Epistemic Game Theory

Lecture 3

ESSLLI'12, Opole

Eric Pacuit   Olivier Roy

TiLPS, Tilburg University   MCMP, LMU Munich

ai.stanford.edu/~epacuit

http://olivier.amonbofis.net

August 8, 2012

Plan for the week

1. **Monday** Basic Concepts.
2. **Tuesday** Epistemics.
3. **Wednesday** Fundamentals of Epistemic Game Theory.
   - Models of all-out attitudes (cnt'd).
   - Common knowledge of Rationality and iterated strict dominance in the matrix.
   - (If time, o/w tomorrow.) Common knowledge of Rationality and backward induction (strict dominance in the tree).
4. **Thursday** Puzzles and Paradoxes.
5. **Friday** Extensions and New Directions.

A family of attitudes

A family of attitudes

- *Conditional Beliefs*: $\mathcal{M}, w \models B_i^{\varphi}\psi$ iff $\mathcal{M}, w' \models \psi$ for all $w' \in max_{\preceq_i}(\pi_i(w) \cap ||\varphi||)$.

A family of attitudes

- *Conditional Beliefs*: $\mathcal{M}, w \models B_i^{\varphi}\psi$ iff $\mathcal{M}, w' \models \psi$ for all $w' \in max_{\preceq_i}(\pi_i(w) \cap ||\varphi||)$.
- *Safe Belief*: $\mathcal{M}, w \models [\preceq]_i\varphi$ iff $\mathcal{M}, w' \models \varphi$ for all $w' \preceq_i$ w.

A family of attitudes

- *Conditional Beliefs*: $\mathcal{M}, w \models B_i^{\varphi}\psi$ iff $\mathcal{M}, w' \models \psi$ for all $w' \in max_{\preceq_i}(\pi_i(w) \cap ||\varphi||)$.
- *Safe Belief*: $\mathcal{M}, w \models [\preceq]_i\varphi$ iff $\mathcal{M}, w' \models \varphi$ for all $w' \preceq_i$ w.
- *Knowledge*: $\mathcal{M}, w \models K_i\varphi$ iff $\mathcal{M}, w' \models \varphi$ for all w' such that $w' \sim_i w$.

A family of attitudes

- *Conditional Beliefs*: $\mathcal{M}, w \models B_i^{\varphi}\psi$ iff $\mathcal{M}, w' \models \psi$ for all $w' \in max_{\preceq_i}(\pi_i(w) \cap ||\varphi||)$.
- *Safe Belief*: $\mathcal{M}, w \models [\preceq]_i\varphi$ iff $\mathcal{M}, w' \models \varphi$ for all $w' \preceq_i$ w.
- *Knowledge*: $\mathcal{M}, w \models K_i\varphi$ iff $\mathcal{M}, w' \models \varphi$ for all w' such that $w' \sim_i w$.

Plain beliefs defined:

$$B_i\psi \Leftrightarrow_{df} B_i^{\top}\varphi$$

A family of attitudes

- *Conditional Beliefs*: $\mathcal{M}, w \models B_i^\varphi \psi$ iff $\mathcal{M}, w' \models \psi$ for all $w' \in max_{\preceq_i}(\pi_i(w) \cap ||\varphi||)$.
- *Safe Belief*: $\mathcal{M}, w \models [\preceq]_i \varphi$ iff $\mathcal{M}, w' \models \varphi$ for all $w' \preceq_i$ w.
- *Knowledge*: $\mathcal{M}, w \models K_i \varphi$ iff $\mathcal{M}, w' \models \varphi$ for all w' such that $w' \sim_i w$.

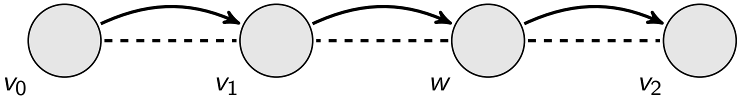
Plain beliefs defined:

$$B_i \psi \Leftrightarrow_{df} B_i^\top \varphi$$

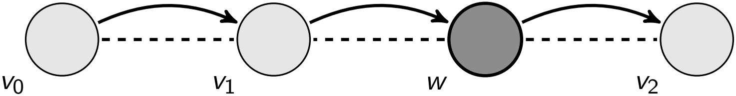
Conditional beliefs defined:

$$B_i^\varphi \psi \Leftrightarrow_{df} \langle K \rangle_i \varphi \rightarrow \langle K \rangle_i (\varphi \wedge [\preceq]_i (\varphi \rightarrow \psi))$$

Soft attitudes

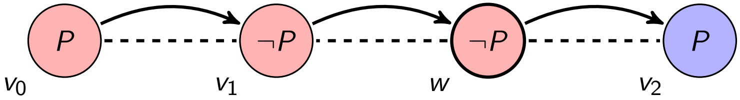


Soft attitudes



Suppose that w is the current state.

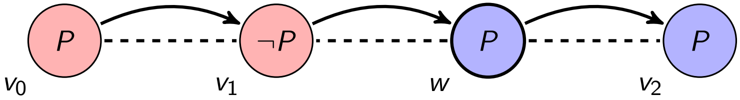
Soft attitudes



Suppose that w is the current state.

- **Belief** ($B_i p$)

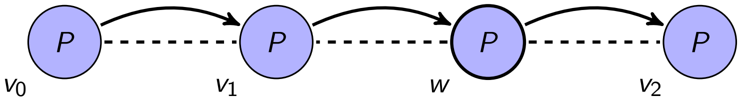
Soft attitudes



Suppose that w is the current state.

- **Belief** ($B_i p$)
- **Safe Belief** ($[\preceq]_i p$)

Soft attitudes

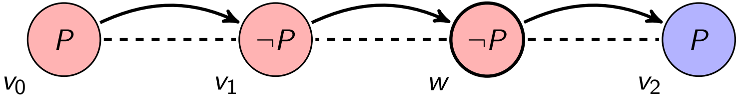


Suppose that w is the current state.

- **Belief** ($B_i p$)
- **Safe Belief** ($[\preceq]_i p$)
- **Knowledge** ($K_i p$)

Properties of Soft Attitudes

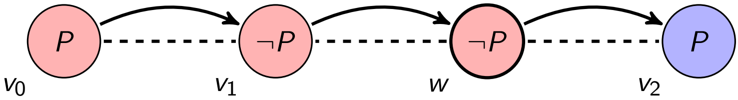
Beliefs and conditional beliefs can be mistaken.



$$\not\models B_i\varphi \rightarrow \varphi$$

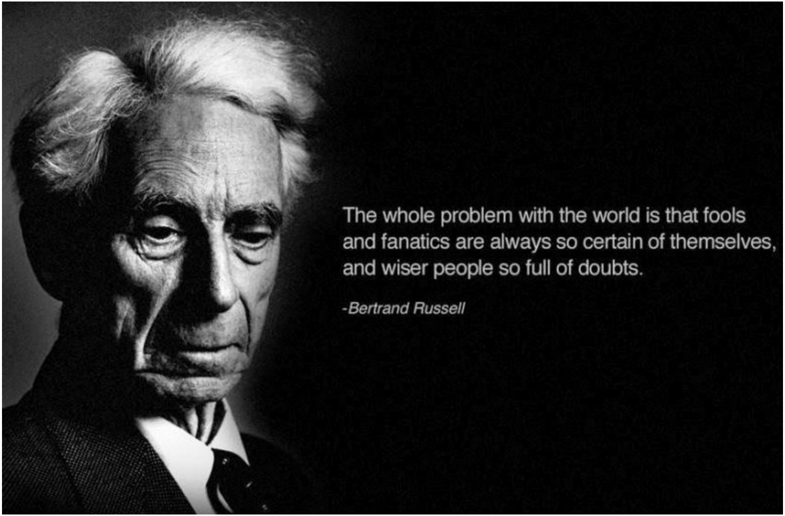
Properties of Soft Attitudes

Beliefs and conditional beliefs are fully introspective.



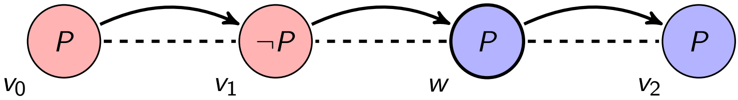
$$\models B_i\varphi \to B_iB_i\varphi$$

$$\models \neg B_i\varphi \to B_i\neg B_i\varphi$$


The whole problem with the world is that fools and fanatics are always so certain of themselves, and wiser people so full of doubts.
-Bertrand Russell

Properties of Soft Attitudes

Safe Belief is truthful and positively introspective.

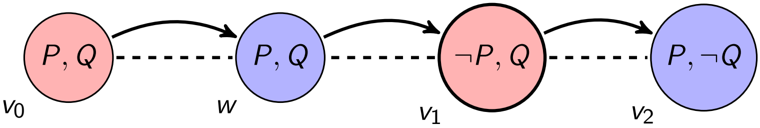


$$\models [\preceq]_i\varphi \rightarrow \varphi$$

$$\models [\preceq]_i\varphi \rightarrow [\preceq]_i[\preceq]_i\varphi$$

Properties of Soft Attitudes

Safe Belief is **not** negatively introspective.



$$\not\models \neg[\preceq]_i\varphi \rightarrow [\preceq]_i\neg[\preceq]_i\varphi$$

but...

$$\models B_i\varphi \leftrightarrow B_i[\preceq]_i\varphi$$

Higher-order attitudes and common knowledge.

"*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).

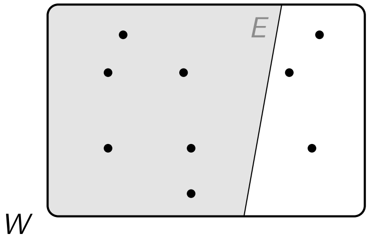
P. Vanderschraaf and G. Sillari. *"Common Knowledge", The Stanford Encyclopedia of Philosophy (2009)*.
http://plato.stanford.edu/entries/common-knowledge/.

The “Standard” Account

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

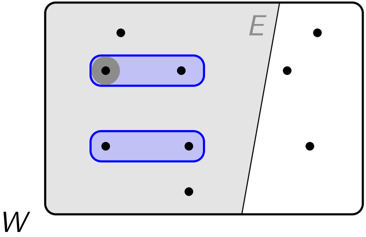
R. Fagin, J. Halpern, Y. Moses and M. Vardi. *Reasoning about Knowledge*. MIT Press, 1995.

The "Standard" Account

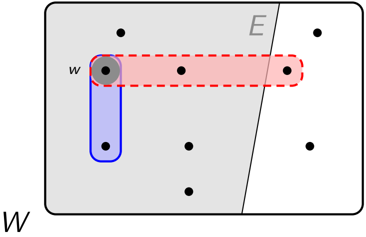


An **event**/**proposition** is any (definable) subset $E \subseteq W$

The "Standard" Account

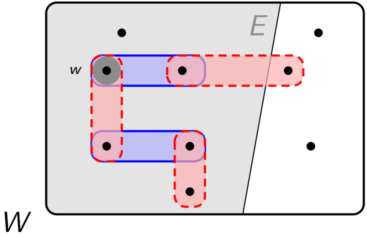


The "Standard" Account



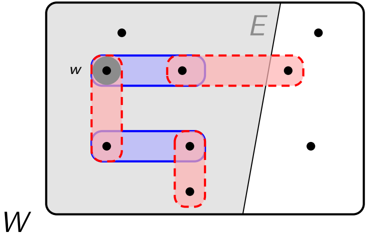
$w \models K_A(E)$ and $w \not\models K_B(E)$

The “Standard” Account



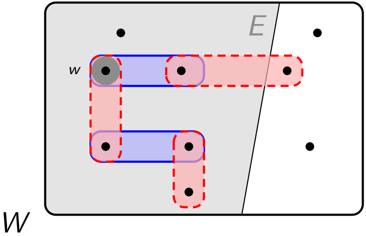
The model also describes the agents’ **higher-order knowledge/beliefs**

The "Standard" Account



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

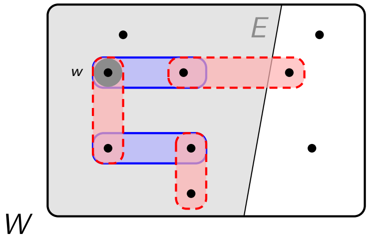
The "Standard" Account



Common Knowledge:

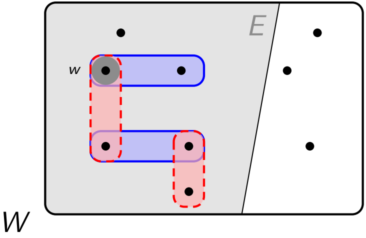
$$C(E) = \bigcap_{m \geq 0} K^m(E)$$

The "Standard" Account



$w \models K(E)$ $\quad$ $w \not\models C(E)$

The "Standard" Account



$w \models C(E)$

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; pause 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, 1999 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:

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

With $E\varphi := \bigwedge_{i \in \mathrm{Ag}} K_i\varphi$.

Some General Remarks

Some General Remarks

- Two broad families of models of higher-order information:
  - Type spaces. (probabilistic)
  - Plausibility models. (all-out)
- There's also a natural notion of qualitative type spaces, just like a natural probabilistic version of plausibility models. No strict separation between the two ways of thinking about information in interaction.

Some General Remarks

- Two broad families of models of higher-order information:
  - Type spaces. (probabilistic)
  - Plausibility models. (all-out)
- There's also a natural notion of qualitative type spaces, just like a natural probabilistic version of plausibility models. No strict separation between the two ways of thinking about information in interaction.
- In both the notion of a state is crucial. A state encodes:
  1. The "non-epistemic facts". Here, mostly: what the agents are playing.
  2. What the agents know and/or believe about 1.
  3. What the agents know and/or believe about 2.
  4. ...

Now let's do epistemics in games...

The Epistemic or Bayesian View on Games

- Traditional game theory:
  Actions, outcomes, preferences, solution concepts.

- Epistemic game theory:
  Actions, outcomes, preferences, beliefs, choice rules.

The Epistemic or Bayesian View on Games

- Traditional game theory:
  Actions, outcomes, preferences, solution concepts.

- Epistemic game theory:
  Actions, outcomes, preferences, beliefs, choice rules.
  $:=$ (interactive) decision problem: choice rule and higher-order information.

The Epistemic or Bayesian View on Games

- Traditional game theory:
  Actions, outcomes, preferences, solution concepts.
- Decision theory:
  Actions, outcomes, preferences, beliefs, choice rules.
- Epistemic game theory:
  Actions, outcomes, preferences, beliefs, choice rules.
  := (interactive) decision problem: choice rule and higher-order information.

Beliefs, Choice Rules, Rationality

What do we mean when we say that a player chooses rationally?
That she follows some given choice rules.

- Maximization of expected utility, (Strict) dominance reasoning, Admissibility, etc.

Beliefs, Choice Rules, Rationality

What do we mean when we say that a player chooses rationally?
That she follows some given choice rules.

- Maximization of expected utility, (Strict) dominance reasoning, Admissibility, etc.

In game models:

- The model describes the choices and (higher-order) beliefs/attitudes at each state.
- It is the choice rules that determine whether the choice made at each state is "rational" or not.
  - An agent can be rational at a state given one choice rule, but irrational given the other.
  - Rationality in this sense is not built in the models.

Rationality

Let $G = \langle N, \{S_i\}_{i \in N}, \{u_i\}_{i \in N} \rangle$ be a strategic game and $\mathcal{T} = \langle \{T_i\}_{i \in N}, \{\lambda_i\}_{i \in N}, S \rangle$ a type space for G.
For each $t_i \in T_i$, we can define a probability measure $p_{t_i} \in \Delta(S_{-i})$:

$$p_{t_i}(s_{-i}) = \sum_{t_{-i} \in T_{-i}} \lambda_i(t_i)(s_{-i}, t_{-i})$$

The set of states (pairs of strategy profiles and type profiles) where player i chooses **rationally** is:

$$\mathrm{Rat_i} := \{(s_i, t_i) \mid s_i \text{ is a best response to } p_{t_i}\}$$

The event that all players are *rational* is
$\mathrm{Rat} = \{(s, t) \mid \text{ for all } i, (s_i, t_i) \in \mathrm{Rat_i}\}$.

Rationality

Let $G = \langle N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N}\rangle$ be a strategic game and $\mathcal{T} = \langle \{T_i\}_{i\in N}, \{\lambda_i\}_{i\in N}, S\rangle$ a type space for G.
For each $t_i \in T_i$, we can define a probability measure $p_{t_i} \in \Delta(S_{-i})$:

$$p_{t_i}(s_{-i}) = \sum_{t_{-i}\in T_{-i}} \lambda_i(t_i)(s_{-i}, t_{-i})$$

The set of states (pairs of strategy profiles and type profiles) where player i chooses **rationally** is:

$$\mathrm{Rat_i} \;:=\; \{(s_i, t_i) \mid s_i \text{ is a best response to } p_{t_i}\}$$

The event that all players are *rational* is
$\mathrm{Rat} = \{(s,t) \mid \text{ for all } i,\ (s_i, t_i) \in \mathrm{Rat_i}\}$.

- **Types, as opposed to players, are rational or not at a given state**.

Rationality and common belief of rationality (RCBR) in the matrix

IESDS

		2		
		l	c	r
1	t	3, 3	1, 1	0, 0
	m	1,1	3, 3	1, 0
	m	0, 4	0, 0	4, 0

IESDS

		2		
		l	c	r
1	t	3, 3	1, 1	0, 0
	m	1,1	3, 3	1, 0
	m	0, 4	0, 0	4, 0

$\rightarrowtail$

		2	
		l	c
1	t	3, 3	1, 1
	m	1,1	3, 3
	b	0, 4	0, 0

IESDS

Player 1 (rows), Player 2 (columns):

	l	c	r
t	3, 3	1, 1	0, 0
m	1,1	3, 3	1, 0
m	0, 4	0, 0	4, 0

$\rightarrowtail$

	l	c
t	3, 3	1, 1
m	1,1	3, 3
b	0, 4	0, 0

$\rightarrowtail$

	l	c
t	3, 3	1, 1
m	1,1	3, 3

1's types

$\lambda_1(t_1)$

	l	c	r
s_1	0.5	0.5	0
s_2	0	0	0
s_3	0	0	0

$\lambda_1(t_2)$

	l	c	r
s_1	0	0.5	0
s_2	0	0	0.5
s_3	0	0	0

2's types

$\lambda_2(s_1)$

	t	m	b
t_1	0.5	0.5	0
t_2	0	0	0

$\lambda_2(s_2)$

	t	m	b
t_1	0.25	0.25	0
t_2	0.25	0.25	0

$\lambda_2(s_3)$

	t	m	b
t_1	0.5	0	0
t_2	0	0	0.5

2

1		l	c	r
	t	3, 3	1, 1	0, 0
	m	1,1	3, 3	1, 0
	b	0, 4	0, 0	4, 0

$\lambda_2(s_1)$

	t	m	b
t_1	0.5	0.5	0
t_2	0	0	0

$\lambda_2(s_2)$

	t	m	b
t_1	0.25	0.25	0
t_2	0.25	0.25	0

$\lambda_2(s_3)$

	t	m	b
t_1	0.5	0	0
t_2	0	0	0.5

Player 1 (rows), Player 2 (columns):

	l	c	r
t	3, 3	1, 1	0, 0
m	1,1	3, 3	1, 0
b	0, 4	0, 0	4, 0

$\lambda_2(s_1)$

	t	m	b
t_1	0.5	0.5	0
t_2	0	0	0

$\lambda_2(s_2)$

	t	m	b
t_1	0.25	0.25	0
t_2	0.25	0.25	0

$\lambda_2(s_3)$

	t	m	b
t_1	0.5	0	0
t_2	0	0	0.5

- l and c are rational for both s_1 and s_2.

1 \ 2	l	c	r
t	3, 3	1, 1	0, 0
m	1,1	3, 3	1, 0
b	0, 4	0, 0	4, 0

$\lambda_2(s_1)$

	t	m	b
t_1	0.5	0.5	0
t_2	0	0	0

$\lambda_2(s_2)$

	t	m	b
t_1	0.25	0.25	0
t_2	0.25	0.25	0

$\lambda_2(s_3)$

	t	m	b
t_1	0.5	0	0
t_2	0	0	0.5

- l and c are rational for both s_1 and s_2.

			2	
		l	c	r
1	t	3, 3	1, 1	0, 0
	m	1,1	3, 3	1, 0
	b	0, 4	0, 0	4, 0

$\lambda_2(s_1)$	t	m	b
t_1	0.5	0.5	0
t_2	0	0	0

$\lambda_2(s_2)$	t	m	b
t_1	0.25	0.25	0
t_2	0.25	0.25	0

$\lambda_2(s_3)$	t	m	b
t_1	0.5	0	0
t_2	0	0	0.5

- l and c are rational for both s_1 and s_2.
- l is the only rational action for s_3.

		2		
		l	c	r
1	t	3, 3	1, 1	0, 0
	m	1,1	3, 3	1, 0
	b	0, 4	0, 0	4, 0

$\lambda_2(s_1)$	t	m	b
t_1	0.5	0.5	0
t_2	0	0	0

$\lambda_2(s_2)$	t	m	b
t_1	0.25	0.25	0
t_2	0.25	0.25	0

$\lambda_2(s_3)$	t	m	b
t_1	0.5	0	0
t_2	0	0	0.5

- l and c are rational for both s_1 and s_2.
- l is the only rational action for s_3.
- Whatever her type, it is never rational to play r for 2.

Player 1 (rows) vs. player 2 (columns):

	l	c	r
t	3, 3	1, 1	0, 0
m	1,1	3, 3	1, 0
b	0, 4	0, 0	4, 0

$\lambda_1(t_1)$

	l	c	r
s_1	0.5	0.5	0
s_2	0	0	0
s_3	0	0	0

$\lambda_1(t_2)$

	l	c	r
s_1	0	0.5	0
s_2	0	0	0.5
s_3	0	0	0

		2		
		l	c	r
1	t	3, 3	1, 1	0, 0
	m	1,1	3, 3	1, 0
	b	0, 4	0, 0	4, 0

$\lambda_1(t_1)$		l	c	r
	s_1	0.5	0.5	0
	s_2	0	0	0
	s_3	0	0	0

$\lambda_1(t_2)$		l	c	r
	s_1	0	0.5	0
	s_2	0	0	0.5
	s_3	0	0	0

- t and m are rational for t_1.

		2		
		l	c	r
1	t	3, 3	1, 1	0, 0
	m	1,1	3, 3	1, 0
	b	0, 4	0, 0	4, 0

$\lambda_1(t_1)$

	l	c	r
s_1	0.5	0.5	0
s_2	0	0	0
s_3	0	0	0

$\lambda_1(t_2)$

	l	c	r
s_1	0	0.5	0
s_2	0	0	0.5
s_3	0	0	0

- t and m are rational for t_1.

		2		
		l	c	r
1	t	3, 3	1, 1	0, 0
	m	1,1	3, 3	1, 0
	b	0, 4	0, 0	4, 0

$\lambda_1(t_1)$

	l	c	r
s_1	0.5	0.5	0
s_2	0	0	0
s_3	0	0	0

$\lambda_1(t_2)$

	l	c	r
s_1	0	0.5	0
s_2	0	0	0.5
s_3	0	0	0

- t and m are rational for t_1.
- m and b are rational for t_2.

$\lambda_2(s_1)$

	t	m	b
t_1	0.5	0.5	0
t_2	0	0	0

$\lambda_2(s_2)$

	t	m	b
t_1	0.25	0.25	0
t_2	0.25	0.25	0

$\lambda_2(s_3)$

	t	m	b
t_1	0.5	0	0
t_2	0	0	0.5

$\lambda_2(s_1)$

	t	m	b
t_1	0.5	0.5	0
t_2	0	0	0

$\lambda_2(s_2)$

	t	m	b
t_1	0.25	0.25	0
t_2	0.25	0.25	0

$\lambda_2(s_3)$

	t	m	b
t_1	0.5	0	0
t_2	0	0	0.5

- All of 2's types believe that 1 is rational.

$\lambda_1(t_1)$

	l	c	r
s_1	0.5	0.5	0
s_2	0	0	0
s_3	0	0	0

$\lambda_1(t_2)$

	l	c	r
s_1	0	0.5	0
s_2	0	0	0.5
s_3	0	0	0

$\lambda_1(t_1)$

	l	c	r
s_1	0.5	0.5	0
s_2	0	0	0
s_3	0	0	0

$\lambda_1(t_2)$

	l	c	r
s_1	0	0.5	0
s_2	0	0	0.5
s_3	0	0	0

- Type t_1 of 1 believes that 2 is rational.

$\lambda_1(t_1)$

	l	c	r
s_1	0.5	0.5	0
s_2	0	0	0
s_3	0	0	0

$\lambda_1(t_2)$

	l	c	r
s_1	0	0.5	0
s_2	0	0	0.5
s_3	0	0	0

- Type t_1 of 1 believes that 2 is rational.
- But type t_2 doesn't! (1/2 probability that 2 is playing r.)

$\lambda_2(s_1)$

	t	m	b
t_1	0.5	0.5	0
t_2	0	0	0

$\lambda_2(s_2)$

	t	m	b
t_1	0.25	0.25	0
t_2	0.25	0.25	0

$\lambda_2(s_3)$

	t	m	b
t_1	0.5	0	0
t_2	0	0	0.5

$\lambda_2(s_1)$

	t	m	b
t_1	0.5	0.5	0
t_2	0	0	0

$\lambda_2(s_2)$

	t	m	b
t_1	0.25	0.25	0
t_2	0.25	0.25	0

$\lambda_2(s_3)$

	t	m	b
t_1	0.5	0	0
t_2	0	0	0.5

- Only type s_1 of 2 believes that 1 is rational and that 1 believes that 2 is also rational.

$\lambda_1(t_1)$

	l	c	r
s_1	0.5	0.5	0
s_2	0	0	0
s_3	0	0	0

$\lambda_1(t_2)$

	l	c	r
s_1	0	0.5	0
s_2	0	0	0.5
s_3	0	0	0

$\lambda_1(t_1)$

	l	c	r
s_1	0.5	0.5	0
s_2	0	0	0
s_3	0	0	0

$\lambda_1(t_2)$

	l	c	r
s_1	0	0.5	0
s_2	0	0	0.5
s_3	0	0	0

- Type t_1 of 1 believes that 2 is rational and that 2 believes that 1 believes that 2 is rational.

2

1		l	c	r
	t	3, 3	1, 1	0, 0
	m	1,1	3, 3	1, 0
	b	0, 4	0, 0	4, 0

$\lambda_1(t_1)$

	l	c	r
s_1	0.5	0.5	0
s_2	0	0	0
s_3	0	0	0

$\lambda_2(s_1)$

	t	m	b
t_1	0.5	0.5	0
t_2	0	0	0

2				
		l	c	r
1	t	3, 3	1, 1	0, 0
	m	1,1	3, 3	1, 0
	b	0, 4	0, 0	4, 0

$\lambda_1(t_1)$

	l	c	r
s_1	0.5	0.5	0
s_2	0	0	0
s_3	0	0	0

$\lambda_2(s_1)$

	t	m	b
t_1	0.5	0.5	0
t_2	0	0	0

- No further iteration of mutual belief in rationality eliminate some types or strategies.

		2		
		l	c	r
1	t	3, 3	1, 1	0, 0
	m	1,1	3, 3	1, 0
	b	0, 4	0, 0	4, 0

$\lambda_1(t_1)$		l	c	r
	s_1	0.5	0.5	0
	s_2	0	0	0
	s_3	0	0	0

$\lambda_2(s_1)$		t	m	b
	t_1	0.5	0.5	0
	t_2	0	0	0

- No further iteration of mutual belief in rationality eliminate some types or strategies.
- So at all the states in $\{(t_1, s_1)\} \times \{t, m\} \times \{l, c\}$ we have rationality and common belief in rationality.

2

1		l	c	r
	t	3, 3	1, 1	0, 0
	m	1,1	3, 3	1, 0
	b	0, 4	0, 0	4, 0

$\lambda_1(t_1)$

	l	c	r
s_1	0.5	0.5	0
s_2	0	0	0
s_3	0	0	0

$\lambda_2(s_1)$

	t	m	b
t_1	0.5	0.5	0
t_2	0	0	0

- No further iteration of mutual belief in rationality eliminate some types or strategies.
- So at all the states in $\{(t_1, s_1)\} \times \{t, m\} \times \{l, c\}$ we have rationality and common belief in rationality.
- But observe that $\{t, m\} \times \{l, c\}$ is precisely the set of profiles that survive IESDS.

The general result: RCBR $\Rightarrow$ IESDS

Suppose that G is a strategic game and $\mathcal{T}$ is any type space for G. If (s, t) is a state in $\mathcal{T}$ in which all the players are rational and there is common belief of rationality, then s is a strategy profile that survives iteratively removal of strictly dominated strategies.

D. Bernheim. *Rationalizable strategic behavior*. Econometrica, 52:1007-1028, 1984.

D. Pearce. *Rationalizable strategic behavior and the problem of perfection*. Econometrica, 52:1029-1050, 1984.

A. Brandenburger and E. Dekel. *Rationalizability and correlated equilibria*. Econometrica, 55:1391-1402, 1987.

Proof: RCBR $\Rightarrow$ IESDS

- We show by induction on n that the if the players have n-level of mutual belief in rationality then they do not play strategies that would be eliminated at the $n + 1^{th}$ round of IESDS.

Proof: RCBR $\Rightarrow$ IESDS

- We show by induction on n that the if the players have n-level of mutual belief in rationality then they do not play strategies that would be eliminated at the $n+1^{th}$ round of IESDS.
- Basic case, $n = 0$. All the players are rational. We know that a strictly dominated strategy, i.e. one that would be eliminated in the 1st round of IESDS, is never a best response. So no player is playing such a strategy.

Proof: RCBR $\Rightarrow$ IESDS

- We show by induction on n that the if the players have n-level of mutual belief in rationality then they do not play strategies that would be eliminated at the $n+1^{th}$ round of IESDS.
- Basic case, $n = 0$. All the players are rational. We know that a strictly dominated strategy, i.e. one that would be eliminated in the 1st round of IESDS, is never a best response. So no player is playing such a strategy.
- Inductive step. Suppose that it is mutual belief up to degree n^{th} that all players are rational.

Proof: RCBR $\Rightarrow$ IESDS

- We show by induction on n that the if the players have n-level of mutual belief in rationality then they do not play strategies that would be eliminated at the $n+1^{th}$ round of IESDS.
- Basic case, $n = 0$. All the players are rational. We know that a strictly dominated strategy, i.e. one that would be eliminated in the 1st round of IESDS, is never a best response. So no player is playing such a strategy.
- Inductive step. Suppose that it is mutual belief up to degree n^{th} that all players are rational. Take any strategy s_i of an agent i that would not survive $n+1$ round of IESDS. This strategy is never a best response to a belief whose support is included in the set of states where the others play strategies that would not survive n^{th} round of IESDS. But by our IH this is precisely the kind of belief that all i's type have by IH, so i is not playing s_i either.

“Converse direction” From IESDS to RCBR

Given any strategy profile that survives IESDS, there is a model in and a state in that model where this profile RCBR holds at that state.

"Converse direction" From IESDS to RCBR

Given any strategy profile that survives IESDS, there is a model in and a state in that model where this profile RCBR holds at that state.

- Trivial? Mathematically, yes.

"Converse direction" From IESDS to RCBR

Given any strategy profile that survives IESDS, there is a model in and a state in that model where this profile RCBR holds at that state.

- Trivial? Mathematically, yes.
- ... but conceptually important. One can always *view* or *interpret* the choice of a strategy profile that would survive the iterative elimination procedure as one that results from RCBR.

"Converse direction" From IESDS to RCBR

Given any strategy profile that survives IESDS, there is a model in and a state in that model where this profile RCBR holds at that state.

- Trivial? Mathematically, yes.
- ... but conceptually important. One can always *view* or *interpret* the choice of a strategy profile that would survive the iterative elimination procedure as one that results from RCBR.

Is the *entire* set of strategy profiles that survive IESDS always consistent with rationality and common belief in rationality? Yes.

- For any game G, there is a type structure for that game in which the strategy profiles consistent with rationality and common belief in rationality is the set of strategies that survive iterative removal of strictly dominated strategies.

A. Friedenberg and J. Kiesler. *Iterated Dominance Revisited*. Working paper, 2011.

Subgames

Let $H = \langle H_1, \ldots, H_n, u_1, \ldots, u_n \rangle$ be an *arbitrary* strategic game.

Subgames

Let $H = \langle H_1, \ldots, H_n, u_1, \ldots, u_n \rangle$ be an *arbitrary* strategic game.

A **restriction** of H is a sequence $G = (G_1, \ldots, G_n)$ such that $G_i \subseteq H_i$ for all $i \in \{1, \ldots, n\}$.

The set of all restrictions of a game H ordered by componentwise set inclusion forms a complete lattice.

Game Models

Relational models: $\langle W, R_i \rangle$ where $R_i \subseteq W \times W$. Write $R_i(w) = \{v \mid wR_i v\}$.

Events: $E \subseteq W$

Knowledge/Belief: $\Box E = \{w \mid R_i(w) \subseteq E\}$

Common knowledge/belief:
$\Box^1 E = \Box E$
$\Box^{k+1} E = \Box\Box^k E$
$\Box^* E = \bigcap_{k=1}^{\infty} \Box^k E$

Fact. An event F is called **evident** provided $F \subseteq \Box F$. $w \in \Box^* E$ provided there is an evident event F such that $w \in F \subseteq \Box E$.

Game Models

Let $G = (G_1, \ldots, G_n)$ be a restriction of a game H.

A **knowledge/belief model of** G is a tuple $\langle W, R_1, \ldots, R_n, \sigma_1, \ldots, \sigma_n \rangle$ where $\langle W, R_1, \ldots, R_n \rangle$ is a knowledge/belief model and $\sigma_i : W \to G_i$.

Game Models

Let $G = (G_1, \ldots, G_n)$ be a restriction of a game H.

A **knowledge/belief model of** G is a tuple $\langle W, R_1, \ldots, R_n, \sigma_1, \ldots, \sigma_n \rangle$ where $\langle W, R_1, \ldots, R_n \rangle$ is a knowledge/belief model and $\sigma_i : W \to G_i$.

Given a model $\langle W, R_1, \ldots, R_n, \sigma_1, \ldots \sigma_n \rangle$ for a restriction G and a sequence $\overline{E} = \{E_1, \ldots, E_n\}$ where $E_i \subseteq W$,

$$G_{\overline{E}} = (\sigma_1(E_1), \ldots, \sigma_n(E_n))$$

Some Lattice Theory

- $(D, \subseteq)$ is a lattice with largest element $\top$. $T : D \to D$ an operator.

Some Lattice Theory

- $(D, \subseteq)$ is a lattice with largest element $\top$. $T : D \rightarrow D$ an operator.
- T is monotonic if for all G, G', $G \subseteq G'$ implies $T(G) \subseteq T(G')$

Some Lattice Theory

- $(D, \subseteq)$ is a lattice with largest element $\top$. $T : D \to D$ an operator.
- T is monotonic if for all G, G', $G \subseteq G'$ implies $T(G) \subseteq T(G')$
- G is a fixed-point if $T(G) = G$

Some Lattice Theory

- $(D, \subseteq)$ is a lattice with largest element $\top$. $T : D \rightarrow D$ an operator.
- T is monotonic if for all G, G', $G \subseteq G'$ implies $T(G) \subseteq T(G')$
- G is a fixed-point if $T(G) = G$
- νT is the largest fixed point of T

Some Lattice Theory

- $(D, \subseteq)$ is a lattice with largest element $\top$. $T : D \to D$ an operator.
- T is monotonic if for all G, G', $G \subseteq G'$ implies $T(G) \subseteq T(G')$
- G is a fixed-point if $T(G) = G$
- νT is the largest fixed point of T
- T^{∞} is the "outcome of T: $T^0 = \top$, $T^{\alpha+1} = T(T^{\alpha})$, $T^{\beta} = \bigcap_{\alpha<\beta} T^{\alpha}$, The outcome of iterating T is the least α such that $T^{\alpha+1} = T^{\alpha}$, denoted T^{∞}

Some Lattice Theory

- $(D, \subseteq)$ is a lattice with largest element $\top$. $T : D \to D$ an operator.
- T is monotonic if for all G, G', $G \subseteq G'$ implies $T(G) \subseteq T(G')$
- G is a fixed-point if $T(G) = G$
- νT is the largest fixed point of T
- T^∞ is the "outcome of T: $T^0 = \top$, $T^{\alpha+1} = T(T^\alpha)$, $T^\beta = \bigcap_{\alpha<\beta} T^\alpha$, The outcome of iterating T is the least α such that $T^{\alpha+1} = T^\alpha$, denoted T^∞
- **Tarski's Fixed-Point Theorem**: Every monotonic operator T has a (least and largest) fixed point $T^\infty = \nu T = \bigcup\{G \mid G \subseteq T(G)\}$.

Some Lattice Theory

- $(D, \subseteq)$ is a lattice with largest element $\top$. $T : D \to D$ an operator.
- T is monotonic if for all G, G', $G \subseteq G'$ implies $T(G) \subseteq T(G')$
- G is a fixed-point if $T(G) = G$
- νT is the largest fixed point of T
- T^∞ is the "outcome of T: $T^0 = \top$, $T^{\alpha+1} = T(T^\alpha)$, $T^\beta = \bigcap_{\alpha<\beta} T^\alpha$, The outcome of iterating T is the least α such that $T^{\alpha+1} = T^\alpha$, denoted T^∞
- **Tarski's Fixed-Point Theorem**: Every monotonic operator T has a (least and largest) fixed point $T^\infty = \nu T = \bigcup\{G \mid G \subseteq T(G)\}$.
- T is contracting if $T(G) \subseteq G$. Every contracting operator has an outcome (T^∞ is well-defined)

Rationality Properties

$\varphi(s_i, G_i, G_{-i})$ holds between a strategy $s_i \in H_i$, a set of strategies G_i for player i and strategies G_{-i} of the opponents. Intuitively s_i is φ-optimal strategy for player i in the restricted game $\langle G_i, G_{-i}, u_1, \ldots, u_n \rangle$ (where the payoffs are suitably restricted).

Rationality Properties

$\varphi(s_i, G_i, G_{-i})$ holds between a strategy $s_i \in H_i$, a set of strategies G_i for player i and strategies G_{-i} of the opponents. Intuitively s_i is φ-optimal strategy for player i in the restricted game $\langle G_i, G_{-i}, u_1, \ldots, u_n \rangle$ (where the payoffs are suitably restricted).

φ_i is **monotonic** if for all G_{-i}, $G'_{-i} \subseteq H_{-i}$ and $s_i \in H_i$

$$G_{-i} \subseteq G'_{-i} \text{ and } \varphi(s_i, H_i, G_{-i}) \text{ implies } \varphi(s_i, H_i, G'_{-i})$$

Removing Strategies

If $\varphi = (\varphi_1, \ldots, \varphi_n)$, then define $T_\varphi(G) = G'$ where

- $G = (G_1, \ldots, G_n)$, $G' = (G'_1, \ldots, G'_n)$,
- for all $i \in \{1, \ldots, n\}$, $\quad G'_i = \{s_i \in G_i \mid \varphi_i(s_i, H_i, G_{-i})\}$

Removing Strategies

If $\varphi = (\varphi_1, \ldots, \varphi_n)$, then define $T_\varphi(G) = G'$ where

- $G = (G_1, \ldots, G_n)$, $G' = (G'_1, \ldots, G'_n)$,
- for all $i \in \{1, \ldots, n\}$, $G'_i = \{s_i \in G_i \mid \varphi_i(s_i, H_i, G_{-i})\}$

T_φ is contracting, so it has an outcome T_φ^∞

Removing Strategies

If $\varphi = (\varphi_1, \ldots, \varphi_n)$, then define $T_\varphi(G) = G'$ where

- $G = (G_1, \ldots, G_n)$, $G' = (G'_1, \ldots, G'_n)$,
- for all $i \in \{1, \ldots, n\}$, $\quad G'_i = \{s_i \in G_i \mid \varphi_i(s_i, H_i, G_{-i})\}$

T_φ is contracting, so it has an outcome T_φ^∞

If each φ_i is monotonic, then νT_φ exists and equals T_φ^∞.

Rational Play

Let $H = \langle H_1, \ldots, H_n, u_1, \ldots, u_n \rangle$ a strategic game and $\langle W, R_1, \ldots, R_n, \sigma_1, \ldots, \sigma_n \rangle$ a model for H.

$\sigma_i(w)$ is the strategy player is using in state w.

$G_{R_i(w)}$ is a restriction of H giving i's view of the game.

Rational Play

Let $H = \langle H_1, \ldots, H_n, u_1, \ldots, u_n \rangle$ a strategic game and $\langle W, R_1, \ldots, R_n, \sigma_1, \ldots, \sigma_n \rangle$ a model for H.

$\sigma_i(w)$ is the strategy player is using in state w.

$G_{R_i(w)}$ is a restriction of H giving i's view of the game.

Player i is φ_i-rational in the state w if $\varphi_i(\sigma_i(w), H_i, (G_{R_i(w)})_{-i})$ holds.

Rational Play

Let $H = \langle H_1, \ldots, H_n, u_1, \ldots, u_n \rangle$ a strategic game and $\langle W, R_1, \ldots, R_n, \sigma_1, \ldots, \sigma_n \rangle$ a model for H.

$\sigma_i(w)$ is the strategy player is using in state w.

$G_{R_i(w)}$ is a restriction of H giving i's view of the game.

Player i is φ_i-rational in the state w if $\varphi_i(\sigma_i(w), H_i, (G_{R_i(w)})_{-i})$ holds.

$\mathbf{Rat}(\varphi) = \{w \in W \mid \text{ each player is } \varphi_i\text{-rational in } w\}$

$\Box\mathbf{Rat}(\varphi)$
$\Box^*\mathbf{Rat}(\varphi)$

Theorem (Apt and Zvesper).

- Suppose that each φ_i is monotonic. Then for all belief models for H,
$$G_{\mathbf{Rat}(\varphi)\cap B^*(\mathbf{Rat}(\varphi))} \subseteq T^\infty_\varphi$$
- Suppose that each φ_i is monotonic. Then for all knowledge models for H,
$$G_{K^*(\mathbf{Rat}(\varphi))} \subseteq T^\infty_\varphi$$
- For some standard knowledge model for H,
$$T^\infty_\varphi \subseteq G_{K^*(\mathbf{Rat}(\varphi))}$$

K. Apt and J. Zvesper. *The Role of Monotonicity in the Epistemic Analysis of Games*. Games, 1(4), pgs. 381-394, 2010.

Claim If each φ_i is monotonic, then $G_{\mathbf{Rat}(\varphi)\cap\Box^*\mathbf{Rat}(\varphi)} \subseteq T_{\varphi}^{\infty}$.

Claim If each φ_i is monotonic, then $G_{\mathbf{Rat}(\varphi)\cap\Box^*\mathbf{Rat}(\varphi)} \subseteq T^{\infty}_{\varphi}$.

Let s_i be an element of the ith component of $G_{\mathbf{Rat}(\varphi)\cap\Box^*\mathbf{Rat}(\varphi)}$:
$s_i = \sigma_i(w)$ for some $w \in \mathbf{Rat}(\varphi) \cap \Box^*\mathbf{Rat}(\varphi)$

Claim If each φ_i is monotonic, then $G_{\mathbf{Rat}(\varphi)\cap\Box^*\mathbf{Rat}(\varphi)} \subseteq T^\infty_\varphi$.

Let s_i be an element of the ith component of $G_{\mathbf{Rat}(\varphi)\cap\Box^*\mathbf{Rat}(\varphi)}$: $s_i = \sigma_i(w)$ for some $w \in \mathbf{Rat}(\varphi) \cap \Box^*\mathbf{Rat}(\varphi)$

there is an F such that $F \subseteq \Box F$ and

$$w \in F \subseteq \Box\mathbf{Rat}(\varphi) = \{v \in W \mid \forall i\ R_i(v) \subseteq \mathbf{Rat}(\varphi)\}$$

Claim If each φ_i is monotonic, then $G_{\mathbf{Rat}(\varphi)\cap\Box^*\mathbf{Rat}(\varphi)} \subseteq T_\varphi^\infty$.

Let s_i be an element of the ith component of $G_{\mathbf{Rat}(\varphi)\cap\Box^*\mathbf{Rat}(\varphi)}$: $s_i = \sigma_i(w)$ for some $w \in \mathbf{Rat}(\varphi) \cap \Box^*\mathbf{Rat}(\varphi)$

there is an F such that $F \subseteq \Box F$ and

$$w \in F \subseteq \Box\mathbf{Rat}(\varphi) = \{v \in W \mid \forall i\ R_i(v) \subseteq \mathbf{Rat}(\varphi)\}$$

Claim. $G_{F\cap\mathbf{Rat}(\varphi)}$ is post-fixed point of T_φ ($G_{F\cap\mathbf{Rat}(\varphi)} \subseteq T_\varphi(G_{F\cap\mathbf{Rat}(\varphi)})$).

Claim If each φ_i is monotonic, then $G_{\mathbf{Rat}(\varphi)\cap\Box^*\mathbf{Rat}(\varphi)} \subseteq T_\varphi^\infty$.

Let s_i be an element of the ith component of $G_{\mathbf{Rat}(\varphi)\cap\Box^*\mathbf{Rat}(\varphi)}$:
$s_i = \sigma_i(w)$ for some $w \in \mathbf{Rat}(\varphi) \cap \Box^*\mathbf{Rat}(\varphi)$

there is an F such that $F \subseteq \Box F$ and

$$w \in F \subseteq \Box\mathbf{Rat}(\varphi) = \{v \in W \mid \forall i \ R_i(v) \subseteq \mathbf{Rat}(\varphi)\}$$

Claim. $G_{F\cap\mathbf{Rat}(\varphi)}$ is post-fixed point of T_φ
$(G_{F\cap\mathbf{Rat}(\varphi)} \subseteq T_\varphi(G_{F\cap\mathbf{Rat}(\varphi)}))$.

Since each φ_i is monotonic, T_φ is monotonic and by Tarski's fixed-point theorem, $G_{F\cap\mathbf{Rat}(\varphi)} \subseteq T_\varphi^\infty$. But $s_i = \sigma_i(w)$ and $w \in F \cap \mathbf{Rat}(\varphi)$, so s_i is the ith component in T_φ^∞.

$F \subseteq \Box F$ and $w \in F \subseteq \Box \mathbf{Rat}(\varphi) = \{v \in W \mid \forall i\ R_i(v) \subseteq \mathbf{Rat}(\varphi)\}$

Claim. $G_{F \cap \mathbf{Rat}(\varphi)}$ is post-fixed point of T_φ ($G_{F \cap \mathbf{Rat}(\varphi)} \subseteq T_\varphi(G_{F \cap \mathbf{Rat}(\varphi)})$).

$F \subseteq \Box F$ and $w \in F \subseteq \Box \mathbf{Rat}(\varphi) = \{v \in W \mid \forall i \; R_i(v) \subseteq \mathbf{Rat}(\varphi)\}$

Claim. $G_{F \cap \mathbf{Rat}(\varphi)}$ is post-fixed point of T_φ ($G_{F \cap \mathbf{Rat}(\varphi)} \subseteq T_\varphi(G_{F \cap \mathbf{Rat}(\varphi)})$).

Let $w' \in F \cap \mathbf{Rat}(\varphi)$ and let $i \in \{1, \ldots, n\}$.

$F \subseteq \Box F$ and $w \in F \subseteq \Box \mathbf{Rat}(\varphi) = \{v \in W \mid \forall i\ R_i(v) \subseteq \mathbf{Rat}(\varphi)\}$

Claim. $G_{F\cap\mathbf{Rat}(\varphi)}$ is post-fixed point of T_φ ($G_{F\cap\mathbf{Rat}(\varphi)} \subseteq T_\varphi(G_{F\cap\mathbf{Rat}(\varphi)})$).

Let $w' \in F \cap \mathbf{Rat}(\varphi)$ and let $i \in \{1, \ldots, n\}$.

Since $w' \in \mathbf{Rat}(\varphi)$, $\varphi_i(\sigma_i(w'), H_i, (G_{R_i(w)})_{-i})$ holds.

$F \subseteq \Box F$ and $w \in F \subseteq \Box \mathbf{Rat}(\varphi) = \{v \in W \mid \forall i\ R_i(v) \subseteq \mathbf{Rat}(\varphi)\}$

Claim. $G_{F \cap \mathbf{Rat}(\varphi)}$ is post-fixed point of T_φ $(G_{F \cap \mathbf{Rat}(\varphi)} \subseteq T_\varphi(G_{F \cap \mathbf{Rat}(\varphi)}))$.

Let $w' \in F \cap \mathbf{Rat}(\varphi)$ and let $i \in \{1, \ldots, n\}$.

Since $w' \in \mathbf{Rat}(\varphi)$, $\varphi_i(\sigma_i(w'), H_i, (G_{R_i(w)})_{-i})$ holds.

F is evident, so $R_i(w') \subseteq F$. We also have $R_i(w') \subseteq \mathbf{Rat}(\varphi)$.

$F \subseteq \Box F$ and $w \in F \subseteq \Box\mathbf{Rat}(\varphi) = \{v \in W \mid \forall i\ R_i(v) \subseteq \mathbf{Rat}(\varphi)\}$

Claim. $G_{F\cap\mathbf{Rat}(\varphi)}$ is post-fixed point of T_φ ($G_{F\cap\mathbf{Rat}(\varphi)} \subseteq T_\varphi(G_{F\cap\mathbf{Rat}(\varphi)})$).

Let $w' \in F \cap \mathbf{Rat}(\varphi)$ and let $i \in \{1, \dots, n\}$.

Since $w' \in \mathbf{Rat}(\varphi)$, $\varphi_i(\sigma_i(w'), H_i, (G_{R_i(w)})_{-i})$ holds.

F is evident, so $R_i(w') \subseteq F$. We also have $R_i(w') \subseteq \mathbf{Rat}(\varphi)$.

Hence, $R_i(w') \subseteq F \cap \mathbf{Rat}(\varphi)$.

$F \subseteq \Box F$ and $w \in F \subseteq \Box\mathbf{Rat}(\varphi) = \{v \in W \mid \forall i\ R_i(v) \subseteq \mathbf{Rat}(\varphi)\}$

Claim. $G_{F\cap\mathbf{Rat}(\varphi)}$ is post-fixed point of T_φ $(G_{F\cap\mathbf{Rat}(\varphi)} \subseteq T_\varphi(G_{F\cap\mathbf{Rat}(\varphi)}))$.

Let $w' \in F \cap \mathbf{Rat}(\varphi)$ and let $i \in \{1, \ldots, n\}$.

Since $w' \in \mathbf{Rat}(\varphi)$, $\varphi_i(\sigma_i(w'), H_i, (G_{R_i(w)})_{-i})$ holds.

F is evident, so $R_i(w') \subseteq F$. We also have $R_i(w') \subseteq \mathbf{Rat}(\varphi)$.

Hence, $R_i(w') \subseteq F \cap \mathbf{Rat}(\varphi)$.

This implies $(G_{R_i(w')}) \subseteq (G_{F\cap\mathbf{Rat}(\varphi)})_{-i}$, and so by monotonicity of φ_i, $\varphi_i(s_i, H_i, (G_{F\cap\mathbf{Rat}(\varphi)})_{-i})$ holds.

$F \subseteq \Box F$ and $w \in F \subseteq \Box \mathbf{Rat}(\varphi) = \{v \in W \mid \forall i\ R_i(v) \subseteq \mathbf{Rat}(\varphi)\}$

Claim. $G_{F \cap \mathbf{Rat}(\varphi)}$ is post-fixed point of T_φ ($G_{F \cap \mathbf{Rat}(\varphi)} \subseteq T_\varphi(G_{F \cap \mathbf{Rat}(\varphi)})$).

Let $w' \in F \cap \mathbf{Rat}(\varphi)$ and let $i \in \{1, \ldots, n\}$.

Since $w' \in \mathbf{Rat}(\varphi)$, $\varphi_i(\sigma_i(w'), H_i, (G_{R_i(w)})_{-i})$ holds.

F is evident, so $R_i(w') \subseteq F$. We also have $R_i(w') \subseteq \mathbf{Rat}(\varphi)$.

Hence, $R_i(w') \subseteq F \cap \mathbf{Rat}(\varphi)$.

This implies $(G_{R_i(w')}) \subseteq (G_{F \cap \mathbf{Rat}(\varphi)})_{-i}$, and so by monotonicity of φ_i, $\varphi_i(s_i, H_i, (G_{F \cap \mathbf{Rat}(\varphi)})_{-i})$ holds.

This means $G_{F \cap \mathbf{Rat}(\varphi)} \subseteq T_\varphi(G_{F \cap \mathbf{Rat}(\varphi)})$

$sd_i(s_i, G_i, G_{-i})$ is $\neg\exists s_i' \in G_i, \forall s_{-i} \in G_{-i} u_i(s_i', s_{-i}) > u_i(s_i, s_{-i})$

$sd_i(s_i, G_i, G_{-i})$ is $\neg\exists s_i' \in G_i, \forall s_{-i} \in G_{-i} u_i(s_i', s_{-i}) > u_i(s_i, s_{-i})$

$br_i(s_i, G_i, G_{-i})$ is $\exists \mu_i \in \mathcal{B}_i(G_{-i}) \forall s_i' \in G_i, U_i(s_i, \mu_i) \geq U_i(s_i', \mu_i)$.

$sd_i(s_i, G_i, G_{-i})$ is $\neg\exists s_i' \in G_i, \forall s_{-i} \in G_{-i} u_i(s_i', s_{-i}) > u_i(s_i, s_{-i})$

$br_i(s_i, G_i, G_{-i})$ is $\exists\mu_i \in \mathcal{B}_i(G_{-i})\forall s_i' \in G_i, U_i(s_i, \mu_i) \geq U_i(s_i', \mu_i)$.

$U_\varphi(G) = G'$ where $G_i' = \{s_i \in G_i \mid \varphi_i(s_i, G_i, G_{-i})\}$.

$sd_i(s_i, G_i, G_{-i})$ is $\neg\exists s_i' \in G_i, \forall s_{-i} \in G_{-i} u_i(s_i', s_{-i}) > u_i(s_i, s_{-i})$

$br_i(s_i, G_i, G_{-i})$ is $\exists \mu_i \in \mathcal{B}_i(G_{-i}) \forall s_i' \in G_i, U_i(s_i, \mu_i) \geq U_i(s_i', \mu_i)$.

$U_\varphi(G) = G'$ where $G_i' = \{s_i \in G_i \mid \varphi_i(s_i, G_i, G_{-i})\}$.

Note: U_φ is *not* monotonic.

Corollary. For all belief models, $G_{\mathbf{Rat}(br)\cap\Box^*\mathbf{Rat}(br)} \subseteq U_{sd}^{\infty}$. For all G, we have

$$T_{br}(G) \subseteq T_{sd}(G)$$

$$T_{sd}(G) \subseteq U_{sd}(G)$$

Then, $T_{sd}^{\infty} \subseteq U_{sd}^{\infty}$.

Corollary. For all belief models, $G_{\mathbf{Rat}(br)\cap\Box^*\mathbf{Rat}(br)} \subseteq U_{sd}^{\infty}$. For all G, we have

$$T_{br}(G) \subseteq T_{sd}(G)$$

$$T_{sd}(G) \subseteq U_{sd}(G)$$

Then, $T_{sd}^{\infty} \subseteq U_{sd}^{\infty}$.

Fact. Consider two operators T_1, T_2 on $(D, \subseteq)$ such that,

- for all G, $T_1(G) \subseteq T_2(G)$
- T_1 is monotonic
- T_2 is contracting

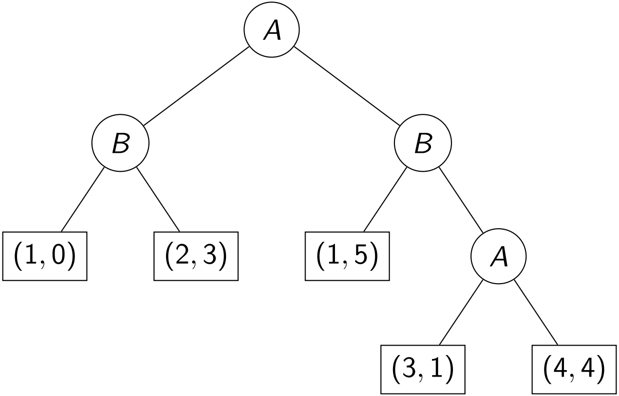
Then, $T_1^{\infty} \subseteq T_2^{\infty}$.

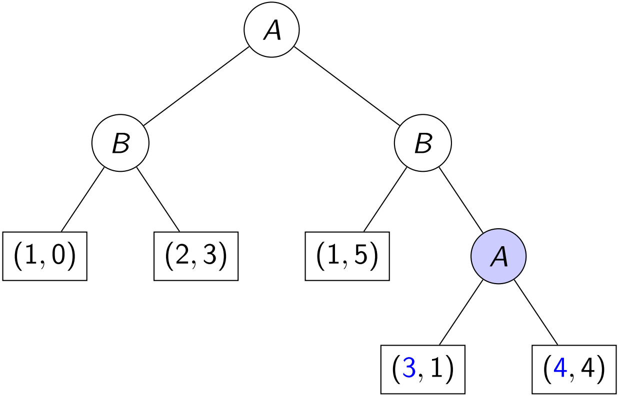
This analysis does not work for weak dominance...

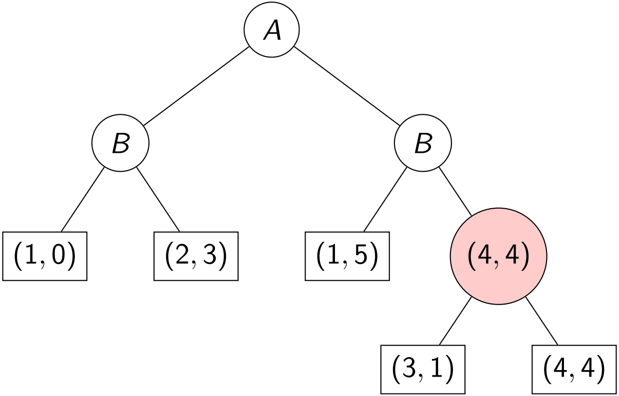
Common knowledge of rationality (CKR) in the tree.

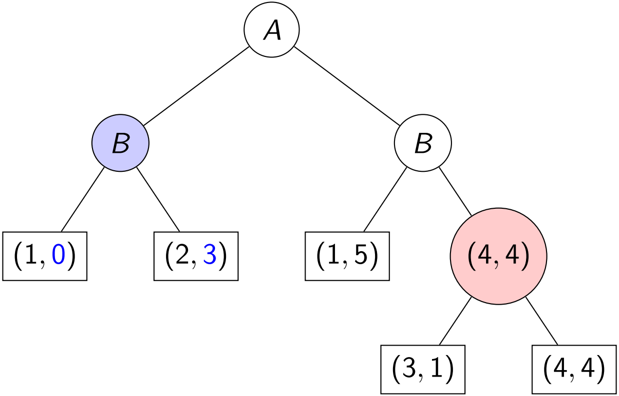
Backwards Induction

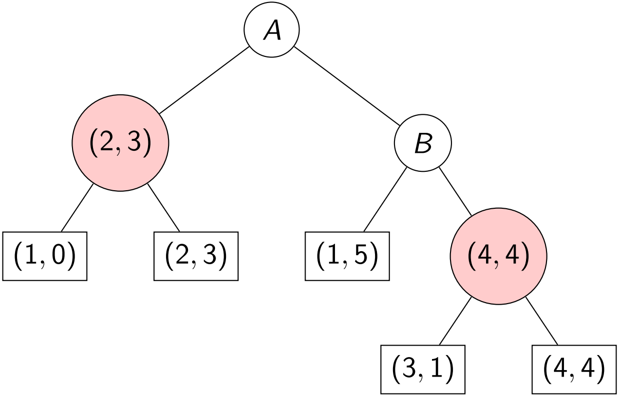
Invented by Zermelo, Backwards Induction is an iterative algorithm for "solving" and extensive game.

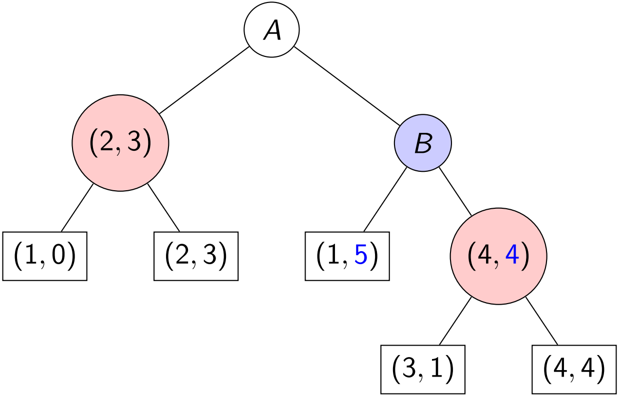

A
B
B
(1, 0)
(2, 3)
(1, 5)
A
(3, 1)
(4, 4)

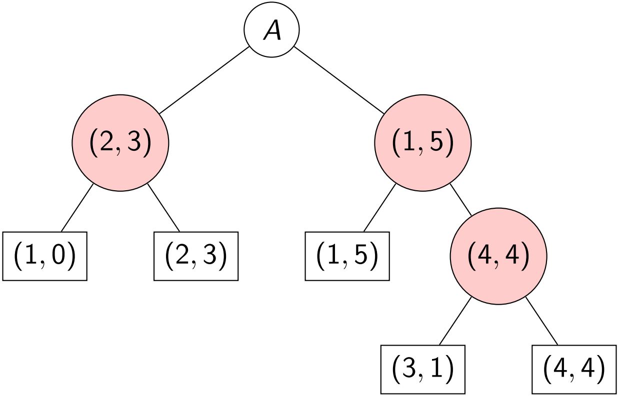

A
B
B
(1, 0)
(2, 3)
(1, 5)
A
(3, 1)
(4, 4)

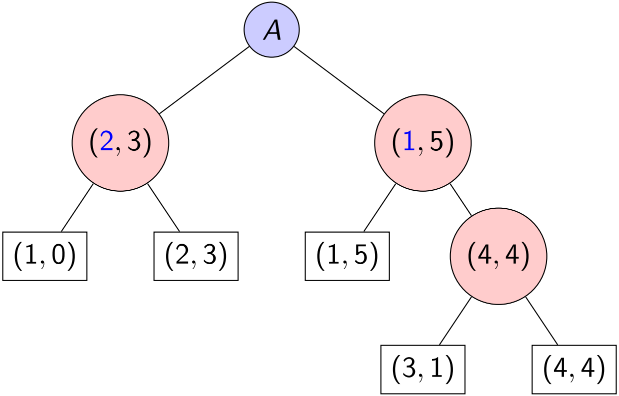

A
B
B
(1, 0)
(2, 3)
(1, 5)
(4, 4)
(3, 1)
(4, 4)

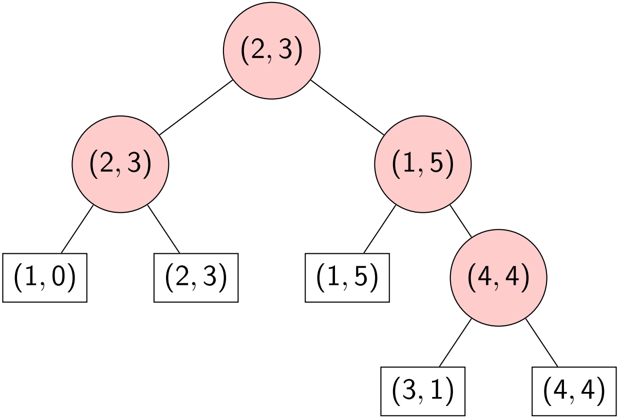

A
B
B
(1, 0)
(2, 3)
(1, 5)
(4, 4)
(3, 1)
(4, 4)


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

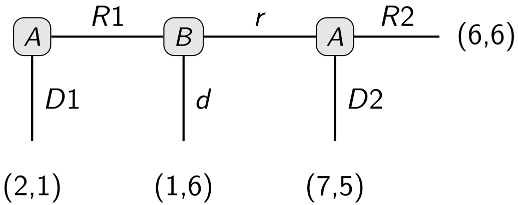

A
(2, 3)
B
(1, 0)
(2, 3)
(1, 5)
(4, 4)
(3, 1)
(4, 4)


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

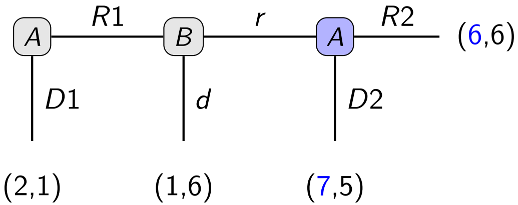

A
(2, 3)
(1, 5)
(1, 0)
(2, 3)
(1, 5)
(4, 4)
(3, 1)
(4, 4)


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

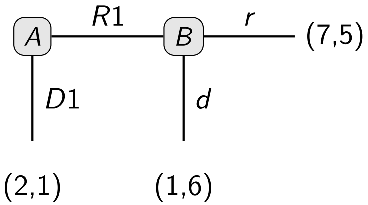
BI Puzzle



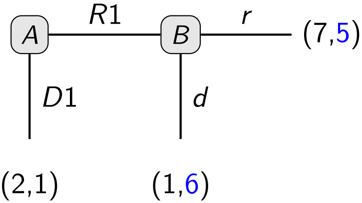
BI Puzzle



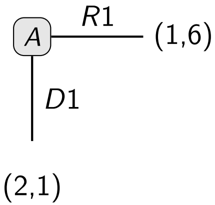
BI Puzzle



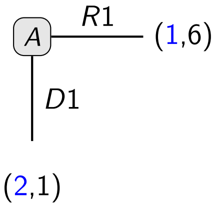
BI Puzzle



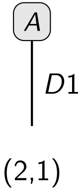
BI Puzzle



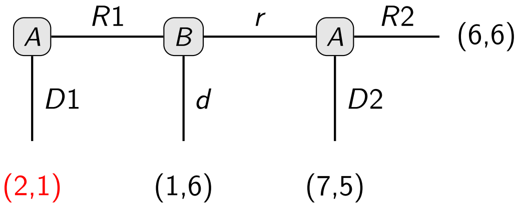
BI Puzzle



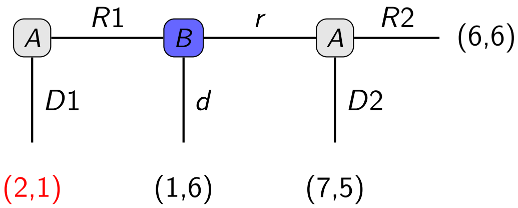
BI Puzzle



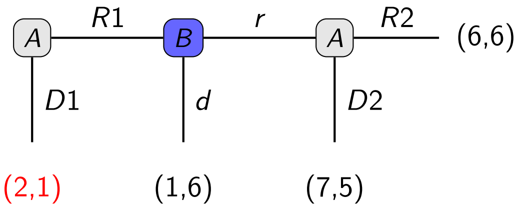
BI Puzzle



But what if Bob has to move?



But what if Bob has to move?



What should Bob thinks of Ann?

- Either she doesn't believe that he is rational and that he believes that she would choose $R2$.
- Or Ann made a "mistake" (= irrational move) at the first turn.

Either way, rationality is not "common knowledge".

R. Aumann. *Backwards induction and common knowledge of rationality*. Games and Economic Behavior, 8, pgs. 6 - 19, 1995.

R. Stalnaker. *Knowledge, belief and counterfactual reasoning in games*. Economics and Philosophy, 12, pgs. 133 - 163, 1996.

J. Halpern. *Substantive Rationality and Backward Induction*. Games and Economic Behavior, 37, pp. 425-435, 1998.

Models of Extensive Games

Let Γ be a *non-degenerate* extensive game with perfect information. Let Γ_i be the set of nodes controlled by player i.

Models of Extensive Games

Let Γ be a *non-degenerate* extensive game with perfect information. Let Γ_i be the set of nodes controlled by player i.

A strategy profile σ describes the choice for each player i at all vertices where i can choose.

Models of Extensive Games

Let Γ be a *non-degenerate* extensive game with perfect information. Let Γ_i be the set of nodes controlled by player i.

A strategy profile σ describes the choice for each player i at all vertices where i can choose.

Given a vertex v in Γ and strategy profile σ, σ specifies a unique path from v to an end-node.

Models of Extensive Games

Let Γ be a *non-degenerate* extensive game with perfect information. Let Γ_i be the set of nodes controlled by player i.

A strategy profile σ describes the choice for each player i at all vertices where i can choose.

Given a vertex v in Γ and strategy profile σ, σ specifies a unique path from v to an end-node.

$\mathcal{M}(\Gamma) = \langle W, \sim_i, \sigma \rangle$ where $\sigma : W \to Strat(\Gamma)$ and $\sim_i \subseteq W \times W$ is an equivalence relation.

Models of Extensive Games

Let Γ be a *non-degenerate* extensive game with perfect information. Let Γ_i be the set of nodes controlled by player i.

A strategy profile σ describes the choice for each player i at all vertices where i can choose.

Given a vertex v in Γ and strategy profile σ, σ specifies a unique path from v to an end-node.

$\mathcal{M}(\Gamma) = \langle W, \sim_i, \sigma \rangle$ where $\sigma : W \to \mathit{Strat}(\Gamma)$ and $\sim_i \subseteq W \times W$ is an equivalence relation.

If $\sigma(w) = \sigma$, then $\sigma_i(w) = \sigma_i$ and $\sigma_{-i}(w) = \sigma_{-i}$

Models of Extensive Games

Let Γ be a *non-degenerate* extensive game with perfect information. Let Γ_i be the set of nodes controlled by player i.

A strategy profile σ describes the choice for each player i at all vertices where i can choose.

Given a vertex v in Γ and strategy profile σ, σ specifies a unique path from v to an end-node.

$\mathcal{M}(\Gamma) = \langle W, \sim_i, \sigma \rangle$ where $\sigma : W \rightarrow Strat(\Gamma)$ and $\sim_i \subseteq W \times W$ is an equivalence relation.

If $\sigma(w) = \sigma$, then $\sigma_i(w) = \sigma_i$ and $\sigma_{-i}(w) = \sigma_{-i}$

(A1) If $w \sim_i w'$ then $\sigma_i(w) = \sigma_i(w')$.

Rationality

$h_i^v(\sigma)$ denote "i's payoff if σ is followed from node v"

Rationality

$h_i^v(\sigma)$ denote "i's payoff if σ is followed from node v"

i **is rational at** v **in** w provided for all strategies $s_i \neq \sigma_i(w)$, $h_i^v(\sigma(w')) \geq h_i^v((\sigma_{-i}(w'), s_i))$ for some $w' \in [w]_i$.

Substantive Rationality

i is **substantively rational** in state w if i is rational at a vertex v in w of every vertex in $v \in \Gamma_i$

Stalnaker Rationality

For every vertex $v \in \Gamma_i$, *if i were to actually reach v, then what he would do in that case would be rational*.

Stalnaker Rationality

For every vertex $v \in \Gamma_i$, *if i were to actually reach v, then what he would do in that case would be rational*.

$f : W \times \Gamma_i \to W$, $f(w, v) = w'$, then w' is the "closest state to w where the vertex v is reached.

Stalnaker Rationality

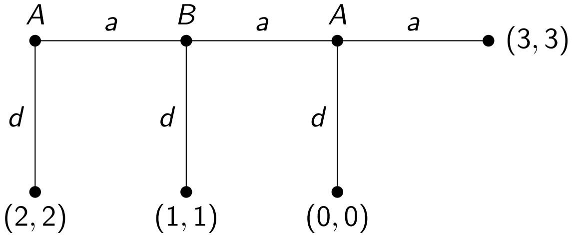
For every vertex $v \in \Gamma_i$, *if i were to actually reach v, then what he would do in that case would be rational*.

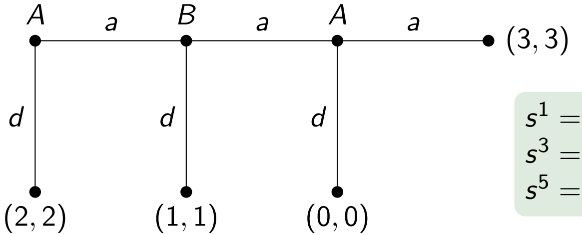
$f : W \times \Gamma_i \to W$, $f(w, v) = w'$, then w' is the "closest state to w where the vertex v is reached.

(F1) v is reached in $f(w, v)$ (i.e., v is on the path determined by $\sigma(f(w, v))$)

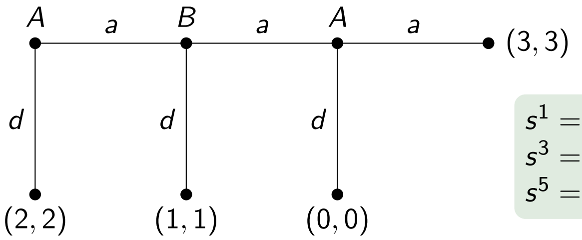
(F2) If v is reached in w, then $f(w, v) = w$

(F3) $\sigma(f(w, v))$ and $\sigma(w)$ agree on the subtree of Γ below v


A
a
B
a
A
a
(3, 3)
d
d
d
(2, 2)
(1, 1)
(0, 0)

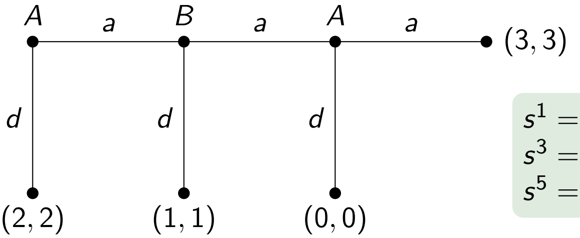


$s^1 = (da, d)$, $s^2 = (aa, d)$,
$s^3 = (ad, d)$, $s^4 = (aa, a)$,
$s^5 = (ad, a)$

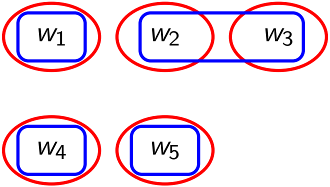


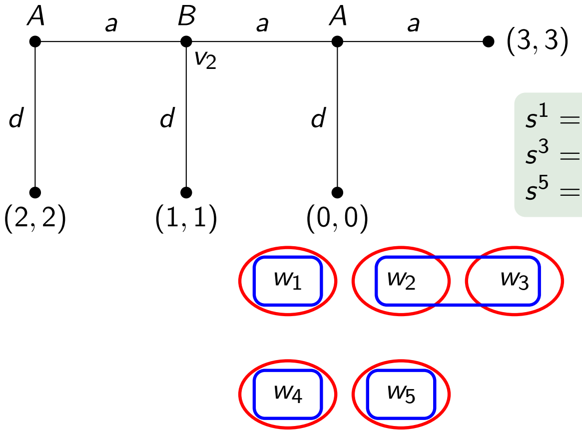
$s^1 = (da, d)$, $s^2 = (aa, d)$,
$s^3 = (ad, d)$, $s^4 = (aa, a)$,
$s^5 = (ad, a)$

- $W = \{w_1, w_2, w_3, w_4, w_5\}$ with $\sigma(w_i) = s^i$
- $[w_i]_A = \{w_i\}$ for $i = 1, 2, 3, 4, 5$
- $[w_i]_B = \{w_i\}$ for $i = 1, 4, 5$ and $[w_2]_B = [w_3]_B = \{w_2, w_3\}$



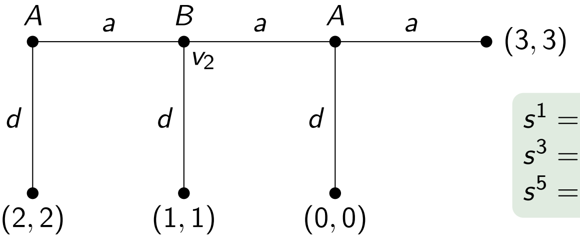
$s^1 = (da, d)$, $s^2 = (aa, d)$,
$s^3 = (ad, d)$, $s^4 = (aa, a)$,
$s^5 = (ad, a)$



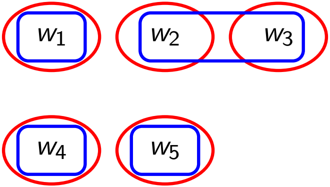


$s^1 = (da, d)$, $s^2 = (aa, d)$,
$s^3 = (ad, d)$, $s^4 = (aa, a)$,
$s^5 = (ad, a)$

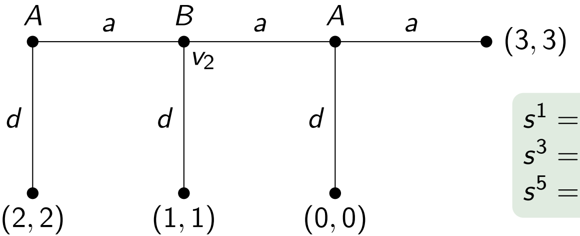
It is **common knowledge** at w_1 that if vertex v_2 were reached, Bob would play down.



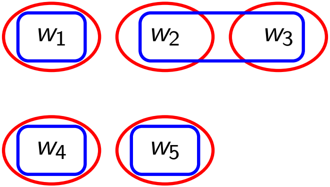
$s^1 = (da, d)$, $s^2 = (aa, d)$,
$s^3 = (ad, d)$, $s^4 = (aa, a)$,
$s^5 = (ad, a)$



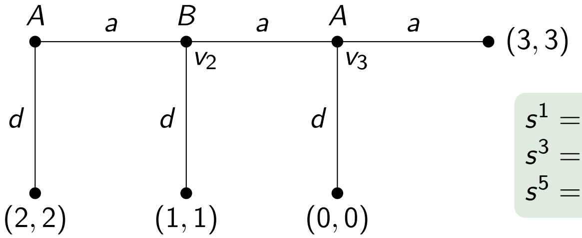
Bob is not rational at v_2 in w_1



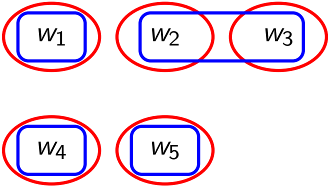
$s^1 = (da, d)$, $s^2 = (aa, d)$,
$s^3 = (ad, d)$, $s^4 = (aa, a)$,
$s^5 = (ad, a)$



Bob is rational at v_2 in w_2



$s^1 = (da, d)$, $s^2 = (aa, d)$,
$s^3 = (ad, d)$, $s^4 = (aa, a)$,
$s^5 = (ad, a)$



Note that $f(w_1, v_2) = w_2$ and $f(w_1, v_3) = w_4$, so there is common knowledge of S-rationality at w_1.

Aumann's Theorem: If Γ is a non-degenerate game of perfect information, then in all models of Γ, we have $C(A-Rat) \subseteq BI$

Stalnaker's Theorem: There exists a non-degenerate game Γ of perfect information and an extended model of Γ in which the selection function satisfies F1-F3 such that $C(S-Rat) \not\subseteq BI$.

Aumann's Theorem: If Γ is a non-degenerate game of perfect information, then in all models of Γ, we have $C(A-Rat) \subseteq BI$

Stalnaker's Theorem: There exists a non-degenerate game Γ of perfect information and an extended model of Γ in which the selection function satisfies F1-F3 such that $C(S-Rat) \not\subseteq BI$.

Revising beliefs during play:

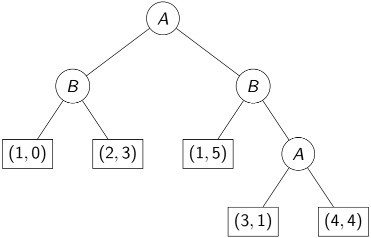
"Although it is common knowledge that Ann would play across if v_3 were reached, if Ann were to play across at v_1, Bob would consider it possible that Ann would play down at v_3"

F4. For all players i and vertices v, if $w' \in [f(w, v)]_i$ then there exists a state $w'' \in [w]_i$ such that $\sigma(w')$ and $\sigma(w'')$ agree on the subtree of Γ below v.

Theorem (Halpern). If Γ is a non-degenerate game of perfect information, then for every extended model of Γ in which the selection function satisfies F1-F4, we have $C(S - Rat) \subseteq BI$. Moreover, there is an extend model of Γ in which the selection function satisfies F1-F4.

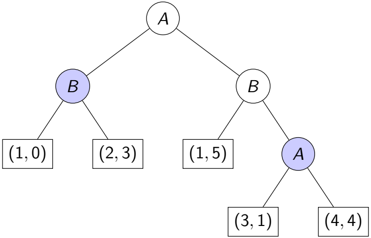
J. Halpern. *Substantive Rationality and Backward Induction*. Games and Economic Behavior, 37, pp. 425-435, 1998.

Proof of Halpern's Theorem



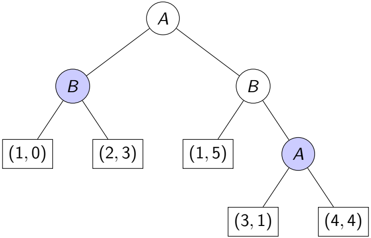
- Suppose $w \in C(S - Rat)$. We show by induction on k that for all w' reachable from w by a finite path along the union of the relations $\sim_i$, if v is at most k moves away from a leaf, then $\sigma_i(w)$ is i's backward induction move at w'.

Proof of Halpern's Theorem



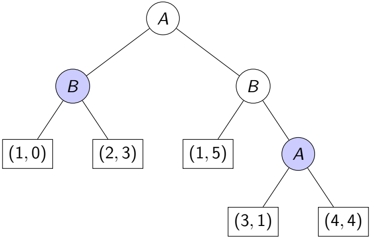
- Base case: we are at most 1 move away from a leaf. Suppose $w \in C(S - Rat)$. Take any w' reachable from w.

Proof of Halpern's Theorem



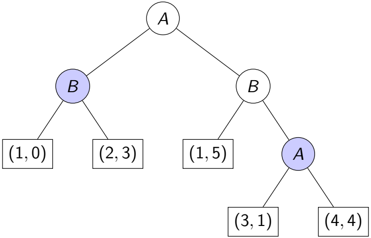
- Base case: we are at most 1 move away from a leaf. Suppose $w \in C(S-Rat)$. Take any w' reachable from w. Since $w \in C(S-Rat)$, we know that $w' \in C(S-Rat)$.

Proof of Halpern's Theorem



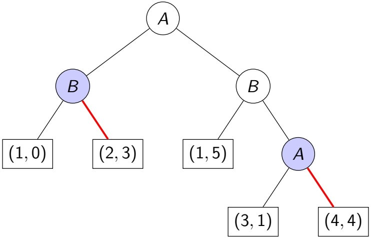
- Base case: we are at most 1 move away from a leaf. Suppose $w \in C(S - Rat)$. Take any w' reachable from w. Since $w \in C(S - Rat)$, we know that $w' \in C(S - Rat)$. So i must play her BI move at $f(w', v)$.

Proof of Halpern's Theorem



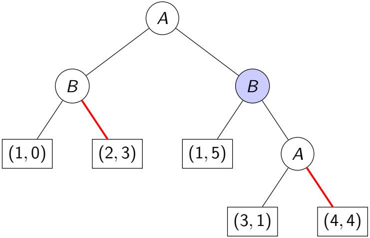
- Base case: we are at most 1 move away from a leaf. Suppose $w \in C(S - Rat)$. Take any w' reachable from w. Since $w \in C(S - Rat)$, we know that $w' \in C(S - Rat)$. So i must play her BI move at $f(w', v)$. But then by F3 this must also be the case at (w', v).

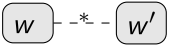
Proof of Halpern's Theorem



- Base case: we are at most 1 move away from a leaf. Suppose $w \in C(S - Rat)$. Take any w' reachable from w. Since $w \in C(S - Rat)$, we know that $w' \in C(S - Rat)$. So i must play her BI move at $f(w', v)$. But then by F3 this must also be the case at (w', v).

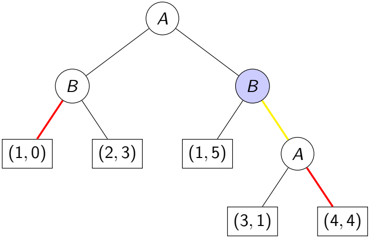
Proof of Halpern's Theorem

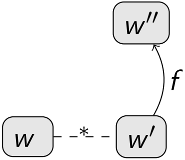




- Suppose $w \in C(S - Rat)$. Take any w' reachable from w. Assume, towards contradiction, that $\sigma(w)_i(v) = a$ is not the BI move for player i.

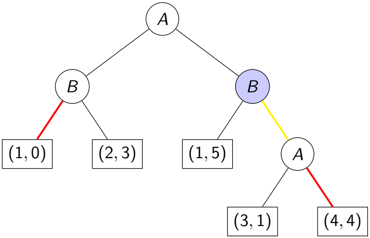
Proof of Halpern's Theorem

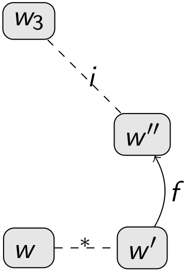




- Induction step. Suppose $w \in C(S - Rat)$. Take any w' reachable from w. Assume, towards contradiction, that $\sigma(w)_i(v) = a$ is not the BI move for player i. Since w is also in $C(S - Rat)$, we know by definition i must be rational at $w'' = f(w', v)$. But then, by F3 and our IH, all players play according to the BI solution after v at w''.

Proof of Halpern's Theorem



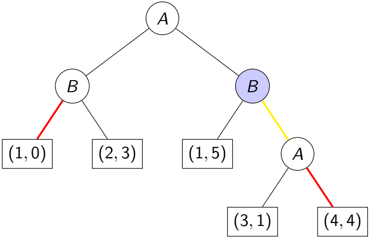


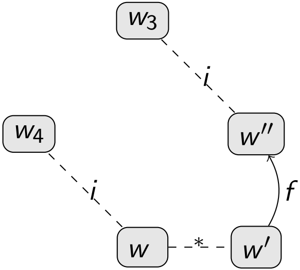
- i's rationality at w'' means, in particular, that there is a $w_3 \in [w'']_i$ such that

$$h_i^v(\sigma_i(w''), \sigma_{-i}(w_3)) \geq h_i^v((bi_i, \sigma_{-i}(w_3)))$$

for bi_i i's backward induction strategy.

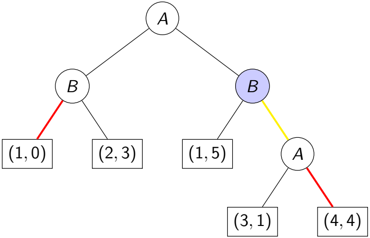
Proof of Halpern's Theorem

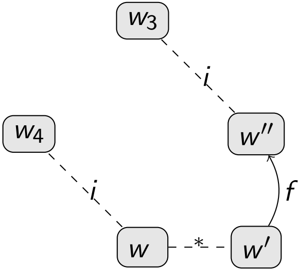




- But then by F4 there must exists $w_4 \in [w]_i$ such that $\sigma(w_4)$ $\sigma(w_3)$ at the same in the sub-tree starting at v.

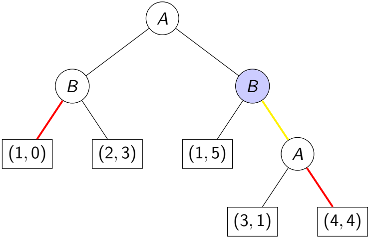
Proof of Halpern's Theorem

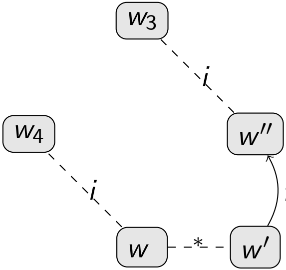




- But then by F4 there must exists $w_4 \in [w]_i$ such that $\sigma(w_4)$ $\sigma(w_3)$ at the same in the sub-tree starting at v. Since w_4 is reachable from w, in that state all players play according to the backward induction after v, and so this is also true of w_3.

Proof of Halpern's Theorem





- But then by F4 there must exists $w_4 \in [w]_i$ such that $\sigma(w_4)$ $\sigma(w_3)$ at the same in the sub-tree starting at v. Since w_4 is reachable from w, in that state all players play according to the backward induction after v, and so this is also true of w_3. But then since the game is non-degenerate, playing something else than bi_i must make i strictly worst off at that state, a contradiction.