Suppose No.

A Paradox

Ann believes that Bob's strongest belief is
that Ann believes that Bob's strongest belief is false.

Does Ann believe that Bob's strongest belief is false? Suppose No.

Then, it's not the case that Ann believes it's not the case that Ann believes that Bob's strongest belief is false.

A Paradox

Ann believes that Bob's strongest belief is
that Ann believes that Bob's strongest belief is false.

Does Ann believe that Bob's strongest belief is false? Suppose No.

Then, it's not the case that Ann believes it's not the case that Ann believes that Bob's strongest belief is false.

So, Ann believes that Bob's strongest belief is false. $(\neg B \neg B\varphi \to B\varphi)$

A Paradox

Ann believes that Bob's strongest belief is
that Ann believes that Bob's strongest belief is false.

Does Ann believe that Bob's strongest belief is false? Suppose No.

Then, it's not the case that Ann believes it's not the case that Ann believes that Bob's strongest belief is false.

So, Ann believes that Bob's strongest belief is false. ($\neg B \neg B\varphi \rightarrow B\varphi$)

So, the answer must be yes.

- strongest belief

- strongest belief
- weakest belief

- strongest belief
- weakest belief
- craziest belief

- strongest belief
- weakest belief
- craziest belief
- all of Bob's belief

Is there a space of all possible interactive beliefs of a game?

Is there a space of all possible interactive beliefs of a game?

Two questions

Is there a space of all possible interactive beliefs of a game?

Two questions

- What exactly does "all possible" mean?

Is there a space of all possible interactive beliefs of a game?

Two questions

- What exactly does "all possible" mean?
  (Complete, Canonical, Universal)

Is there a space of all possible interactive beliefs of a game?

Two questions

- What exactly does "all possible" mean?
  (Complete, Canonical, Universal)
- Who cares?

Who Cares?

A. Brandenburger and E. Dekel. *Hierarchies of Beliefs and Common Knowledge*. Journal of Economic Theory (1993).

A. Heifetz and D. Samet. *Knoweldge Spaces with Arbitrarily High Rank*. Games and Economic Behavior (1998).

L. Moss and I. Viglizzo. *Harsanyi type spaces and final coalgebras constructed from satisfied theories*. EN in Theoretical Computer Science (2004).

A. Friendenberg. *When do type structures contain all hierarchies of beliefs?*. working paper (2007).

Who cares?

We think of a particular incomplete structure as giving the "context" in which the game is played.

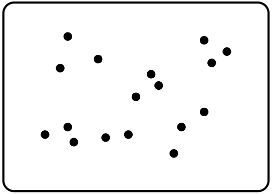
Who cares?

We think of a particular incomplete structure as giving the "context" in which the game is played. In line with Savage's Small-Worlds idea in decision theory [...], who the players are in the given game can be seen as a shorthand for their experiences before the game.

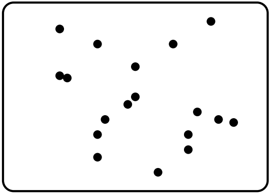
Who cares?

We think of a particular incomplete structure as giving the "context" in which the game is played. In line with Savage's Small-Worlds idea in decision theory [...], who the players are in the given game can be seen as a shorthand for their experiences before the game. The players' possible characteristics — including their possible types — then reflect the prior history or context. (Seen in this light, complete structures represent a special "context-free" case, in which there has been no narrowing down of types.) (pg. 319)

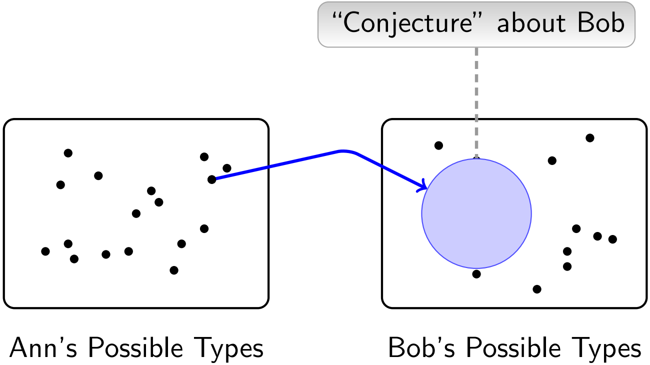
A. Brandenburger, A. Friedenberg, H. J. Keisler. *Admissibility in Games.* Econometrica (2008).



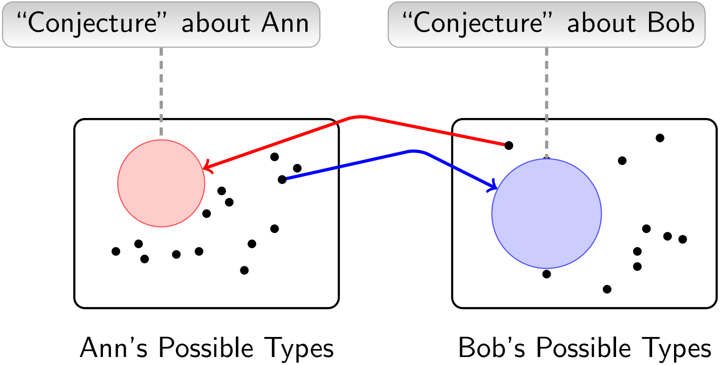
Ann's Possible Types



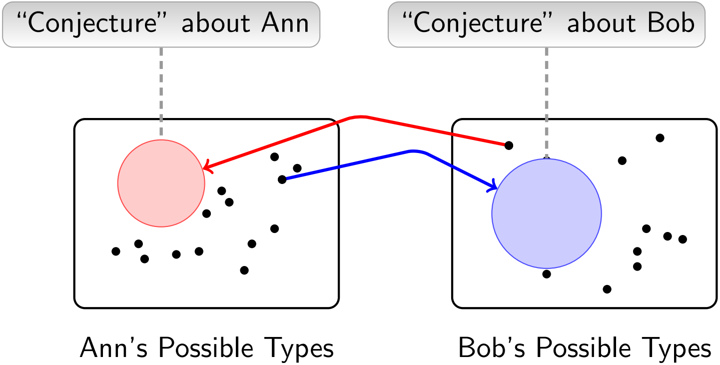
Bob's Possible Types



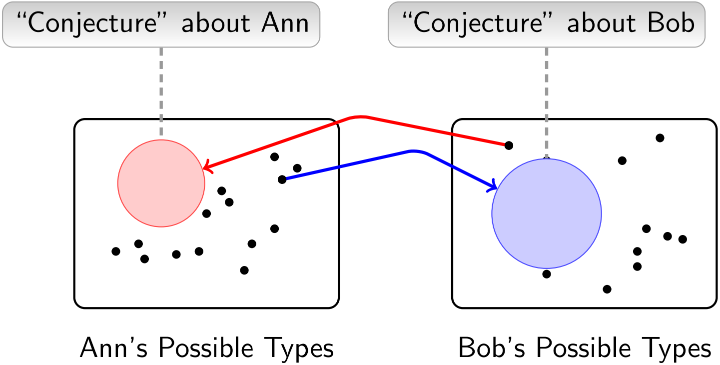
"Conjecture" about Bob
Ann's Possible Types
Bob's Possible Types



"Conjecture" about Ann
"Conjecture" about Bob
Ann's Possible Types
Bob's Possible Types



Is there a space where every *possible* conjecture is considered by *some* type?



Is there a space where every *possible* conjecture is considered by *some* type? It depends...

S. Abramsky and J. Zvesper. *From Lawvere to Brandenburger-Keisler: interactive forms of diagonalization and self-reference*. Proceedings of LOFT 2010.

EP. *Understanding the Brandenburger Keisler Pardox*. Studia Logica (2007).

Impossibility Results

Language: the (formal) language used by the players to formulate conjectures about their opponents.

Impossibility Results

Language: the (formal) language used by the players to formulate conjectures about their opponents.

Completeness: A model is **complete for a language** if every (consistent) statement in a player's language about an opponent is *considered* by some type.

Qualitative Type Spaces: $\langle T_a, T_b, \lambda_a, \lambda_b \rangle$

$\lambda_a : T_a \rightarrow \wp(T_b)$
$\lambda_b : T_b \rightarrow \wp(T_a)$

Qualitative Type Spaces: $\langle T_a, T_b, \lambda_a, \lambda_b \rangle$

$\lambda_a : T_a \rightarrow \wp(T_b)$
$\lambda_b : T_b \rightarrow \wp(T_a)$

x **believes** a set $Y \subseteq T_b$ if $\lambda_a(x) \subseteq Y$

x **assumes** a set $Y \subseteq T_b$ if $\lambda_a(x) = Y$

Impossibility Results

Impossibility 1 There is no complete interactive belief structure for the *powerset language*.

Proof. Cantor: there is no onto map from X to the nonempty subsets of X.

Impossibility Results

Impossibility 1 There is no complete interactive belief structure for the *powerset language*.

Proof. Cantor: there is no onto map from X to the nonempty subsets of X.

Impossibility 2 (Brandenburger and Keisler) There is no complete interactive belief structure for *first-order logic*.

Suppose that $\mathcal{C}_A \subseteq \wp(T_A)$ is a set of *conjectures* about Ann and $\mathcal{C}_B \subseteq \wp(T_B)$ a set of conjectures about Bob states.

Assume For all $X \in \mathcal{C}_A$ there is a $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$: "in state x_0, Ann has consistent beliefs"
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$: "in state x_0, Ann believes that Bob's strongest belief is that X"

Lemma. Under the above assumption, for each $X \in \mathcal{C}_A$ there is an x_0 such that

$$x_0 \in X \text{ iff there is a } y \in T_B \text{ such that } y \in \lambda_A(x_0) \text{ and } x_0 \in \lambda_B(y)$$

Claim. $x_0 \in X$ iff $\exists y \in T_B$, $y \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y)$

Assumption: For all $X \in \mathcal{C}_A$ there is a $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$

Suppose that $X \in \mathcal{C}_A$. Then there is an $x_0 \in T_A$ satisfying 1 and 2.

Claim. $x_0 \in X$ iff $\exists y \in T_B$, $y \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y)$

Assumption: For all $X \in \mathcal{C}_A$ there is a $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$

Suppose that $X \in \mathcal{C}_A$. Then there is an $x_0 \in T_A$ satisfying 1 and 2.

Suppose that $x_0 \in X$.

Claim. $x_0 \in X$ iff $\exists y \in T_B$, $y \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y)$

Assumption: For all $X \in \mathcal{C}_A$ there is a $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$

Suppose that $X \in \mathcal{C}_A$. Then there is an $x_0 \in T_A$ satisfying 1 and 2.

Suppose that $x_0 \in X$. By 1., $\lambda_A(x_0) \neq \emptyset$ so there is a $y_0 \in T_B$ such that $y_0 \in \lambda_A(x_0)$.

Claim. $x_0 \in X$ iff $\exists y \in T_B$, $y \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y)$

Assumption: For all $X \in \mathcal{C}_A$ there is a $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$

Suppose that $X \in \mathcal{C}_A$. Then there is an $x_0 \in T_A$ satisfying 1 and 2.

Suppose that $x_0 \in X$. By 1., $\lambda_A(x_0) \neq \emptyset$ so there is a $y_0 \in T_B$ such that $y_0 \in \lambda_A(x_0)$. We show that $x_0 \in \lambda_B(y_0)$.

Claim. $x_0 \in X$ iff $\exists y \in T_B$, $y \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y)$

Assumption: For all $X \in \mathcal{C}_A$ there is a $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$

Suppose that $X \in \mathcal{C}_A$. Then there is an $x_0 \in T_A$ satisfying 1 and 2.

Suppose that $x_0 \in X$. By 1., $\lambda_A(x_0) \neq \emptyset$ so there is a $y_0 \in T_B$ such that $y_0 \in \lambda_A(x_0)$. We show that $x_0 \in \lambda_B(y_0)$. By 2., we have $y_0 \in \lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$. Hence, $x_0 \in X = \lambda_B(y_0)$.

Claim. $x_0 \in X$ iff $\exists y \in T_B$, $y \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y)$

Assumption: For all $X \in \mathcal{C}_A$ there is a $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$

Suppose that $X \in \mathcal{C}_A$. Then there is an $x_0 \in T_A$ satisfying 1 and 2.

Suppose that $x_0 \in X$. By 1., $\lambda_A(x_0) \neq \emptyset$ so there is a $y_0 \in T_B$ such that $y_0 \in \lambda_A(x_0)$. We show that $x_0 \in \lambda_B(y_0)$. By 2., we have $y_0 \in \lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$. Hence, $x_0 \in X = \lambda_B(y_0)$.

Suppose that there is a $y_0 \in T_B$ such that $y_0 \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y_0)$.

Claim. $x_0 \in X$ iff $\exists y \in T_B$, $y \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y)$

Assumption: For all $X \in \mathcal{C}_A$ there is a $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$

Suppose that $X \in \mathcal{C}_A$. Then there is an $x_0 \in T_A$ satisfying 1 and 2.

Suppose that $x_0 \in X$. By 1., $\lambda_A(x_0) \neq \emptyset$ so there is a $y_0 \in T_B$ such that $y_0 \in \lambda_A(x_0)$. We show that $x_0 \in \lambda_B(y_0)$. By 2., we have $y_0 \in \lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$. Hence, $x_0 \in X = \lambda_B(y_0)$.

Suppose that there is a $y_0 \in T_B$ such that $y_0 \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y_0)$. By 2., $y_0 \in \lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$.

Claim. $x_0 \in X$ iff $\exists y \in T_B$, $y \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y)$

Assumption: For all $X \in \mathcal{C}_A$ there is a $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$

Suppose that $X \in \mathcal{C}_A$. Then there is an $x_0 \in T_A$ satisfying 1 and 2.

Suppose that $x_0 \in X$. By 1., $\lambda_A(x_0) \neq \emptyset$ so there is a $y_0 \in T_B$ such that $y_0 \in \lambda_A(x_0)$. We show that $x_0 \in \lambda_B(y_0)$. By 2., we have $y_0 \in \lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$. Hence, $x_0 \in X = \lambda_B(y_0)$.

Suppose that there is a $y_0 \in T_B$ such that $y_0 \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y_0)$. By 2., $y_0 \in \lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$. Hence, $x_0 \in \lambda_B(y_0) = X$.

Consider a first-order language $\mathcal{L}$ containing binary relational symbols $R_A(x, y)$ and $R_B(x, y)$ defining λ_A and λ_B, respectively.

Consider a first-order language $\mathcal{L}$ containing binary relational symbols $R_A(x, y)$ and $R_B(x, y)$ defining λ_A and λ_B, respectively.

$\mathcal{L}$ is interpreted over qualitative type structures where the interpretation of R_A is $\{(t, s) \mid t \in T_A, s \in T_B, \text{ and } s \in \lambda_A(t)\}$.

Consider a first-order language $\mathcal{L}$ containing binary relational symbols $R_A(x, y)$ and $R_B(x, y)$ defining λ_A and λ_B, respectively.

$\mathcal{L}$ is interpreted over qualitative type structures where the interpretation of R_A is $\{(t, s) \mid t \in T_A, s \in T_B, \text{ and } s \in \lambda_A(t)\}$.

Consider the formula φ in $\mathcal{L}$:

$$\varphi(x) := \exists y(R_A(x, y) \wedge R_B(y, x))$$

Consider a first-order language $\mathcal{L}$ containing binary relational symbols $R_A(x, y)$ and $R_B(x, y)$ defining λ_A and λ_B, respectively.

$\mathcal{L}$ is interpreted over qualitative type structures where the interpretation of R_A is $\{(t, s) \mid t \in T_A, s \in T_B, \text{ and } s \in \lambda_A(t)\}$.

Consider the formula φ in $\mathcal{L}$:

$$\varphi(x) := \exists y(R_A(x, y) \wedge R_B(y, x))$$

$\neg\varphi(x) := \forall y(R_A(x, y) \rightarrow \neg R_B(y, x))$: "Ann believes that Bob's strongest belief is *false*."

Proof of the Theorem

Suppose that $X \in \mathcal{C}_A$ is defined by the formula
$\neg\varphi(x) := \neg\exists y(R_A(x, y) \wedge R_B(y, x))$.

Proof of the Theorem

Suppose that $X \in \mathcal{C}_A$ is defined by the formula $\neg\varphi(x) := \neg\exists y(R_A(x,y) \wedge R_B(y,x))$.

There is an $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$: Ann's beliefs at x_0 are consistent.
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$: At x_0, Ann believes that Bob's strongest belief is that $X = \{x \mid \neg\varphi(x)\}$ (i.e., Ann believes that Bob's strongest belief is that Ann believes that Bob's strongest belief is false.)

Proof of the Theorem

Suppose that $X \in \mathcal{C}_A$ is defined by the formula
$\neg\varphi(x) := \neg\exists y(R_A(x,y) \wedge R_B(y,x))$.

There is an $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$: Ann's beliefs at x_0 are consistent.
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$: At x_0, Ann believes that Bob's strongest belief is that $X = \{x \mid \neg\varphi(x)\}$ (i.e., Ann believes that Bob's strongest belief is that Ann believes that Bob's strongest belief is false.)

$\neg\varphi(x_0)$ is true iff (def. of X) $x_0 \in X$

Proof of the Theorem

Suppose that $X \in \mathcal{C}_A$ is defined by the formula $\neg\varphi(x) := \neg\exists y(R_A(x,y) \wedge R_B(y,x))$.

There is an $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$: Ann's beliefs at x_0 are consistent.
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$: At x_0, Ann believes that Bob's strongest belief is that $X = \{x \mid \neg\varphi(x)\}$ (i.e., Ann believes that Bob's strongest belief is that Ann believes that Bob's strongest belief is false.)

$\neg\varphi(x_0)$ is true	iff (def. of X)	$x_0 \in X$
	iff (Lemma)	there is a $y \in T_B$ with $y \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y)$

Proof of the Theorem

Suppose that $X \in \mathcal{C}_A$ is defined by the formula $\neg\varphi(x) := \neg\exists y(R_A(x,y) \wedge R_B(y,x))$.

There is an $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$: Ann's beliefs at x_0 are consistent.
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$: At x_0, Ann believes that Bob's strongest belief is that $X = \{x \mid \neg\varphi(x)\}$ (i.e., Ann believes that Bob's strongest belief is that Ann believes that Bob's strongest belief is false.)

$\neg\varphi(x_0)$ is true	iff (def. of X)	$x_0 \in X$
	iff (Lemma)	there is a $y \in T_B$ with $y \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y)$
	iff (def. of $\varphi(x)$)	$\varphi(x_0)$ is true.