Puzzles and Paradoxes from Decision and Game Theory

Eric Pacuit
University of Maryland
pacuit.org

July 21, 2017

Plan

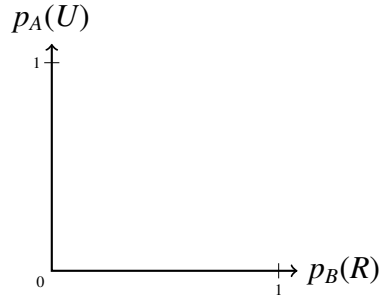
- ✓ Day 1: Rational Choice Theory, Decision Theory
- ✓ Day 2: Expected Utility Theory, Allais Paradox
- ✓ Day 3: Evidential and Causal Decision Theory,
- ✓ Day 4: Introduction to (Epistemic) Game Theory, Common Knowledge, Backward Induction
- ▸ Day 5: Epistemic Game Theory, Paradoxes of Interactive Epistemology, Imperfect Recall

A \ B	L	R
U	2,1	0,0
D	0,0	1,2

Strategic Game

		B	
		L	R
A	U	2,1	0,0
	D	0,0	1,2

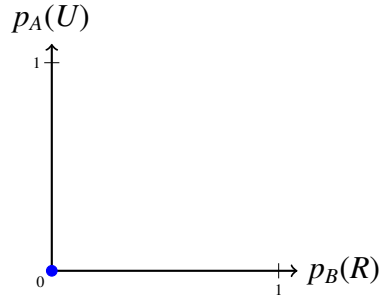
Strategic Game



Solution Space

		B: L	B: R
A	U	2,1	0,0
A	D	0,0	1,2

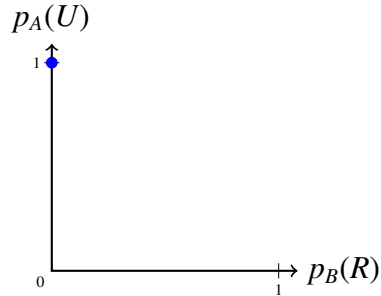
Strategic Game



Solution Space

		B	
		L	R
A	U	2,1	0,0
	D	0,0	1,2

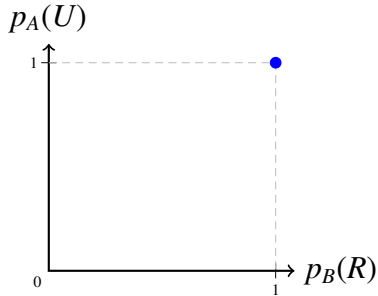
Strategic Game



Solution Space

		B	
		L	R
A	U	2,1	0,0
	D	0,0	1,2

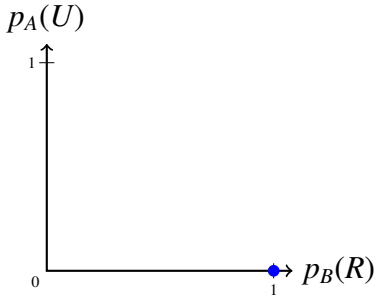
Strategic Game



Solution Space

		B	
		L	R
A	U	2,1	0,0
	D	0,0	1,2

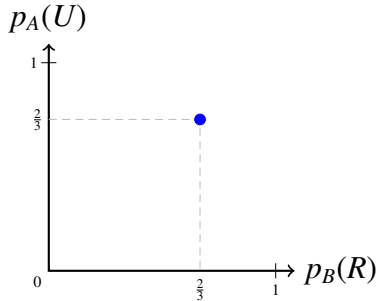
Strategic Game



Solution Space

		B: L	B: R
A	U	2,1	0,0
A	D	0,0	1,2

Strategic Game



Solution Space

Game Models

- A game is a *partial* description of a set (or sequence) of interdependent **(Bayesian) decision problems**.

Game Models

- A game is a *partial* description of a set (or sequence) of interdependent **(Bayesian) decision problems**.

  A game will not normally contain enough information to determine what the players *believe* about each other.

Game Models

- A game is a *partial* description of a set (or sequence) of interdependent **(Bayesian) decision problems**.

  A game will not normally contain enough information to determine what the players *believe* about each other.

- A **model of a game** is a completion of the partial specification of the Bayesian decision problems *and* a representation of a particular play of the game.

Game Models

- A game is a *partial* description of a set (or sequence) of interdependent **(Bayesian) decision problems**.

  A game will not normally contain enough information to determine what the players *believe* about each other.

- A **model of a game** is a completion of the partial specification of the Bayesian decision problems *and* a representation of a particular play of the game.

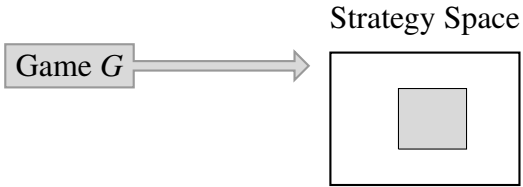
- There are no special rules of rationality telling one what to do in the absence of degrees of belief except: decide what you believe, and then **maximize (subjective) expected utility**.

The Epistemic Program in Game Theory

Game G

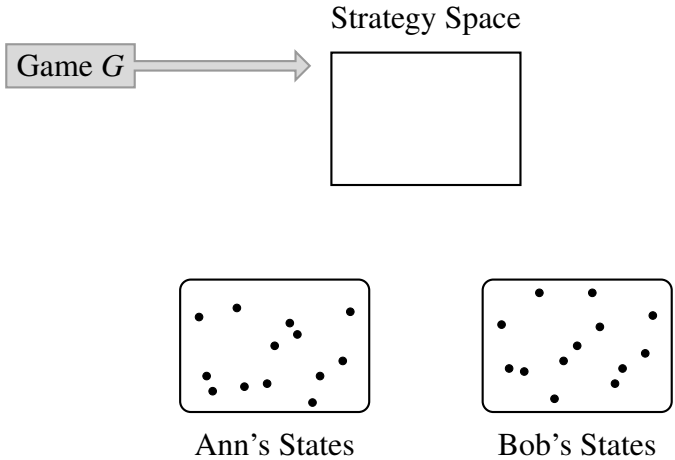
G: available actions, payoffs, structure of the decision problem

The Epistemic Program in Game Theory



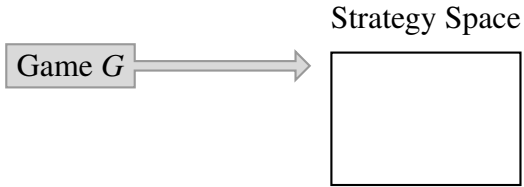
solution concepts are systematic descriptions of what players *do*

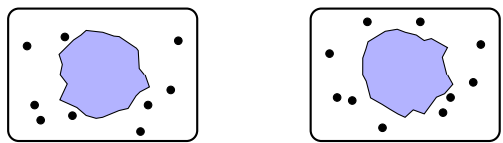
The Epistemic Program in Game Theory



The game model includes *information states* of the players

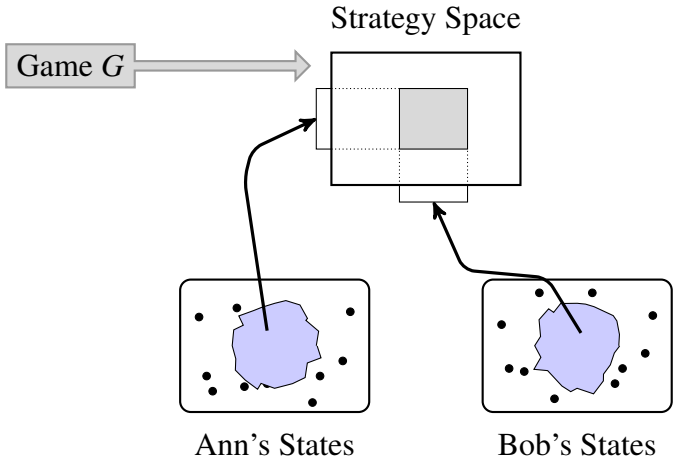
The Epistemic Program in Game Theory





Restrict to information states satisfying some rationality condition

The Epistemic Program in Game Theory



Project onto the strategy space

Models of Games

Suppose that G is a game.

- Outcomes of the game: $S = \Pi_{i \in N} S_i$
- A profile is a vector $\vec{s} \in S$, specifying an action for each player
- Player i's partial beliefs (or conjecture): $P_i \in \Delta(S_{-i})$

  $\Delta(X)$ *is the set of probabilities measures over* X

Models of Games, continued

$G = \langle N, \{S_i, u_i\}_{i \in N} \rangle$ is a strategic (form of a) game.

Models of Games, continued

$G = \langle N, \{S_i, u_i\}_{i \in N} \rangle$ is a strategic (form of a) game.

- W is a set of *possible worlds* (possible outcomes of the game)
- $\mathbf{s}$ is a function $\mathbf{s} : W \to \Pi_{i \in N} S_i$, write $\mathbf{s}_i(w)$ for the ith component of $\mathbf{s}(w)$

Models of Games, continued

$G = \langle N, \{S_i, u_i\}_{i \in N} \rangle$ is a strategic (form of a) game.

- W is a set of *possible worlds* (possible outcomes of the game)
- $\mathbf{s}$ is a function $\mathbf{s} : W \to \Pi_{i \in N} S_i$, write $\mathbf{s}_i(w)$ for the ith component of $\mathbf{s}(w)$
- If $\vec{s} \in \Pi_{i \in N} S_i$, then $[\vec{s}] = \{w \mid \mathbf{s}(w) = \vec{s}\}$; if $s_i \in S_i$, then $[s_i] = \{w \mid \mathbf{s}_i(w) = s_i\}$; and if $X \subseteq S$, $[X] = \bigcup_{s \in X} [s]$.

Models of Games, continued

$G = \langle N, \{S_i, u_i\}_{i \in N} \rangle$ is a strategic (form of a) game.

- W is a set of *possible worlds* (possible outcomes of the game)
- $\mathbf{s}$ is a function $\mathbf{s} : W \to \Pi_{i \in N} S_i$, write $\mathbf{s}_i(w)$ for the ith component of $\mathbf{s}(w)$
- If $\vec{s} \in \Pi_{i \in N} S_i$, then $[\vec{s}] = \{w \mid \mathbf{s}(w) = \vec{s}\}$; if $s_i \in S_i$, then $[s_i] = \{w \mid \mathbf{s}_i(w) = s_i\}$; and if $X \subseteq S$, $[X] = \bigcup_{s \in X} [s]$.
- ***ex ante* beliefs**: For each $i \in N$, let $P_i \in \Delta(W)$ (the set of probability measures on W). Two assumptions:
  - $[s]$ is measurable for all strategy profiles $s \in S$
  - $P_i([s_i]) > 0$ for all $s_i \in S_i$

***ex interim* beliefs**: $P_{i,w} \in \Delta(S_{-i})$

- ...given player i's choice: $P_{i,w}(\cdot) = P_i(\cdot \mid [\mathbf{s}_i(w)])$
- ...given all player i *knows*: $P_{i,w}(\cdot) = P_i(\cdot \mid K_i)$, $\quad K_i \subseteq [\mathbf{s}_i(w)]$
- ...given all player i *fully believes*: $P_{i,w}(\cdot) = P_i(\cdot \mid B_i)$, $\quad B_i \subseteq [\mathbf{s}_i(w)]$

***ex interim* beliefs**: $P_{i,w} \in \Delta(S_{-i})$

- ...given player i's choice: $P_{i,w}(\cdot) = P_i(\cdot \mid [\mathbf{s}_i(w)])$
- ...given all player i *knows*: $P_{i,w}(\cdot) = P_i(\cdot \mid K_i)$, $K_i \subseteq [\mathbf{s}_i(w)]$
- ...given all player i *fully believes*: $P_{i,w}(\cdot) = P_i(\cdot \mid B_i)$, $B_i \subseteq [\mathbf{s}_i(w)]$

Expected utility of strategy $s_i \in S_i$: Given $P \in \Delta(S_{-i})$,

$$EU_{i,P}(s_i) = \sum_{s_{-i} \in S_{-i}} P(s_{-i}) u_i(s_i, s_{-i})$$

***ex interim* beliefs**: $P_{i,w} \in \Delta(S_{-i})$

- ...given player i's choice: $P_{i,w}(\cdot) = P_i(\cdot \mid [\mathbf{s}_i(w)])$
- ...given all player i *knows*: $P_{i,w}(\cdot) = P_i(\cdot \mid K_i), \qquad K_i \subseteq [\mathbf{s}_i(w)]$
- ...given all player i *fully believes*: $P_{i,w}(\cdot) = P_i(\cdot \mid B_i), \quad B_i \subseteq [\mathbf{s}_i(w)]$

Expected utility of strategy $s_i \in S_i$: Given $w \in W$,

$$EU_{i,w}(s_i) = \sum_{s_{-i} \in S_{-i}} P_{i,w}([s_{-i}]) u_i(s_i, s_{-i})$$

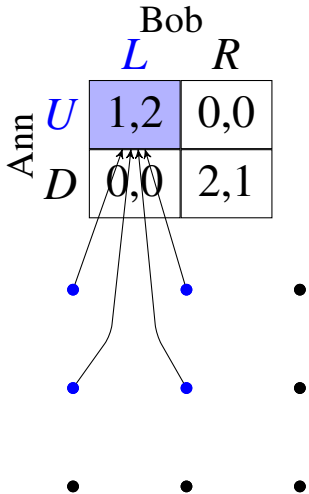
An Example

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

An Example

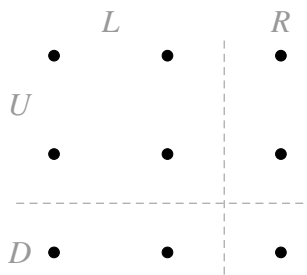
Ann \ Bob	L	R
U	1,2	0,0
D	0,0	2,1

An Example



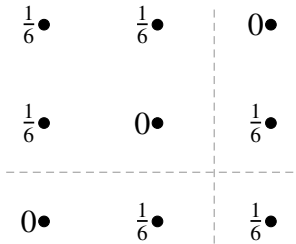
An Example

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



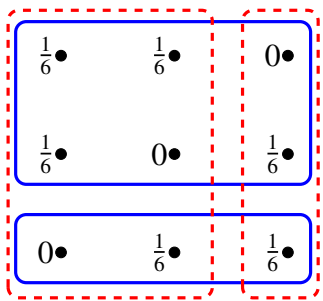
An Example

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



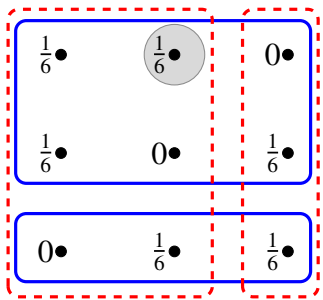
An Example

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



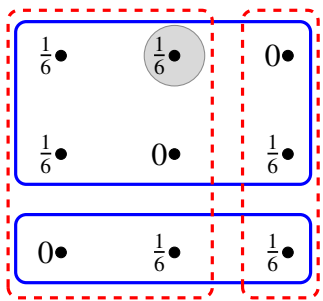
An Example

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



An Example

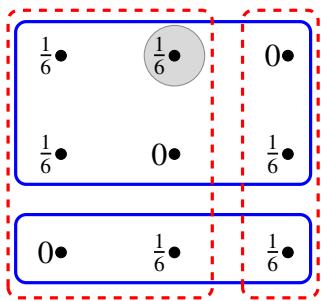
		Bob L	Bob R
Ann	U	1,2	0,0
	D	0,0	2,1



- Ann's choice is *optimal* (given her information)

An Example

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

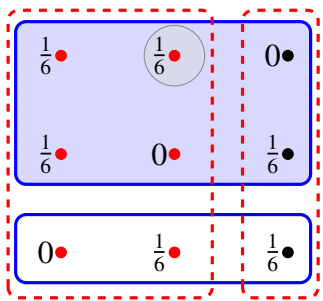


- Ann's choice is *optimal* (given her information)

$$1 \cdot P_A(L) + 0 \cdot P_A(R) \geq 0 \cdot P_A(L) + 2 \cdot P_A(R)$$

An Example

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

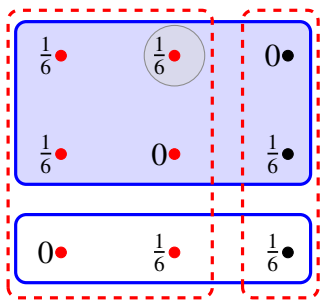


- Ann's choice is *optimal* (given her information)

$$1 \cdot P_A(L) + 0 \cdot P_A(R) \geq 0 \cdot P_A(L) + 2 \cdot P_A(R)$$

An Example

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

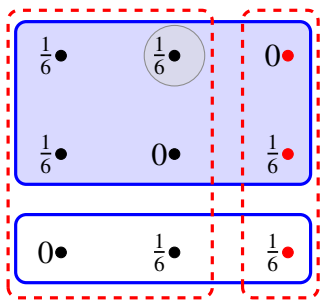


- Ann's choice is *optimal* (given her information)

$$1 \cdot \tfrac{3}{4} + 0 \cdot P_A(R) \geq 0 \cdot \tfrac{3}{4} + 2 \cdot P_A(R)$$

An Example

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

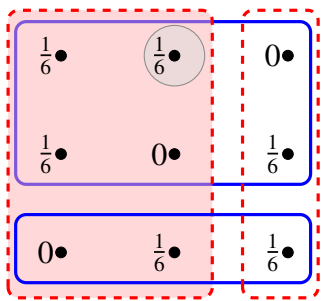


- Ann's choice is *optimal* (given her information)

$$1 \cdot \tfrac{3}{4} + 0 \cdot \tfrac{1}{4} \geq 0 \cdot \tfrac{3}{4} + 2 \cdot \tfrac{1}{4}$$

An Example

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

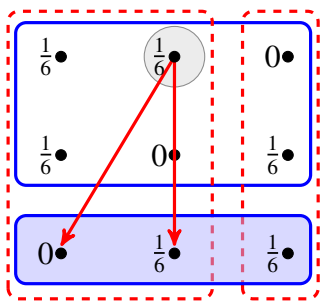


- Ann's choice is *optimal* (given her information)
- Bob's choice is *optimal* (given her information)

$$2 \cdot \tfrac{3}{4} + 0 \cdot \tfrac{1}{4} \geq 0 \cdot \tfrac{3}{4} + 1 \cdot \tfrac{1}{4}$$

An Example

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



- Ann's choice is *optimal* (given her information)
- Bob's choice is *optimal* (given her information)
- Bob *considers it possible* Ann is *irrational*

$$1 \cdot \tfrac{1}{2} + 0 \cdot \tfrac{1}{2} \not\geq 0 \cdot \tfrac{1}{2} + 2 \cdot \tfrac{1}{2}$$

For any $P \in \Delta(S_{-i})$ and $s_i \in S_i$, $EU_{i,P}(s_i) = \sum_{s_{-i} \in S_{-i}} P(s_{-i}) u_i(s_i, s_{-i})$

For any $P \in \Delta(S_{-i})$ and $s_i \in S_i$, $EU_{i,P}(s_i) = \sum_{s_{-i} \in S_{-i}} P(s_{-i}) u_i(s_i, s_{-i})$

For any $w \in W$ and $s_i \in S_i$, $EU_{i,w}(s_i) = \sum_{s_{-i} \in S_{-i}} P_{i,w}([s_{-i}]) u_i(s_i, s_{-i})$

For any $P \in \Delta(S_{-i})$ and $s_i \in S_i$, $EU_{i,P}(s_i) = \sum_{s_{-i} \in S_{-i}} P(s_{-i}) u_i(s_i, s_{-i})$

For any $w \in W$ and $s_i \in S_i$, $EU_{i,w}(s_i) = \sum_{s_{-i} \in S_{-i}} P_{i,w}([s_{-i}]) u_i(s_i, s_{-i})$

$\mathsf{Rat}_i = \{w \mid EU_{i,w}(\mathbf{s}_i(w)) \geq EU_{i,w}(s_i) \text{ for all } s_i \in S_i\}$

For any $P \in \Delta(S_{-i})$ and $s_i \in S_i$, $EU_{i,P}(s_i) = \sum_{s_{-i} \in S_{-i}} P(s_{-i}) u_i(s_i, s_{-i})$

For any $w \in W$ and $s_i \in S_i$, $EU_{i,w}(s_i) = \sum_{s_{-i} \in S_{-i}} P_{i,w}([s_{-i}]) u_i(s_i, s_{-i})$

$\mathsf{Rat}_i = \{w \mid EU_{i,w}(\mathbf{s}_i(w)) \geq EU_{i,w}(s_i) \text{ for all } s_i \in S_i\}$

Each $P \in \Delta(W)$ is associated with $P^S \in \Delta(S)$ as follows: for all $s \in S$, $P^S(s) = P([s])$

For any $P \in \Delta(S_{-i})$ and $s_i \in S_i$, $EU_{i,P}(s_i) = \sum_{s_{-i} \in S_{-i}} P(s_{-i}) u_i(s_i, s_{-i})$

For any $w \in W$ and $s_i \in S_i$, $EU_{i,w}(s_i) = \sum_{s_{-i} \in S_{-i}} P_{i,w}([s_{-i}]) u_i(s_i, s_{-i})$

$\mathsf{Rat}_i = \{w \mid EU_{i,w}(\mathbf{s}_i(w)) \geq EU_{i,w}(s_i) \text{ for all } s_i \in S_i\}$

Each $P \in \Delta(W)$ is associated with $P^S \in \Delta(S)$ as follows: for all $s \in S$, $P^S(s) = P([s])$

A mixed strategy $\sigma \in \Pi_{i \in N} \Delta(S_i)$, $P_\sigma \in \Delta(S)$, $P_\sigma(s) = \sigma_1(s_1) \cdots \sigma_n(s_n)$

Characterizing Nash Equilibria

Theorem (Aumann). σ is a Nash equilibrium of G iff there exists a model $\mathcal{M}^G = \langle W, \{P_i\}_{i \in N}, \mathbf{s} \rangle$ such that:

- for all $i \in N$, $\mathsf{Rat}_i = W$;
- for all $i, j \in N$, $P_i = P_j$; and
- for all $i \in N$, $P_i^S = P_\sigma$.

Rationalizability

A **best reply set** (BRS) is a sequence $(B_1, B_2, \ldots, B_n) \subseteq S = \Pi_{i \in N} S_i$ such that for all $i \in N$, and all $b_i \in B_i$, there exists $\mu_{-i} \in \Delta(B_{-i})$ such that s_i is a best response to μ_{-i}: I.e.,

$$b_i = \arg\max_{s_i \in S_i} EU_{i,\mu_{-i}}(s_i)$$

		2			
		b_1	b_2	b_3	b_4
1	a_1	0, 7	2, 5	7, 0	0, 1
	a_2	5, 2	3, 3	5, 2	0, 1
	a_3	7, 0	2, 5	0, 7	0, 1
	a_4	0, 0	0, -2	0, 0	10, -1

		2			
		b_1	b_2	b_3	b_4
1	a_1	0, 7	2, 5	7, 0	0, 1
	a_2	5, 2	3, 3	5, 2	0, 1
	a_3	7, 0	2, 5	0, 7	0, 1
	a_4	0, 0	0, -2	0, 0	10, -1

- (a_2, b_2) is the unique Nash equilibria, hence $(\{a_2\}, \{b_2\})$ is a BRS

		2			
		b_1	b_2	b_3	b_4
1	a_1	0, 7	2, 5	7, 0	0, 1
	a_2	5, 2	3, 3	5, 2	0, 1
	a_3	7, 0	2, 5	0, 7	0, 1
	a_4	0, 0	0, -2	0, 0	10, -1

- (a_2, b_2) is the unique Nash equilibria, hence $(\{a_2\}, \{b_2\})$ is a BRS
- $(\{a_1, a_3\}, \{b_1, b_3\})$ is a BRS

1 \ 2	b_1	b_2	b_3	b_4
a_1	0, 7	2, 5	7, 0	0, 1
a_2	5, 2	3, 3	5, 2	0, 1
a_3	7, 0	2, 5	0, 7	0, 1
a_4	0, 0	0, -2	0, 0	10, -1

- (a_2, b_2) is the unique Nash equilibria, hence $(\{a_2\}, \{b_2\})$ is a BRS
- $(\{a_1, a_3\}, \{b_1, b_3\})$ is a BRS
- $(\{a_1, a_2, a_3\}, \{b_1, b_2, b_3\})$ is a full BRS

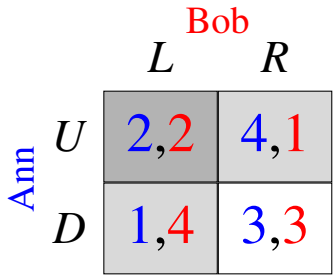
Theorem (Bernheim; Pearce; Brandenburger and Dekel; ...). $(B_1, B_2, \ldots, B_n)$ is a BRS for G iff there exists a model $\mathcal{M}^G = \langle W, \{P_i\}_{i \in N}, \mathbf{s} \rangle$ such that for all $i \in N$, $\mathsf{Rat}_i = W$ and $[B_1 \times \cdots \times B_n] = W$.

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

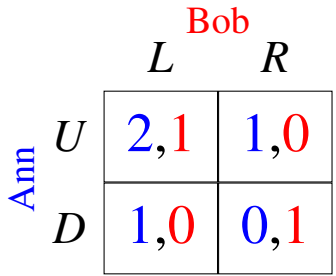
Game 1

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

Game 2

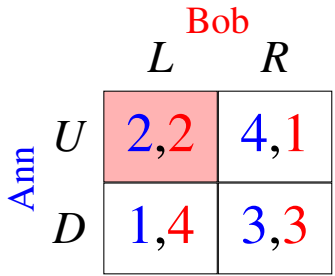


Game 1

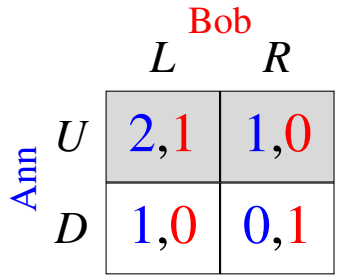


Game 2

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



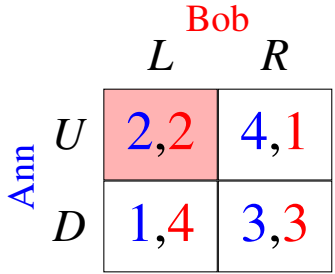
Game 1



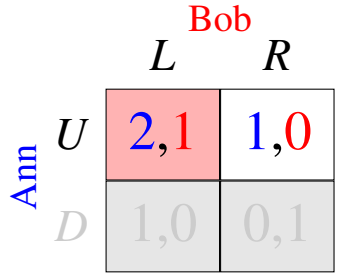
Game 2

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

Game 2: U strictly dominates D



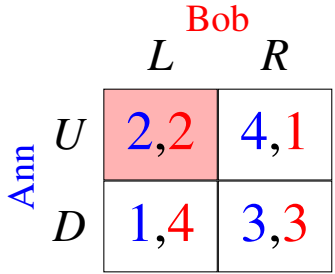
Game 1



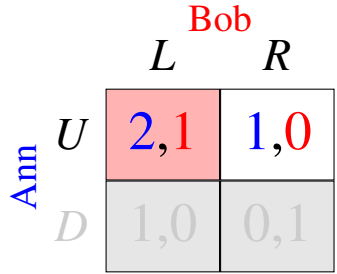
Game 2

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.



Game 1



Game 2

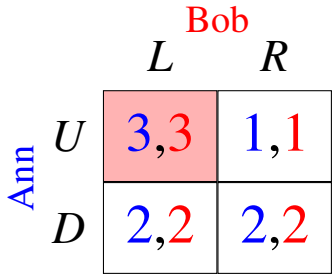
Game 1: U strictly dominates D and L strictly dominates R.

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

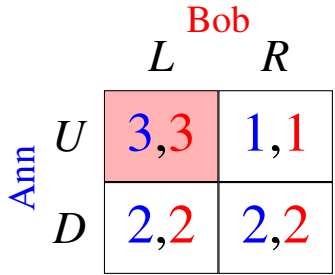
Theorem. In all models where the players are *rational* and there is *common belief of rationality*, the players choose strategies that survive iterative removal of strictly dominated strategies (and, conversely...).

Let $P \in \Delta(X)$ be a probability measure, the **support** of P is $\mathsf{supp}(P) = \{x \in X \mid P(x) > 0\}$.

A probability measure $P \in \Delta(X)$ is said to be a **full support** probability measure on X provided $\mathsf{supp}(P) = X$.

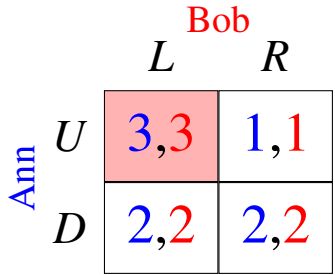


Is R rationalizable?



Is R rationalizable?

There is no *full support* probability such that R is a best response



Is R rationalizable?

There is no *full support* probability such that R is a best response

Should Ann assign probability 0 to R or probability > 0 to R?

Strategic Reasoning and Admissibility

"The argument for deletion of a weakly dominated strategy for player i is that he contemplates the possibility that every strategy combination of his rivals occurs with positive probability. However, this hypothesis clashes with the logic of iterated deletion, which assumes, precisely, that eliminated strategies are not expected to occur."

Mas-Colell, Whinston and Green. *Introduction to Microeconomics*. 1995.

A Puzzle

R. Cubitt and R. Sugden. *Rationally Justifiable Play and the Theory of Non-cooperative games*. Economic Journal, 104, pgs. 798 - 803, 1994.

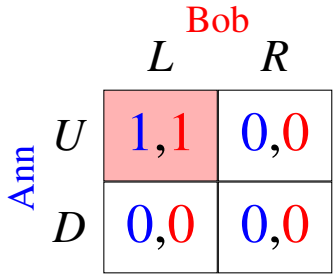
R. Cubitt and R. Sugden. *Common reasoning in games: A Lewisian analysis of common knowledge of rationality*. Manuscript, 2011.

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

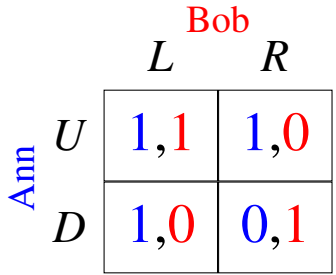
Game 1

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

Game 2

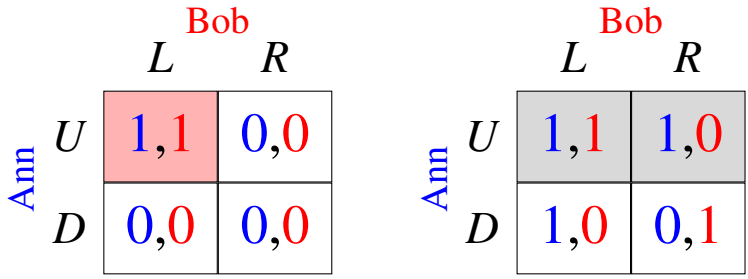


Game 1



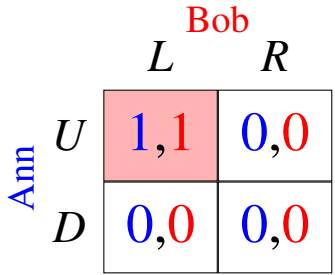
Game 2

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

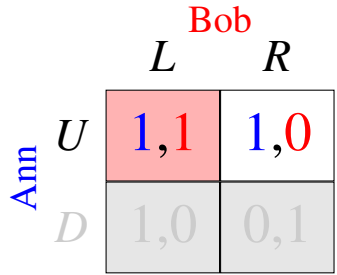


Game 1: U weakly dominates D and L weakly dominates R.

Game 2: U weakly dominates D



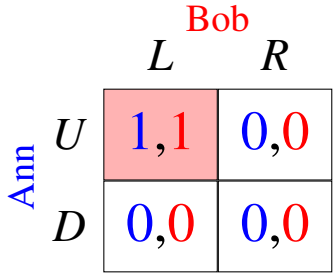
Game 1



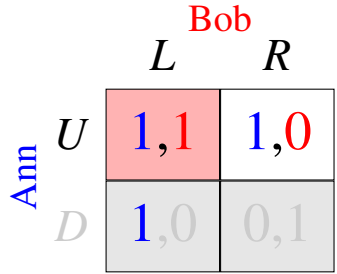
Game 2

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

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



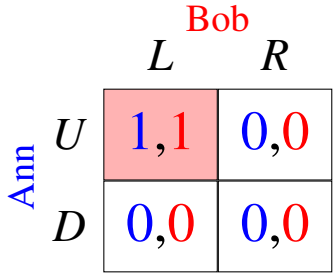
Game 1



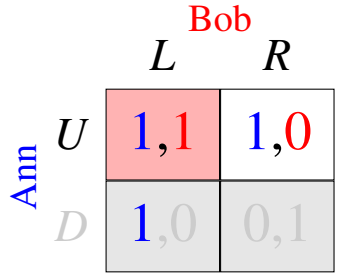
Game 2

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

Game 2: But, now what is the reason for not playing D?



Game 1



Game 2

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

Game 2: But, now what is the reason for not playing D?

Theorem (Samuelson). There is no model of Game 2 satisfying common knowledge of rationality (where rationality incorporates weak dominance).

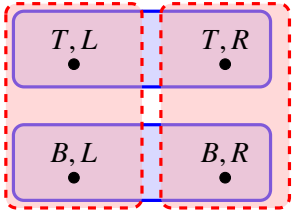
Common Knowledge of Admissibility

Ann \ Bob	*L*	*R*
T	1,1	1,0
B	1,0	0,1

There is no model of this game with *common knowledge* of admissibility.

Common Knowledge of Admissibility

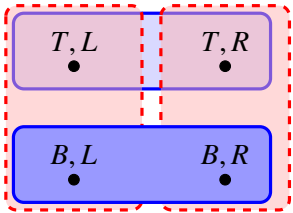
Ann \ Bob	L	R
T	1,1	1,0
B	1,0	0,1



The ”full” model of the game

Common Knowledge of Admissibility

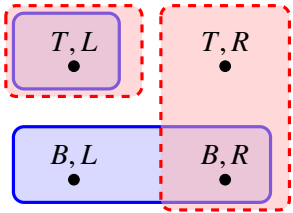
Ann \ Bob	L	R
T	1,1	1,0
B	1,0	0,1



The "full" model of the game: *B is not admissible given Ann's information*

Common Knowledge of Admissibility

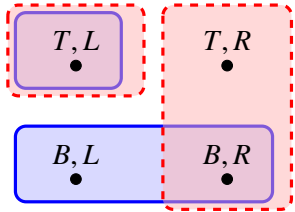
Ann \ Bob	L	R
T	1,1	1,0
B	1,0	0,1



What is wrong with this model?

Common Knowledge of Admissibility

Ann \ Bob	L	R
T	1,1	1,0
B	1,0	0,1



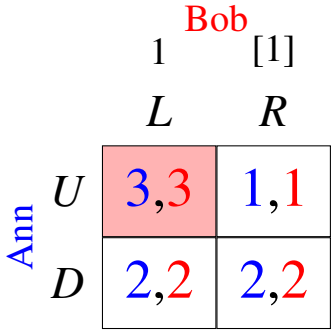
Privacy of Tie-Breaking/No Extraneous Beliefs: If a strategy is *rational* for an opponent, then it cannot be "ruled out".

Both Including and Excluding a Strategy

Returning to the problem of weakly dominated strategies and rationalizability, one solution is to assume that players consider some strategies *infinitely more likely than other strategies*.

Both Including and Excluding a Strategy

Returning to the problem of weakly dominated strategies and rationalizability, one solution is to assume that players consider some strategies *infinitely more likely than other strategies*.



L. Blume, A. Brandenburger, E. Dekel. *Lexicographic probabilities and choice under uncertainty*. Econometrica, 59(1), pgs. 61 - 79, 1991.

In a game model $\mathcal{M}^G = \langle W, \{P_i\}_{i \in N}, \mathbf{s} \rangle$, different states represent different beliefs only when the agent is doing something different.

$$P_{i,w}(E) = P_i(E \mid [\mathbf{s}_i(w)])$$

To represent different *explanations* (i.e., beliefs) for the same strategy choice, we would need a set of models $\{\mathcal{M}^G_1, \mathcal{M}^G_2, \ldots\}$.

In a game model $\mathcal{M}^G = \langle W, \{P_i\}_{i \in N}, \mathbf{s} \rangle$, different states represent different beliefs only when the agent is doing something different.

$$P_{i,w}(H) = P_i(H \mid B_{i,w}), \quad B_{i,w} \subseteq [\mathbf{s}_i(w)]$$

To represent different *explanations* (i.e., beliefs) for the same strategy choice, we would need a set of models $\{\mathcal{M}^G_1, \mathcal{M}^G_2, \ldots\}$.

In a game model $\mathcal{M}^G = \langle W, \{P_i\}_{i \in N}, \mathbf{s} \rangle$, different states represent different beliefs only when the agent is doing something different.

$$P_{i,w}(H) = P_i(H \mid B_{i,w}), \quad B_{i,w} \subseteq [\mathbf{s}_i(w)]$$

Two way to change beliefs: $P_i(\cdot \mid E \cap B_{i,w})$ and $P_i(\cdot \mid B'_{i,w})$ (conditioning on 0 events).

Game Models

Richer models of a game: lexicographic probabilities, conditional probability systems, non-standard probabilities, plausibility models, ...
(type spaces)

Game Models

Richer models of a game: lexicographic probabilities, conditional probability systems, non-standard probabilities, plausibility models, ...
(type spaces)

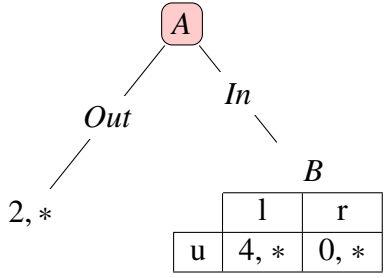
"The aim in giving the general definition of a model is not to propose an original explanatory hypothesis, or any explanatory hypothesis, for the behavior of players in games, but only to provide a descriptive framework for the representation of considerations that are relevant to such explanations, a framework that is as *general* and as *neutral* as we can make it." (pg. 35)

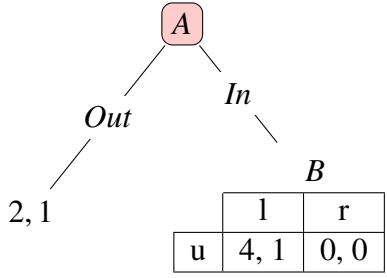
R. Stalnaker. *Knowledge, Belief and Counterfactual Reasoning in Games*. Economics and Philosophy, 12(1), pgs. 133 - 163, 1996.

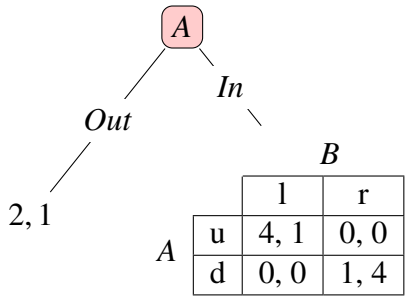
Rationalizing Observed Actions

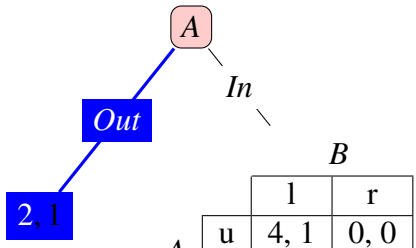
After observing an (unexpected) move by some player, you could:

1. Change your belief about the player's rationality, but maintain your beliefs about the player's *passive beliefs*.
2. Change your belief about the player's passive beliefs, but maintain your belief in the player's rationality.
3. Conclude that the player perceives the game differently.


A
Out
In
2, ∗
B
l
r
u
4, ∗
0, ∗


A
Out
In
2, 1
B
l
r
u
4, 1
0, 0

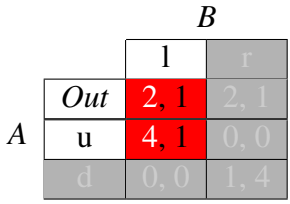

A
Out
In
2, 1
B
l
r
A
u
4, 1
0, 0
d
0, 0
1, 4

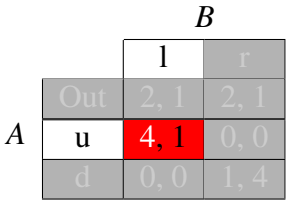

A
In
Out
2, 1
B
l
r
A
u
4, 1
0, 0
d
0, 0
1, 4

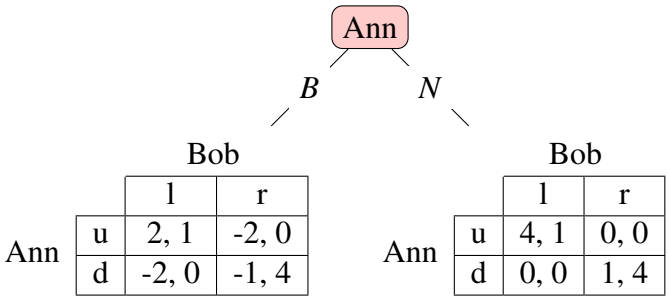
		B	
		l	r
A	*Out*	2, 1	2, 1
	u	4, 1	0, 0
	d	0, 0	1, 4

		B	
		l	r
A	*Out*	2, 1	2, 1
	u	4, 1	0, 0
	d	0, 0	1, 4

		B	
		l	r
A	*Out*	2, 1	2, 1
	u	4, 1	0, 0
	d	0, 0	1, 4


B
l
r
A
Out
2, 1
2, 1
u
4, 1
0, 0
d
0, 0
1, 4


B
l
r
A
Out
2, 1
2, 1
u
4, 1
0, 0
d
0, 0
1, 4


Ann
B
N
Bob
l
r
Ann
u
2, 1
-2, 0
d
-2, 0
-1, 4
Bob
l
r
Ann
u
4, 1
0, 0
d
0, 0
1, 4

		Bob			
		ll	lr	rl	rr
Ann	Bu	2, 1	2, 1	-2, 0	-2, 0
	Bd	-2, 0	-2, 0	-1, 4	-1, 4
	Nu	4, 1	0, 0	4, 1	0, 0
	Nd	0, 0	1, 4	0, 0	1, 4

		Bob			
		ll	lr	rl	rr
Ann	Bu	2, 1	2, 1	-2, 0	-2, 0
	Bd	-2, 0	-2, 0	-1, 4	-1, 4
	Nu	4, 1	0, 0	4, 1	0, 0
	Nd	0, 0	1, 4	0, 0	1, 4

		Bob			
		ll	lr	rl	rr
Ann	Bu	2, 1	2, 1	-2, 0	-2, 0
	Bd	-2, 0	-2, 0	-1, 4	-1, 4
	Nu	4, 1	0, 0	4, 1	0, 0
	Nd	0, 0	1, 4	0, 0	1, 4

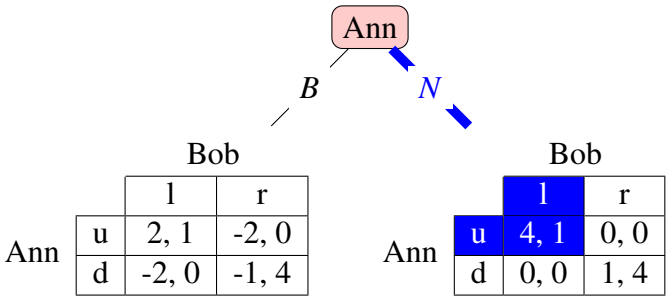
		Bob			
		ll	lr	rl	rr
Ann	Bu	2, 1	2, 1	-2, 0	-2, 0
	Bd	-2, 0	-2, 0	-1, 4	-1, 4
	Nu	4, 1	0, 0	4, 1	0, 0
	Nd	0, 0	1, 4	0, 0	1, 4

		Bob			
		ll	lr	rl	rr
Ann	Bu	2, 1	2, 1	-2, 0	-2, 0
	Bd	-2, 0	-2, 0	-1, 4	-1, 4
	Nu	4, 1	0, 0	4, 1	0, 0
	Nd	0, 0	1, 4	0, 0	1, 4

		Bob			
		ll	lr	rl	rr
Ann	Bu	2, 1	2, 1	-2, 0	-2, 0
	Bd	-2, 0	-2, 0	-1, 4	-1, 4
	Nu	4, 1	0, 0	4, 1	0, 0
	Nd	0, 0	1, 4	0, 0	1, 4

		Bob			
		ll	lr	rl	rr
Ann	Bu	2, 1	2, 1	-2, 0	-2, 0
	Bd	-2, 0	-2, 0	-1, 4	-1, 4
	Nu	4, 1	0, 0	4, 1	0, 0
	Nd	0, 0	1, 4	0, 0	1, 4

		Bob			
		ll	lr	rl	rr
Ann	Bu	2, 1	2, 1	-2, 0	-2, 0
	Bd	-2, 0	-2, 0	-1, 4	-1, 4
	Nu	4, 1	0, 0	4, 1	0, 0
	Nd	0, 0	1, 4	0, 0	1, 4

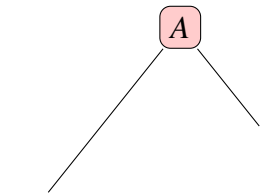

Ann
B
N
Bob
l
r
Ann
u
2, 1
-2, 0
d
-2, 0
-1, 4
Bob
l
r
Ann
u
4, 1
0, 0
d
0, 0
1, 4

What is forward induction reasoning?

Forward Induction Principle: a player should use all information she acquired about her opponents' past behavior in order to improve her prediction of their future simultaneous and past (unobserved) behavior, relying on the assumption that they are rational.

P. Battigalli. *On Rationalizability in Extensive Games*. Journal of Economic Theory, 74, pgs. 40 - 61, 1997.

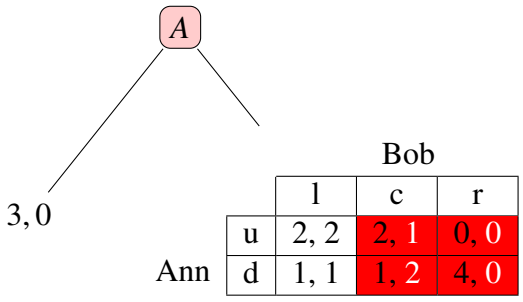
Backward *versus* Forward Induction



		Bob		
		l	c	r
Ann	u	2, 2	2, 1	0, 0
	d	1, 1	1, 2	4, 0

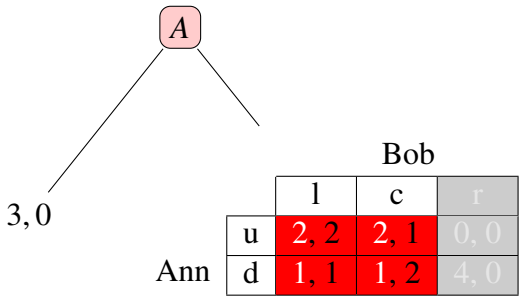
A. Perea. *Backward Induction* versus *Forward Induction Reasoning*. Games, 1, pgs. 168 - 188, 2010.

Backward *versus* Forward Induction



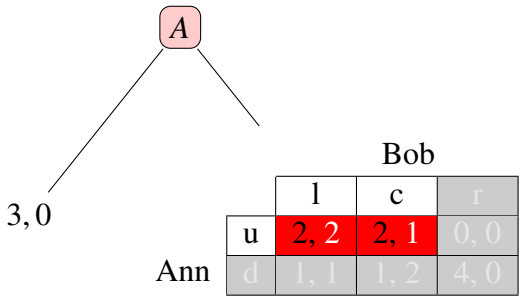
A. Perea. *Backward Induction* versus *Forward Induction Reasoning*. Games, 1, pgs. 168 - 188, 2010.

Backward *versus* Forward Induction



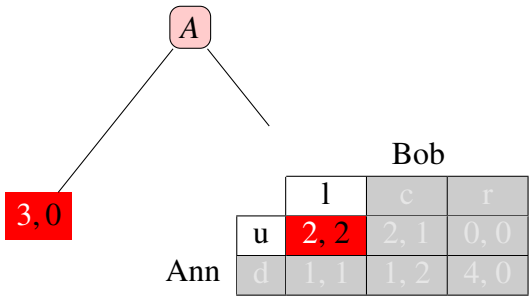
A. Perea. *Backward Induction* versus *Forward Induction Reasoning*. Games, 1, pgs. 168 - 188, 2010.

Backward *versus* Forward Induction



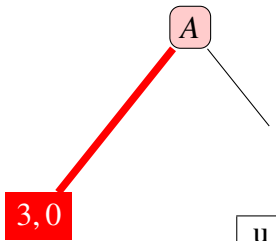
A. Perea. *Backward Induction* versus *Forward Induction Reasoning*. Games, 1, pgs. 168 - 188, 2010.

Backward *versus* Forward Induction



A. Perea. *Backward Induction* versus *Forward Induction Reasoning*. Games, 1, pgs. 168 - 188, 2010.

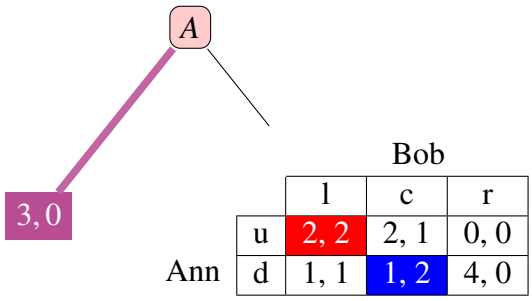
Backward *versus* Forward Induction



		Bob		
		l	c	r
Ann	u	2, 2	2, 1	0, 0
	d	1, 1	1, 2	4, 0

A. Perea. *Backward Induction* versus *Forward Induction Reasoning*. Games, 1, pgs. 168 - 188, 2010.

Backward *versus* Forward Induction



A. Perea. *Backward Induction* versus *Forward Induction Reasoning*. Games, 1, pgs. 168 - 188, 2010.

A. Knoks and EP. *Interpreting Mistakes in Games: From Beliefs about Mistakes to Mistaken Beliefs*. manuscript, 2016.

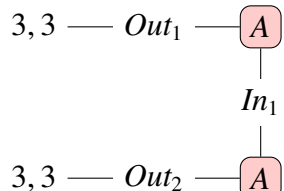
Rationalization *versus* Mistakes

In

B

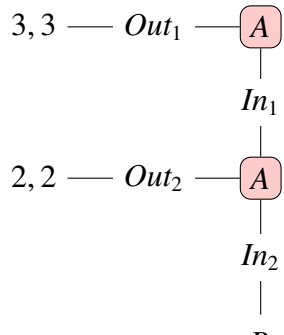
		l	r
A	u	4, 1	0, 0
	d	0, 0	1, 4

Rationalization *versus* Mistakes



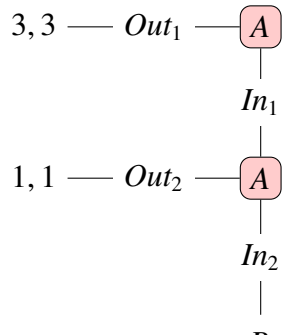
		l	r
A	u	4, 1	0, 0
	d	0, 0	1, 4

Rationalization *versus* Mistakes



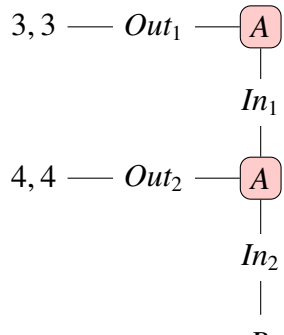
A		l	r
	u	4, 1	0, 0
	d	0, 0	1, 4

Rationalization *versus* Mistakes



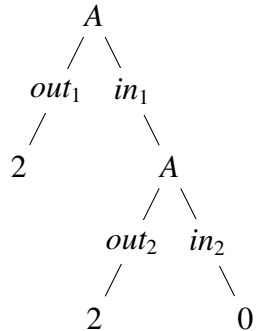
A		l	r
	u	4, 1	0, 0
	d	0, 0	1, 4

Rationalization *versus* Mistakes

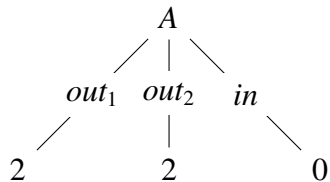


		l	r
A	u	4, 1	0, 0
	d	0, 0	1, 4

Allowing for mistakes



D_1



D_2

Interpreting Mistakes

How should Bob respond given evidence that Ann's moves seem irrational?

Interpreting Mistakes

How should Bob respond given evidence that Ann's moves seem irrational?

1. Bob's beliefs about Ann's perception of the game are incorrect.
2. Bob's assumption about Ann's decision procedure is incorrect.
3. Bob's belief about Ann's assumptions about him is incorrect.
4. Ann's moves are an attempt to influence Bob's behavior in the game.

Interpreting Mistakes

How should Bob respond given evidence that Ann's moves seem irrational?

1. Bob's beliefs about Ann's perception of the game are incorrect.
2. Bob's assumption about Ann's decision procedure is incorrect.
3. Bob's belief about Ann's assumptions about him is incorrect.
4. Ann's moves are an attempt to influence Bob's behavior in the game.
5. Ann simply failed to successfully implement her adopted strategy, i.e., Ann made a "trembling hand mistake".

Backward and Forward Induction

- There are many epistemic characterizations (Aumann, Stalnaker, Battigalli & Siniscalchi, Friedenberg & Siniscalchi, Perea, Baltag & Smets, Bonanno, van Benthem,...)

- How should we compare the two "styles of reasoning" about games? (Heifetz & Perea, Reny, Battigalli & Siniscalchi, Knoks & EP, Perea)

  "When all is said and done, how should we play and what should we expect".

Issues

- The players' conditional beliefs must be *rich enough* to employ the forward induction principle.
- Do the players robustly believe the forward induction principle?
- Can players become more/less confident in the forward induction principle?

Is there a space of all possible interactive beliefs (about a game)?

Is there a space of all possible interactive beliefs (about a game)?

Two questions

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

Is there a space of all possible interactive beliefs (about 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. *Knowledge Spaces with Arbitrarily High Rank*. Games and Economic Behavior (1998).

L. Moss and I. Viglizzo. *Final coalgebras for functors on measurable spaces*. Information and Computation 204(4), pgs. 610-636, 2006.

A. Friendenberg. *When Do Type Structures Contain All Hierarchies of Beliefs?*. Games and Economic Behavior, Vol. 68, 2010.

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).

Richness Conditions

Many epistemic characterization results make a *richness* assumption about the epistemic models.

- What is a "good" epistemic characterization result?
- Players need "enough" conditional beliefs to "make sense of" observed behavior.

Richness Conditions

Many epistemic characterization results make a *richness* assumption about the epistemic models.

- What is a "good" epistemic characterization result?
- Players need "enough" conditional beliefs to "make sense of" observed behavior.

Richness Conditions

Many epistemic characterization results make a *richness* assumption about the epistemic models.

- What is a "good" epistemic characterization result?
- Players need "enough" conditional beliefs to "make sense of" observed behavior.

A. Brandenburger, H. J. Keisler, and A. Friedenberg. *Admissibility in Games*. Econometrica 76(2), pgs. 307-352., 2008.

A. Friedenberg and H. J. Keisler. *Iterated Dominance Revisited*. manuscript, 2010.

J. Halpern and R. Pass . *A logical characterization of iterated admissibility*. in Proceedings of Twelfth Conference on Theoretical, pgs. 146-155, 2009.

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).

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).

EP. *Understanding the Brandenburger-Keisler Paradox*. Studia Logica, 86(3), pgs. 435 - 454, 2007.

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

C. Baskent. *Some non-classical approaches to the BrandenburgerKeisler paradox* . Logic Journal of the IGPL, 23(4): 533-552, 2015.

- Belief paradoxes: The knower paradox, Buriden-Burge sentences, Anti-expert sentences
- The Brandenburger-Keisler (BK) paradox
- Logic of beliefs with definite descriptions for propositions
- Formalizing the paradoxes

The Knower Paradox

Let T be a theory in the language of arithmetic that can prove the Gödel-Carnap fixed-point theorem and K a (perhaps complex) unary predicate in the language of T, such that, for every sentence φ in the language of T, T satisfies:

- $K\varphi \rightarrow \varphi$
- If $T \vdash \varphi$, then $T \vdash K\varphi$

Then, T is inconsistent.

D. Kaplan and R. Montague. *A Paradox Regained*. Notre Dame Journal of Formal Logic, 1, 1960, pp. 79-90.

The Knower Paradox

P. Egré. *The Knower Paradox in the Light of Provability Interpretations of Modal Logic*. Journal of Logic, Language and Information, 14, pgs. 13-48, 2005.

The Knower Paradox

1. $\gamma \leftrightarrow \neg K\ulcorner\gamma\urcorner$ Gödel-Carnap Fixed-Point Lemma

The Knower Paradox

1. $\gamma \leftrightarrow \neg K\ulcorner\gamma\urcorner$ Gödel-Carnap Fixed-Point Lemma
1. $\gamma \leftrightarrow \neg K\gamma$ (Forget Gödel numbering)

The Knower Paradox

1.	$\gamma \leftrightarrow \neg K\ulcorner\gamma\urcorner$	Gödel-Carnap Fixed-Point Lemma
1.	$\gamma \leftrightarrow \neg K\gamma$	(Forget Gödel numbering)
2.	$\gamma \rightarrow \neg K\gamma$	Prop. Reasoning
3.	$K\gamma \rightarrow K\neg K\gamma$	Modal Reasoning
4.	$K\neg K\gamma \rightarrow \neg K\gamma$	T
5.	$K\gamma \rightarrow \neg K\gamma$	Prop. Reasoning
6.	$\neg K\gamma$	Prop. Reasoning

The Knower Paradox

1.	$\gamma \leftrightarrow \neg K\ulcorner\gamma\urcorner$	Gödel-Carnap Fixed-Point Lemma
1.	$\gamma \leftrightarrow \neg K\gamma$	(Forget Gödel numbering)
2.	$\gamma \rightarrow \neg K\gamma$	Prop. Reasoning
3.	$K\gamma \rightarrow K\neg K\gamma$	Modal Reasoning
4.	$K\neg K\gamma \rightarrow \neg K\gamma$	T
5.	$K\gamma \rightarrow \neg K\gamma$	Prop. Reasoning
6.	$\neg K\gamma$	Prop. Reasoning
7.	$\neg K\gamma \rightarrow \gamma$	Prop. Reasoning
8.	γ	Prop. Reasoning
9.	$K\gamma$	Nec
	$\lightning$	

Let T be a theory in the language of arithmetic that can prove the Gödel-Carnap fixed-point theorem and B a (perhaps complex) unary predicate in the language of T, such that, for every sentence φ and ψ in the language of T, T satisfies:

- $B\neg B\varphi \rightarrow \neg B\varphi$
- If $T \vdash \varphi$, then $T \vdash B\varphi$
- If $T \vdash \varphi \leftrightarrow \psi$, then $T \vdash B\varphi \leftrightarrow B\psi$

then T is inconsistent.

R. Thomason. *A note on syntactical treatments of modality*. Synthese 44, pgs. 391 - 395, 1980.

1.	$\gamma \leftrightarrow \neg K\ulcorner\gamma\urcorner$	Gödel-Carnap Fixed-Point Lemma
1.	$\gamma \leftrightarrow \neg K\gamma$	(Forget Gödel numbering)
2.	$\gamma \rightarrow \neg K\gamma$	Prop. Reasoning
3.	$K\gamma \rightarrow K\neg K\gamma$	Modal Reasoning
4.	$K\neg K\gamma \rightarrow \neg K\gamma$	T
5.	$K\gamma \rightarrow \neg K\gamma$	Prop. Reasoning
6.	$\neg K\gamma$	Prop. Reasoning
7.	$\neg K\gamma \rightarrow \gamma$	Prop. Reasoning
8.	γ	Prop. Reasoning
9.	$K\gamma$	Nec
	$\lightning$	

1.	$\gamma \leftrightarrow \neg B^{\ulcorner}\gamma^{\urcorner}$	Gödel-Carnap Fixed-Point Lemma
1.	$\gamma \leftrightarrow \neg B\gamma$	(Forget Gödel numbering)
2.	$\gamma \rightarrow \neg B\gamma$	Prop. Reasoning
3.	$B\gamma \rightarrow B\neg B\gamma$	Modal Reasoning
4.	$B\neg B\gamma \rightarrow \neg B\gamma$	T
5.	$B\gamma \rightarrow \neg B\gamma$	Prop. Reasoning
6.	$\neg B\gamma$	Prop. Reasoning
7.	$\neg B\gamma \rightarrow \gamma$	Prop. Reasoning
8.	γ	Prop. Reasoning
9.	$B\gamma$	Nec
	$\lightning$	

1.	$\gamma \leftrightarrow \neg B^{\ulcorner}\gamma^{\urcorner}$	Gödel-Carnap Fixed-Point Lemma
1.	$\gamma \leftrightarrow \neg B\gamma$	(Forget Gödel numbering)
2.	$\gamma \rightarrow \neg B\gamma$	Prop. Reasoning
3.	$B\gamma \rightarrow B\neg B\gamma$	Modal Reasoning
4.	$B\neg B\gamma \rightarrow \neg B\gamma$	T
5.	$B\gamma \rightarrow \neg B\gamma$	Prop. Reasoning
6.	$\neg B\gamma$	Prop. Reasoning
7.	$\neg B\gamma \rightarrow \gamma$	Prop. Reasoning
8.	γ	Prop. Reasoning
9.	$B\gamma$	Nec
	$\lightning$	

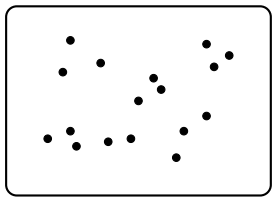
1.	$\gamma \leftrightarrow \neg B^{\ulcorner}\gamma^{\urcorner}$	Gödel-Carnap Fixed-Point Lemma
1.	$\gamma \leftrightarrow \neg B\gamma$	(Forget Gödel numbering)
2.	$\gamma \rightarrow \neg B\gamma$	Prop. Reasoning
3.	$B\gamma \rightarrow B\neg B\gamma$	Modal Reasoning
4.	$B\neg B\gamma \rightarrow \neg B\gamma$	Corr_N
5.	$B\gamma \rightarrow \neg B\gamma$	Prop. Reasoning
6.	$\neg B\gamma$	Prop. Reasoning
7.	$\neg B\gamma \rightarrow \gamma$	Prop. Reasoning
8.	γ	Prop. Reasoning
9.	$B\gamma$	Nec
	$\lightning$	

Buridan-Burge: $p \leftrightarrow \neg B_a p$

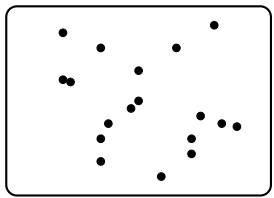
Anti-Expert: $p \leftrightarrow B_a \neg p$

Buridan-Burge: $p \leftrightarrow \neg B_a p$
$B_a(p \leftrightarrow \neg B_a p)$

Anti-Expert: $p \leftrightarrow B_a \neg p$
$B_a(p \leftrightarrow B_a \neg p)$



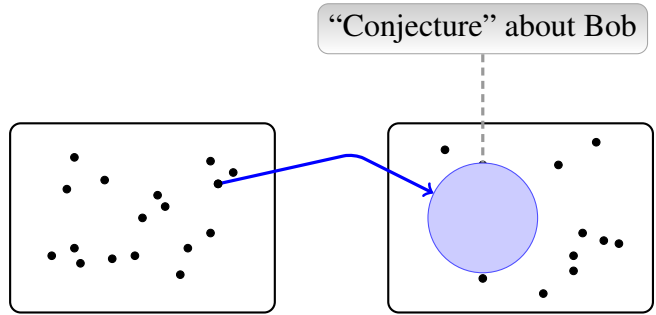
Ann's Possible Types

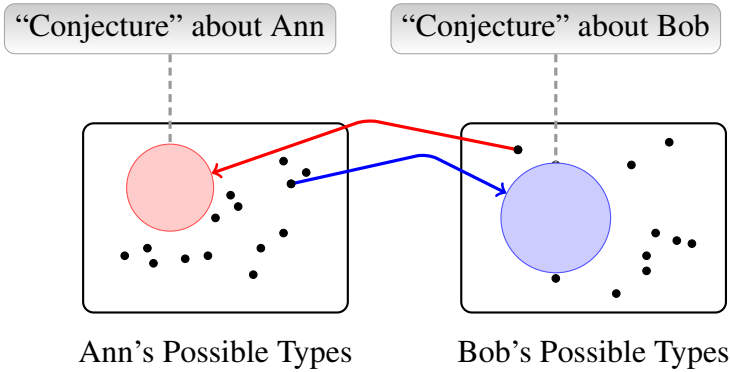


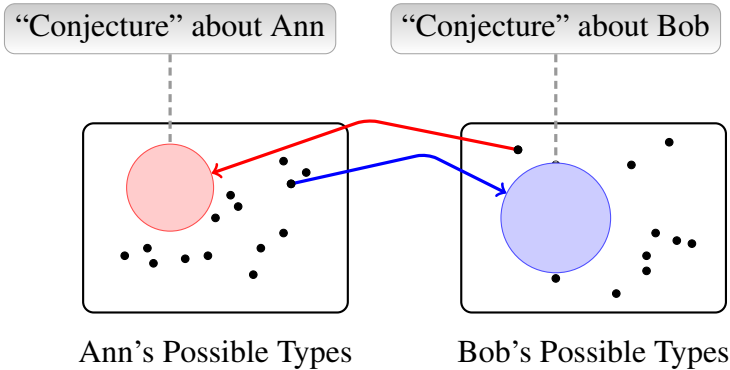
Bob's Possible Types

A (game-theoretic) **type** of a player summarizes everything the player knows privately at the beginning of the game which could affect his beliefs about payoffs in the game and about all other players' types.

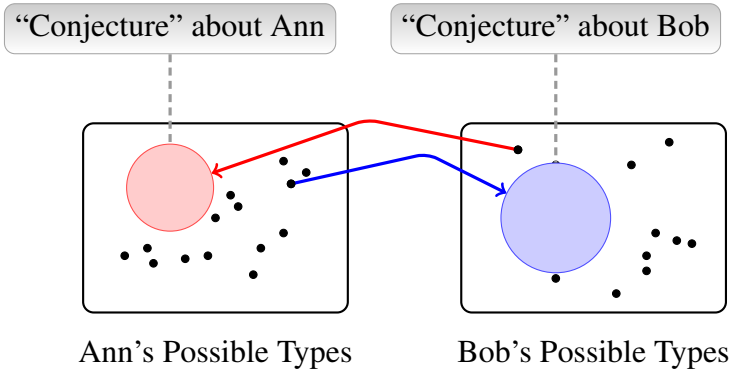
(Harsanyi argued that all uncertainty in a game can be equivalently modeled as uncertainty about payoff functions.)


“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...

Results

Language for i: A set $\mathcal{L}_i \subseteq \wp(T_{-i})$.

Results

Language for i: A set $\mathcal{L}_i \subseteq \wp(T_{-i})$.

Richness Property for $\mathcal{L}_i$: For all $X \in \mathcal{L}_i$, if $X \neq \emptyset$, then there is some type $t \in T_i$ of player i such that X describes t's conjecture about $-i$.

Results

Language for i: A set $\mathcal{L}_i \subseteq \wp(T_{-i})$.

Richness Property for $\mathcal{L}_i$: For all $X \in \mathcal{L}_i$, if $X \neq \emptyset$, then there is some type $t \in T_i$ of player i such that X describes t's conjecture about $-i$.

- $\mathcal{L}_i$ **can't** be the set of all non-empty subsets. (Brandenburger, 2003)
- $\mathcal{L}_i$ **can't** be the set of sets that are definable in first-order logic. (Brandenburger and Keisler, 2006)
- $\mathcal{L}_i$ **can't** be the set of sets definable in a propositional modal logic with an assumption modality. (Brandenburger and Keisler, 2006)
- $\mathcal{L}_i$ **can** be the set of compact subsets of a topological space (Mariotti, Meier and Piccione, 2005)
- ...

The BK Paradox

> *Ann believes that Bob's assumption is that Ann believes that Bob's assumption is wrong.*

Does Ann believe that Bob's assumption is wrong?

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

The BK Paradox

Ann believes that the strongest proposition that Bob believes is that Ann believes that the strongest proposition that Bob believes is false.

Does Ann believe that the strongest proposition that Bob believes is false?

The BK Paradox

Ann believes that the strongest proposition that Bob believes is that Ann believes that the strongest proposition that Bob believes is false.

Does Ann believe that the strongest proposition that Bob believes is false?

The BK Paradox

Ann believes that the strangest proposition that Bob believes is that Ann believes that the strangest proposition that Bob believes is false.

Does Ann believe that the strangest proposition that Bob believes is false?

The BK Paradox

Ann believes that the most interesting proposition that Bob believes is that Ann believes that the most interesting proposition that Bob believes is false.

Does Ann believe that the most interesting proposition that Bob believes is false?

The BK Paradox

Goal: A modal logic of belief (for many agents) with formulas that may contain definition descriptions of propositions (which may or may not denote).

Modal logics of belief...

$$\langle W, \{R_i\}_{i \in \mathcal{A}}, V \rangle$$

States/possible worlds: $W \neq \emptyset$

Quasi-partitions: $R_i \subseteq W \times W$ is serial, transitive and Euclidean

Belief operators: $\mathcal{M}, w \models B_i \varphi$ iff for all v, if $w \; R_i \; v$, then $\mathcal{M}, v \models \varphi$.

Modal logics of belief...

$$\langle W, \{R_i\}_{i \in \mathcal{A}}, V \rangle$$

States/possible worlds: $W \neq \emptyset$

Quasi-partitions: $R_i \subseteq W \times W$ is serial, transitive and Euclidean

Belief operators: $\mathcal{M}, w \models B_i\varphi$ iff for all v, if $w \; R_i \; v$, then $\mathcal{M}, v \models \varphi$.

$$\mathcal{M}, w \models B_i\varphi \text{ iff } R_i(w) \subseteq [\![\varphi]\!]_{\mathcal{M}}$$

Modal logics of belief...

$$\langle W, \{R_i\}_{i \in \mathcal{A}}, V \rangle$$

States/possible worlds: $W \neq \emptyset$

Quasi-partitions: $R_i \subseteq W \times W$ is serial, transitive and Euclidean

Belief operators: $\mathcal{M}, w \models B_i\varphi$ iff for all v, if $w \; R_i \; v$, then $\mathcal{M}, v \models \varphi$.

$$\mathcal{M}, w \models B_i\varphi \text{ iff } R_i(w) \subseteq [\![\varphi]\!]_{\mathcal{M}}$$

$\{v \mid w \; R_i \; v\}$ $\qquad$ $\{v \mid \mathcal{M}, w \models \varphi\}$

Assumption

The BK Paradox

> *Ann believes that Bob's assumption is that Ann believes that Bob's assumption is wrong.*

Does Ann believe that Bob's assumption is wrong?

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

Concluding Remarks: Assumption

Two players: Ann (a) and Bob (b)

Standard model of beliefs: $\langle W, \{R_i\}_{i \in \mathcal{A}}, V \rangle$, where R_i is a quasi-partition, $W \neq \emptyset$ and V a valuation function.

$w \models B_i \varphi$ iff $R_i(w) \subseteq [\![\varphi]\!]_{\mathcal{M}} = \{w \mid \mathcal{M}, w \models \varphi\}$

Concluding Remarks: Assumption

Two players: Ann (a) and Bob (b)

Standard model of beliefs: $\langle W, \{R_i\}_{i \in \mathcal{A}}, V \rangle$, where R_i is a quasi-partition, $W \neq \emptyset$ and V a valuation function.

$w \models B_i \varphi$ iff $R_i(w) \subseteq [\![\varphi]\!]_{\mathcal{M}} = \{w \mid \mathcal{M}, w \models \varphi\}$

Assumption operator $\boxplus_i \varphi$: "i's assumption is φ".

(window modality, "all the agent knows")

$w \models \boxplus_i \varphi$ iff $R_i(w) = [\![\varphi]\!]_{\mathcal{M}} = \{w \mid \mathcal{M}, w \models \varphi\}$

Concluding Remarks: Assumption

Ann believes that Bob's assumption is that Ann believes that Bob's assumption is wrong.

$$\Box_a \boxplus_b D \wedge \Diamond_a \top \quad \Longrightarrow \quad D \leftrightarrow \neg D$$

Definite descriptions of propositions

Suppose that At is a set of atomic propositions, $\mathcal{A}$ is a set of agents, and Lab is a set of "labels" for propositions. The language $\mathcal{L}$:

$$\varphi \quad ::= \quad p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid B_i\varphi \mid \gamma \text{ is } \varphi$$

where $p \in \mathsf{At}$, $i \in \mathcal{A}$, and $\gamma \in \mathsf{Lab}$.

- $B_i\varphi$: "agent i believes that φ"
- γ is φ: "the definite description γ denotes the proposition expressed by φ" (or "the γ-proposition is φ").

Models

$$\mathcal{M} = \langle W, \{R_i\}_{i \in \mathcal{A}}, \{N_\gamma\}_{\gamma \in \mathsf{Lab}}, V \rangle$$

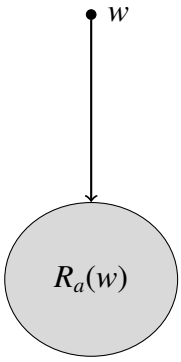
- W is a nonempty set (the *set of worlds*)
- for each $i \in \mathcal{A}$, R_i is a serial, transitive and Euclidean relation on W (a quasi-partition)
- for each $\gamma \in \mathsf{Lab}$, $N_\gamma \colon W \dashrightarrow \wp(W)$ is a partial function (the *denotation function*)
- $V \colon \mathsf{At} \to \wp(W)$ (the *valuation function*)

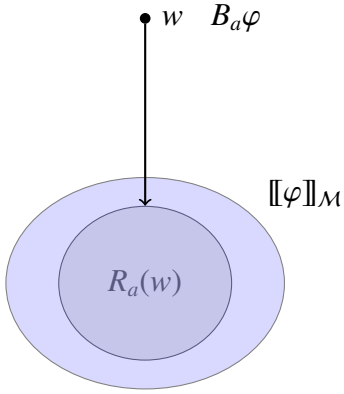
The truth of a formula $\varphi \in \mathcal{L}$ at a world w in a model $\mathcal{M}$ (notation: $\mathcal{M}, w \models \varphi$), and the set $[\![\varphi]\!]_{\mathcal{M}}$ of worlds at which φ is true in $\mathcal{M}$, is defined by recursion as follows:

- $\mathcal{M}, w \models p$ iff $w \in V(p)$, where $p \in \mathsf{At}$
- $\mathcal{M}, w \models \neg\varphi$ iff $\mathcal{M}, w \not\models \varphi$
- $\mathcal{M}, w \models \varphi \wedge \psi$ iff $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models B_i\varphi$ iff $R_i(w) \subseteq [\![\varphi]\!]_{\mathcal{M}}$
- $\mathcal{M}, w \models \gamma$ is φ iff $D_\gamma(w)$ is defined and $D_\gamma(w) = [\![\varphi]\!]_{\mathcal{M}}$

The truth of a formula $\varphi \in \mathcal{L}$ at a world w in a model $\mathcal{M}$ (notation: $\mathcal{M}, w \models \varphi$), and the set $[\![\varphi]\!]_{\mathcal{M}}$ of worlds at which φ is true in $\mathcal{M}$, is defined by recursion as follows:

- $\mathcal{M}, w \models B_i^{re}(\mathsf{T}(\gamma))$ iff $D_\gamma(w)$ is defined and $R_i(w) \subseteq N_\gamma(w)$
- $\mathcal{M}, w \models B_i^{re}(\mathsf{F}(\gamma))$ iff $D_\gamma(w)$ is defined and $R_i(w) \subseteq W \setminus N_\gamma(w)$
- $\mathcal{M}, w \models B_i^{dicto}(\mathsf{T}(\gamma))$ iff for all $v \in R_i(w)$, $N_\gamma(v)$ is defined and $R_i(w) \subseteq N_\gamma(v)$
- $\mathcal{M}, w \models B_i^{dicto}(\mathsf{F}(\gamma))$ iff for all $v \in R_i(w)$, $N_\gamma(v)$ is defined and $R_i(w) \subseteq W \setminus N_\gamma(v)$.


w
$R_a(w)$

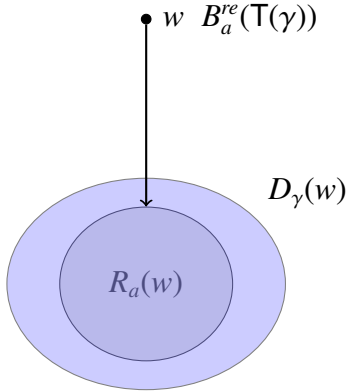

w
$B_a\varphi$
$[\![\varphi]\!]_{\mathcal{M}}$
$R_a(w)$

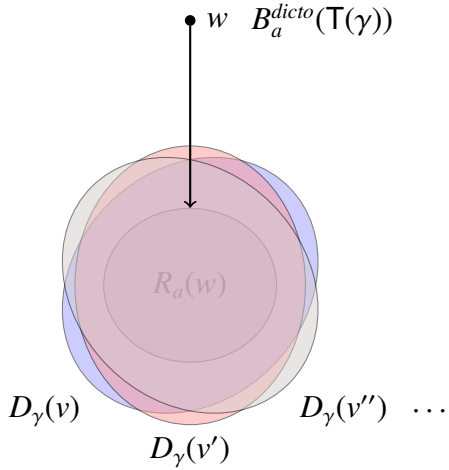
$$B_a(\gamma \text{ is } \varphi)$$

$$B_a(\gamma)$$

$$B_a(\gamma \text{ is } \varphi)$$

~~$B_a(\gamma)$~~ $\qquad\qquad$ $B_a(\mathsf{T}(\gamma)),\ \ B_a(\mathsf{F}(\gamma))$


w
$B_a^{re}(\mathsf{T}(\gamma))$
$D_\gamma(w)$
$R_a(w)$


w
$B_a^{dicto}(\mathsf{T}(\gamma))$
$R_a(w)$
$D_\gamma(v)$
$D_\gamma(v')$
$D_\gamma(v'')$
$\cdots$

Suppose that At is a set of atomic propositions, $\mathcal{A}$ is a set of agents, and Lab is a set of labels for propositions. The language $\mathcal{L}$:

$$\begin{aligned}\varphi \quad ::= \quad & p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid B_i\varphi \mid \gamma \text{ is } \varphi \\ & B_i^{dicto}\mathsf{T}(\gamma) \mid B_i^{dicto}\mathsf{F}(\gamma) \mid B_i^{re}\mathsf{T}(\gamma) \mid B_i^{re}\mathsf{F}(\gamma)\end{aligned}$$

where $p \in \mathsf{At}$, $i \in \mathcal{A}$, and $\gamma \in \mathsf{Lab}$.

- *de re belief*: $B_i^{re}\mathsf{T}(\gamma)$ ($B_i^{re}\mathsf{F}(\gamma)$): "$i$ believes of the proposition denoted by γ that it is correct (wrong)."
- *de dicto belief*: $B_i^{dicto}\mathsf{T}(\gamma)$ ($B_i^{dicto}\mathsf{F}(\gamma)$): "$i$ believes that γ is correct (wrong), *whatever proposition γ turns out to denote*."

Suppose that At is a set of atomic propositions, $\mathcal{A}$ is a set of agents, and Lab is a set of labels for propositions. The language $\mathcal{L}$:

$$\begin{aligned}\varphi \quad ::= \quad & p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid B_i\varphi \mid \gamma \text{ is } \varphi \\ & B_i^{dicto}\mathsf{T}(\gamma) \mid B_i^{dicto}\mathsf{F}(\gamma) \mid B_i^{re}\mathsf{T}(\gamma) \mid B_i^{re}\mathsf{F}(\gamma)\end{aligned}$$

where $p \in \mathsf{At}$, $i \in \mathcal{A}$, and $\gamma \in \mathsf{Lab}$.

- $\mathsf{T}(\gamma) \leftrightarrow \gamma$ and $\mathsf{F}(\gamma) \leftrightarrow \neg\gamma$ are not formulas (γ may not denote).
- $\mathsf{T}(\gamma)$ and $\mathsf{F}(\gamma)$ are always proceeded by expressions B_i^{re} or B_i^{dicto}.
- γ is $\neg\gamma$ is not well-formed. (no liar-sentences)
- γ is $\neg B_i^{re}\mathsf{T}(\gamma)$ is a well-formed formula. (there is some self-reference)

The truth of a formula $\varphi \in \mathcal{L}$ at a world w in a model $\mathcal{M}$ (notation: $\mathcal{M}, w \models \varphi$), and the set $[\![\varphi]\!]_{\mathcal{M}}$ of worlds at which φ is true in $\mathcal{M}$, is defined by recursion as follows:

- $\mathcal{M}, w \models B_i^{re}(\mathsf{T}(\gamma))$ iff $D_\gamma(w)$ is defined and $R_i(w) \subseteq N_\gamma(w)$
- $\mathcal{M}, w \models B_i^{re}(\mathsf{F}(\gamma))$ iff $D_\gamma(w)$ is defined and $R_i(w) \subseteq W \setminus N_\gamma(w)$
- $\mathcal{M}, w \models B_i^{dicto}(\mathsf{T}(\gamma))$ iff for all $v \in R_i(w)$, $N_\gamma(v)$ is defined and $R_i(w) \subseteq N_\gamma(v)$
- $\mathcal{M}, w \models B_i^{dicto}(\mathsf{F}(\gamma))$ iff for all $v \in R_i(w)$, $N_\gamma(v)$ is defined and $R_i(w) \subseteq W \setminus N_\gamma(v)$.

The BK Paradox

Ann believes that the strongest proposition that Bob believes is that Ann believes that the strongest proposition that Bob believes is wrong.

Does Ann believe that the strongest proposition that Bob believes is wrong?

The BK Paradox

Ann believes that the γ-proposition is that Ann believes that the γ-proposition is wrong.

- $B_a(\gamma \text{ is } B_a^{dicto}\mathsf{F}(\gamma))$
- $B_a(\gamma \text{ is } B_a^{re}\mathsf{F}(\gamma))$

Logic

Individual beliefs are consistent and deductively closed.

(D)	$B_i\varphi \to \neg B_i \neg \varphi$
(K)	$B_i(\varphi \to \psi) \to (B_i\varphi \to B_i\psi)$
(Nec)	if φ is a theorem, so is $B_i\varphi$

Everyone is correct about their own beliefs.

(CorP)	$B_i B_i \varphi \to B_i \varphi$
(CorN)	$B_i \neg B_i \varphi \to \neg B_i \varphi$

Everyone is perfectly introspective.

(PI)	$B_i\varphi \to B_i B_i \varphi$
(NI)	$\neg B_i \varphi \to B_i \neg B_i \varphi$

Logic

Note that $B_i B_i^{re}\mathsf{T}(\gamma) \rightarrow B_i^{re}\mathsf{T}(\gamma)$ is *not* an instance of (*CorP*). Similarly, $B_i^{re}\mathsf{T}(\gamma) \rightarrow B_i B_i^{re}\mathsf{T}(\gamma)$ is *not* an instance of (*PI*).

(Cor_P)	$B_i\chi \rightarrow \chi$
(Cor_N)	$B_i\neg\chi \rightarrow \neg\chi$
	For each $\chi \in \{B_i^{dicto}\mathsf{T}(\gamma), B_i^{dicto}\mathsf{F}(\gamma), B_i^{re}\mathsf{T}(\gamma), B_i^{re}\mathsf{F}(\gamma)\}$
(I_P)	$\chi \rightarrow B_i\chi$
(I_N)	$\neg\chi \rightarrow B_i\neg\chi$
	For each $\chi \in \{B_i^{dicto}\mathsf{T}(\gamma), B_i^{dicto}\mathsf{F}(\gamma), B_i^{re}\mathsf{T}(\gamma), B_i^{re}\mathsf{F}(\gamma)\}$

Logic

Substitution axioms:

$$(S1^{dicto}) \qquad B_i(\gamma \text{ is } \varphi) \rightarrow (B_i^{dicto}\mathsf{T}(\gamma) \leftrightarrow B_i\varphi)$$

$$(S2^{dicto}) \qquad B_i(\gamma \text{ is } \varphi) \rightarrow (B_i^{dicto}\mathsf{F}(\gamma) \leftrightarrow B_i\neg\varphi)$$

$$(S1^{re}) \qquad (\gamma \text{ is } \varphi) \rightarrow (B_i^{re}\mathsf{T}(\gamma) \leftrightarrow B_i\varphi)$$

$$(S2^{re}) \qquad (\gamma \text{ is } \varphi) \rightarrow (B_i^{re}\mathsf{F}(\gamma) \leftrightarrow B_i\neg\varphi)$$

Proposition The set $\{\gamma \text{ is } \neg B_i^{re}\mathsf{T}(\gamma)\}$ is inconsistent in any propositional modal logic containing S1^{re}, Cor_N and I_N.

Proposition The set $\{\gamma \text{ is } \neg B_i^{re}\mathsf{T}(\gamma)\}$ is inconsistent in any propositional modal logic containing S1^{re}, Cor_N and I_N.

1. γ is $\neg B_i^{re}\mathsf{T}(\gamma)$ (assumption)

Proposition The set $\{\gamma \text{ is } \neg B_i^{re}\mathsf{T}(\gamma)\}$ is inconsistent in any propositional modal logic containing S1^{re}, Cor_N and I_N.

1. γ is $\neg B_i^{re}\mathsf{T}(\gamma)$ (assumption)
2. γ is $\neg B_i^{re}\mathsf{T}(\gamma) \rightarrow ((B_i^{re}\mathsf{T}(\gamma) \leftrightarrow B_i \neg B_i^{re}\mathsf{T}(\gamma))$ (S1^{re})

Proposition The set $\{\gamma \text{ is } \neg B_i^{re}\mathsf{T}(\gamma)\}$ is inconsistent in any propositional modal logic containing S1^{re}, Cor_N and I_N.

1.	γ is $\neg B_i^{re}\mathsf{T}(\gamma)$	(assumption)
2.	γ is $\neg B_i^{re}\mathsf{T}(\gamma) \rightarrow ((B_i^{re}\mathsf{T}(\gamma) \leftrightarrow B_i \neg B_i^{re}\mathsf{T}(\gamma))$	(S1^{re})
3.	$B_i^{re}\mathsf{T}(\gamma) \leftrightarrow B_i(\neg B_i^{re}\mathsf{T}(\gamma))$	(Prop Reas, 1, 2)

Proposition The set $\{\gamma \text{ is } \neg B_i^{re}\mathsf{T}(\gamma)\}$ is inconsistent in any propositional modal logic containing S1^{re}, Cor_N and I_N.

1.	γ is $\neg B_i^{re}\mathsf{T}(\gamma)$	(assumption)
2.	γ is $\neg B_i^{re}\mathsf{T}(\gamma) \rightarrow ((B_i^{re}\mathsf{T}(\gamma) \leftrightarrow B_i \neg B_i^{re}\mathsf{T}(\gamma))$	(S1^{re})
3.	$B_i^{re}\mathsf{T}(\gamma) \leftrightarrow B_i(\neg B_i^{re}\mathsf{T}(\gamma))$	(Prop Reas, 1, 2)
4.	$B_i \neg B_i^{re}\mathsf{T}(\gamma) \rightarrow \neg B_i^{re}\mathsf{T}(\gamma)$	(Cor_N)

Proposition The set $\{\gamma \text{ is } \neg B_i^{re}\mathsf{T}(\gamma)\}$ is inconsistent in any propositional modal logic containing S1^{re}, Cor_N and I_N.

1.	γ is $\neg B_i^{re}\mathsf{T}(\gamma)$	(assumption)
2.	γ is $\neg B_i^{re}\mathsf{T}(\gamma) \rightarrow ((B_i^{re}\mathsf{T}(\gamma) \leftrightarrow B_i \neg B_i^{re}\mathsf{T}(\gamma))$	(S1^{re})
3.	$B_i^{re}\mathsf{T}(\gamma) \leftrightarrow B_i(\neg B_i^{re}\mathsf{T}(\gamma))$	(Prop Reas, 1, 2)
4.	$B_i \neg B_i^{re}\mathsf{T}(\gamma) \rightarrow \neg B_i^{re}\mathsf{T}(\gamma)$	(Cor_N)
5.	$B_i^{re}\mathsf{T}(\gamma) \rightarrow \neg B_i^{re}\mathsf{T}(\gamma)$	(Prop Reas, 3, 4)

Proposition The set $\{\gamma \text{ is } \neg B_i^{re}\mathsf{T}(\gamma)\}$ is inconsistent in any propositional modal logic containing S1^{re}, Cor_N and I_N.

1.	γ is $\neg B_i^{re}\mathsf{T}(\gamma)$	(assumption)
2.	γ is $\neg B_i^{re}\mathsf{T}(\gamma) \rightarrow ((B_i^{re}\mathsf{T}(\gamma) \leftrightarrow B_i \neg B_i^{re}\mathsf{T}(\gamma))$	(S1^{re})
3.	$B_i^{re}\mathsf{T}(\gamma) \leftrightarrow B_i(\neg B_i^{re}\mathsf{T}(\gamma))$	(Prop Reas, 1, 2)
4.	$B_i \neg B_i^{re}\mathsf{T}(\gamma) \rightarrow \neg B_i^{re}\mathsf{T}(\gamma)$	(Cor_N)
5.	$B_i^{re}\mathsf{T}(\gamma) \rightarrow \neg B_i^{re}\mathsf{T}(\gamma)$	(Prop Reas, 3, 4)
6.	$\neg B_i^{re}\mathsf{T}(\gamma)$	(Prop Reas, 5)

Proposition The set $\{\gamma \text{ is } \neg B_i^{re}\mathsf{T}(\gamma)\}$ is inconsistent in any propositional modal logic containing S1^{re}, Cor_N and I_N.

1.	γ is $\neg B_i^{re}\mathsf{T}(\gamma)$	(assumption)
2.	γ is $\neg B_i^{re}\mathsf{T}(\gamma) \rightarrow ((B_i^{re}\mathsf{T}(\gamma) \leftrightarrow B_i \neg B_i^{re}\mathsf{T}(\gamma))$	(S1^{re})
3.	$B_i^{re}\mathsf{T}(\gamma) \leftrightarrow B_i(\neg B_i^{re}\mathsf{T}(\gamma))$	(Prop Reas, 1, 2)
4.	$B_i \neg B_i^{re}\mathsf{T}(\gamma) \rightarrow \neg B_i^{re}\mathsf{T}(\gamma)$	(Cor_N)
5.	$B_i^{re}\mathsf{T}(\gamma) \rightarrow \neg B_i^{re}\mathsf{T}(\gamma)$	(Prop Reas, 3, 4)
6.	$\neg B_i^{re}\mathsf{T}(\gamma)$	(Prop Reas, 5)
7.	$\neg B_i^{re}\mathsf{T}(\gamma) \rightarrow B_i \neg B_i^{re}\mathsf{T}(\gamma)$	(I_N)

Proposition The set $\{\gamma \text{ is } \neg B_i^{re}\mathsf{T}(\gamma)\}$ is inconsistent in any propositional modal logic containing S1^{re}, Cor_N and I_N.

1.	γ is $\neg B_i^{re}\mathsf{T}(\gamma)$	(assumption)
2.	γ is $\neg B_i^{re}\mathsf{T}(\gamma) \rightarrow ((B_i^{re}\mathsf{T}(\gamma) \leftrightarrow B_i \neg B_i^{re}\mathsf{T}(\gamma))$	(S1^{re})
3.	$B_i^{re}\mathsf{T}(\gamma) \leftrightarrow B_i(\neg B_i^{re}\mathsf{T}(\gamma))$	(Prop Reas, 1, 2)
4.	$B_i \neg B_i^{re}\mathsf{T}(\gamma) \rightarrow \neg B_i^{re}\mathsf{T}(\gamma)$	(Cor_N)
5.	$B_i^{re}\mathsf{T}(\gamma) \rightarrow \neg B_i^{re}\mathsf{T}(\gamma)$	(Prop Reas, 3, 4)
6.	$\neg B_i^{re}\mathsf{T}(\gamma)$	(Prop Reas, 5)
7.	$\neg B_i^{re}\mathsf{T}(\gamma) \rightarrow B_i \neg B_i^{re}\mathsf{T}(\gamma)$	(I_N)
8.	$B_i \neg B_i^{re}\mathsf{T}(\gamma)$	(Prop Reas, 6, 7)

Proposition The set $\{\gamma \text{ is } \neg B_i^{re}\mathsf{T}(\gamma)\}$ is inconsistent in any propositional modal logic containing S1^{re}, Cor_N and I_N.

1.	γ is $\neg B_i^{re}\mathsf{T}(\gamma)$	(assumption)
2.	γ is $\neg B_i^{re}\mathsf{T}(\gamma) \rightarrow ((B_i^{re}\mathsf{T}(\gamma) \leftrightarrow B_i \neg B_i^{re}\mathsf{T}(\gamma))$	(S1^{re})
3.	$B_i^{re}\mathsf{T}(\gamma) \leftrightarrow B_i(\neg B_i^{re}\mathsf{T}(\gamma))$	(Prop Reas, 1, 2)
4.	$B_i \neg B_i^{re}\mathsf{T}(\gamma) \rightarrow \neg B_i^{re}\mathsf{T}(\gamma)$	(Cor_N)
5.	$B_i^{re}\mathsf{T}(\gamma) \rightarrow \neg B_i^{re}\mathsf{T}(\gamma)$	(Prop Reas, 3, 4)
6.	$\neg B_i^{re}\mathsf{T}(\gamma)$	(Prop Reas, 5)
7.	$\neg B_i^{re}\mathsf{T}(\gamma) \rightarrow B_i \neg B_i^{re}\mathsf{T}(\gamma)$	(I_N)
8.	$B_i \neg B_i^{re}\mathsf{T}(\gamma)$	(Prop Reas, 6, 7)
9.	$B_i^{re} C(\gamma)$	(Prop Reas, 8, 3)

Proposition The set $\{\gamma \text{ is } \neg B_i^{re}\mathsf{T}(\gamma)\}$ is inconsistent in any propositional modal logic containing S1^{re}, Cor_N and I_N.

1.	γ is $\neg B_i^{re}\mathsf{T}(\gamma)$	(assumption)
2.	γ is $\neg B_i^{re}\mathsf{T}(\gamma) \rightarrow ((B_i^{re}\mathsf{T}(\gamma) \leftrightarrow B_i \neg B_i^{re}\mathsf{T}(\gamma))$	(S1^{re})
3.	$B_i^{re}\mathsf{T}(\gamma) \leftrightarrow B_i(\neg B_i^{re}\mathsf{T}(\gamma))$	(Prop Reas, 1, 2)
4.	$B_i \neg B_i^{re}\mathsf{T}(\gamma) \rightarrow \neg B_i^{re}\mathsf{T}(\gamma)$	(Cor_N)
5.	$B_i^{re}\mathsf{T}(\gamma) \rightarrow \neg B_i^{re}\mathsf{T}(\gamma)$	(Prop Reas, 3, 4)
6.	$\neg B_i^{re}\mathsf{T}(\gamma)$	(Prop Reas, 5)
7.	$\neg B_i^{re}\mathsf{T}(\gamma) \rightarrow B_i \neg B_i^{re}\mathsf{T}(\gamma)$	(I_N)
8.	$B_i \neg B_i^{re}\mathsf{T}(\gamma)$	(Prop Reas, 6, 7)
9.	$B_i^{re} C(\gamma)$	(Prop Reas, 8, 3)
10.	Contradiction	(6, 9)

Proposition The set $\{\gamma \text{ is } B_i^{re}\mathsf{F}(\gamma)\}$ is inconsistent in any propositional modal logic containing S2^{re}, Cor and I.

Proposition The set $\{\gamma \text{ is } B_i^{re}\mathsf{F}(\gamma)\}$ is inconsistent in any propositional modal logic containing S2^{re}, Cor and I.

This does not imply the BK Paradox: γ is $B_i^{re}\mathsf{F}(\gamma)$ and $B_i(\gamma \text{ is } B_i^{re}\mathsf{F}(\gamma))$ are logically independent. What about $B_i(\gamma \text{ is } B_i^{dicto}\mathsf{F}(\gamma))$?

The BK Paradox

Ann believes that the strongest proposition that Bob believes is that Ann believes that the strongest proposition that Bob believes is false.

Does Ann believe that the strongest proposition that Bob believes is false?

The BK Paradox

Ann believes that the γ-proposition is that Ann believes that the γ-proposition is false.

Claim 1. $\{B_a(\gamma \text{ is } B_a^{dicto}\mathsf{F}(\gamma))\}$ is inconsistent in any modal logic containing K, Nec, Cor, I, $S2^{dicto}$

Claim 2. $\{B_a(\gamma \text{ is } B_a^{re}\mathsf{F}(\gamma))\}$ is inconsistent in any modal logic containing K, Nec, Cor, I, $S2^{re}$

The BK Paradox

Proposition. $\{B_a(\gamma \text{ is } B_a^{dicto}\mathsf{F}(\gamma))\}$ is inconsistent in any modal logic containing K, Nec, Cor_N, I_N, $S2^{dicto}$.

The BK Paradox

1.	$B_i(\gamma \text{ is } B_i^{dicto}\mathsf{F}(\gamma))$	(assumption)
2.	$B_i(\gamma \text{ is } B_i^{dicto}\mathsf{F}(\gamma)) \rightarrow (B_i^{dicto}\mathsf{F}(\gamma) \leftrightarrow B_i(\neg B_i^{dicto}\mathsf{F}(\gamma)))$	(S2^{dicto})
3.	$B_i^{dicto}\mathsf{F}(\gamma) \leftrightarrow B_i(\neg B_i^{dicto}\mathsf{F}(\gamma))$	(MP, 1, 2)
4.	$B_i(\neg B_i^{dicto}\mathsf{F}(\gamma)) \rightarrow \neg B_i^{dicto}\mathsf{F}(\gamma)$	(Cor_N)
5.	$B_i^{dicto}\mathsf{F}(\gamma) \rightarrow \neg B_i^{dicto}\mathsf{F}(\gamma)$	(Prop Reas, 3, 4)
6.	$\neg B_i^{dicto}\mathsf{F}(\gamma)$	(Prop Reas, 7)
7.	$\neg B_i^{dicto}\mathsf{F}(\gamma) \rightarrow B_i\neg B_i^{dicto}\mathsf{F}(\gamma)$	(I_N)
8.	$B_i\neg B_i^{dicto}\mathsf{F}(\gamma)$	(MP, 6, 7)
9.	$B_i^{dicto}\mathsf{F}(\gamma)$	(MP, 3, 8)
10.	Contradiction	(6, 9)

The BK Paradox

Proposition. $\{B_a(\gamma \text{ is } B_a^{re}\mathsf{F}(\gamma))\}$ is inconsistent in any modal logic containing K, Nec, Cor, I, $S2^{re}$.

1. $B_i(\gamma \text{ is } B_i^{re}\mathsf{F}(\gamma))$ (assumption)
2. $(\gamma \text{ is } B_i^{re}\mathsf{F}(\gamma)) \rightarrow$
   $(B_i^{re}\mathsf{F}(\gamma) \leftrightarrow B_i(\neg B_i^{re}\mathsf{F}(\gamma)))$ ($S2^{re}$)
3. $B_i(\gamma \text{ is } B_i^{re}\mathsf{F}(\gamma)) \rightarrow$
   $B_i(B_i^{re}\mathsf{F}(\gamma) \leftrightarrow B_i(\neg B_i^{re}\mathsf{F}(\gamma)))$ (Mon, 2)

   ⋮

22. Contradiction

Taking Stock

- Propositional modal logic with definition descriptions for propositions.
- On this analysis, the BK Paradox is **not** a paradox of *interactive* beliefs.
- The proof of the BK Paradox is similar to the proof of the Knower Paradox.

Interpretation shifts

$$\langle W, \{R_i\}_{i \in \mathcal{A}}, V \rangle$$

Interpretation shifts

$$\langle W, \{R_i\}_{i \in \mathcal{A}}, V \rangle$$

Propositional Valuation: $V : \mathsf{At} \to \wp(W)$

Interpretation shifts

$$\langle W, \{R_i\}_{i \in \mathcal{A}}, V \rangle$$

Propositional Valuation: $V : \mathsf{At} \to \wp(W)$

Ambiguity/Propositional Control: $V : \mathsf{At} \times \mathcal{A} \to \wp(W)$

i and j disagree about the interpretation of p when $V(p, i) \neq V(p, j)$

Interpretation shifts

$$\langle W, \{R_i\}_{i \in \mathcal{A}}, V \rangle$$

Propositional Valuation: $V : \mathsf{At} \to \wp(W)$

Ambiguity/Propositional Control: $V : \mathsf{At} \times \mathcal{A} \to \wp(W)$

i and j disagree about the interpretation of p when $V(p, i) \neq V(p, j)$

Interpretation shifts

- Actions: $[p := \varphi]\psi$

Interpretation shifts

$$\langle W, \{R_i\}_{i \in \mathcal{A}}, V \rangle$$

Propositional Valuation: $V : \mathsf{At} \to \wp(W)$

Ambiguity/Propositional Control: $V : \mathsf{At} \times \mathcal{A} \to \wp(W)$

i and j disagree about the interpretation of p when $V(p, i) \neq V(p, j)$

Interpretation shifts

- Actions: $[p := \varphi]\psi$
- Becoming aware (W becomes more "fine-grained")

Interpretation shifts

$$\langle W, \{R_i\}_{i \in \mathcal{A}}, V \rangle$$

Propositional Valuation: $V : \mathsf{At} \to \wp(W)$

Ambiguity/Propositional Control: $V : \mathsf{At} \times \mathcal{A} \to \wp(W)$

i and j disagree about the interpretation of p when $V(p, i) \neq V(p, j)$

Interpretation shifts

- Actions: $[p := \varphi]\psi$
- Becoming aware (W becomes more "fine-grained")
- $V : \mathsf{At} \times W \to \wp(W)$

Related Models

- Dynamic logics with factual change $[p := \varphi]\psi$
- Logics of beliefs with *ambiguity* (or propositional control)
- Second-order propositional modal logic: $B_a \exists p B_b \neg p$, $\exists p B_a B_b \neg p$
- FOIL (First-order intensional logic): “The king of Sweden could be taller than he is.”

Knower	BB/AE	BK
γ is $\neg B_i \mathsf{T}(\gamma)$	$p \leftrightarrow \neg B_i p$	$B_i(\gamma \text{ is } \neg B_i \mathsf{T}(\gamma))$
γ is $B_i \mathsf{F}(\gamma)$	$p \leftrightarrow B_i \neg p$	$B_i(\gamma \text{ is } B_i \mathsf{F}(\gamma))$

$$\gamma \text{ is } B_i\mathsf{F}(\gamma) \quad \lightning \qquad B_i(\gamma \text{ is } B_i\mathsf{F}(\gamma)) \quad \lightning$$

$$p \leftrightarrow B_i\neg p \qquad\qquad B_i(p \leftrightarrow B_i\neg p) \quad \lightning$$

Are definition descriptions essential?

$$B_a \exists p (B_b p \wedge \forall q (B_b q \wedge \Box (p \rightarrow q)) \wedge$$
$$(p \leftrightarrow B_a \exists r (B_b r \wedge \forall q (B_b q \wedge \Box (r \rightarrow q)) \wedge \neg r)$$

Are definition descriptions essential? (No)

$$B_a \exists p(B_b p \wedge \forall q(B_b q \wedge \Box(p \rightarrow q)) \wedge$$
$$(p \leftrightarrow B_a \exists r(B_b r \wedge \forall q(B_b q \wedge \Box(r \rightarrow q)) \wedge \neg r)$$

Concluding Remarks

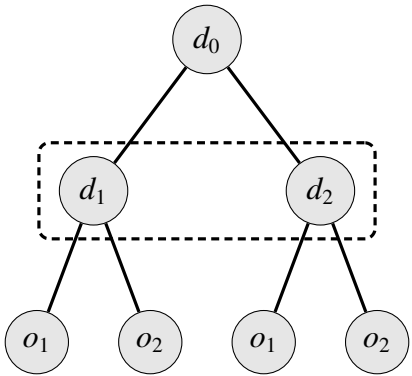
- ✓ Paradoxes of expected utility: St. Petersberg paradox, Pasadena game, The Two-envelop paradox
- ✓ Allais and Ellsberg paradox
- ✓ Newcomb's paradox and the psychopath button problem
- ✓ Puzzling games: the Prisoner's Dilemma and the Traveler's Dilemma
- ▸ The absent-minded driver problem
- ▸ Rubinstein's email game and the general's problem
- ✓ Backward induction and common knowledge of rationality
- ✓ The Brandenburger-Keisler paradox
- ▸ Framing in decision and game theory: language-dependent decisions and games, coordination problems and the theory of focal points.

Concluding Remarks

- ✓ Paradoxes of expected utility: St. Petersberg paradox, Pasadena game, The Two-envelop paradox
- ✓ Allais and Ellsberg paradox
- ✓ Newcomb's paradox and the psychopath button problem
- ✓ Puzzling games: the Prisoner's Dilemma and the Traveler's Dilemma
- ▸ The absent-minded driver problem
- ▸ Rubinstein's email game and the general's problem
- ✓ Backward induction and common knowledge of rationality
- ✓ The Brandenburger-Keisler paradox
- ✓ **Introduced some new people to one of the best scenes in modern movie history!**

The Absent-Minded Driver

Games of Imperfect Information



The Absent-Minded Driver

An individual is sitting late at night in a bar planning his midnight trip home. In order to get home he has to take the highway and get off at the second exit.

The Absent-Minded Driver

An individual is sitting late at night in a bar planning his midnight trip home. In order to get home he has to take the highway and get off at the second exit. Turning at the first exit leads into a disastrous area (payoff 0). Turning at the second exit yields the highest reward (payoff 4).

The Absent-Minded Driver

An individual is sitting late at night in a bar planning his midnight trip home. In order to get home he has to take the highway and get off at the second exit. Turning at the first exit leads into a disastrous area (payoff 0). Turning at the second exit yields the highest reward (payoff 4). If he continues beyond the second exit, he cannot go back and at the end of the highway he will find a motel where he can spend the night (payoff 1).

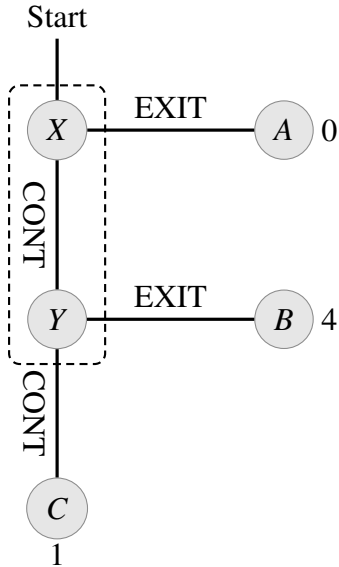
The Absent-Minded Driver

The driver is absentminded and is aware of this fact. At an intersection, he cannot tell whether it is the first or the second intersection and he cannot remember how many he has passed (one can make the situation more realistic by referring to the 17th intersection).

The Absent-Minded Driver

The driver is absentminded and is aware of this fact. At an intersection, he cannot tell whether it is the first or the second intersection and he cannot remember how many he has passed (one can make the situation more realistic by referring to the 17th intersection). While sitting at the bar, all he can do is to decide whether or not to exit at an intersection. (pg. 7)

M. Piccione and A. Rubinstein. *On the Interpretation of Decision Problems with Imperfect Recall*. Games and Econ Behavior, 20, pgs. 3- 24, 1997.


Start
X
EXIT
A
0
CONT
Y
EXIT
B
4
CONT
C
1

Planning stage: While planning his trip home at the bar, the decision maker is faced with a choice between "Continue; Continue" and "Exit". Since he cannot distinguish between the two intersections, he cannot plan to "Exit" at the second intersection (he must plan the same behavior at both X and Y). Since "Exit" will lead to the worst outcome (with a payoff of 0), the optimal strategy is "Continue; Continue" with a guaranteed payoff of 1.

Action stage: When arriving at an intersection, the decision maker is faced with a local choice of either “Exit” or “Continue” (possibly followed by another decision). Now the decision maker knows that since he committed to the plan of choosing “Continue” at each intersection, it is possible that he is at the second intersection. Indeed, the decision maker concludes that he is at the first intersection with probability 1/2. But then, his expected payoff for “Exit” is 2, which is greater than the payoff guaranteed by following the strategy he previously committed to. Thus, he chooses to “Exit”.

Have we captured *strategic reasoning*?

Strategic reasoning vs. Bayesian rationality

- Normal form vs. Extensive Form: Should the analysis take place on the tree or the matrix? (plans vs. strategies)
- There is an important different between what I would believe given E is true and what I believe after *learning* E
- What should I assume about my opponents?
- What is the role of *higher-order beliefs*? (Common knowledge, common belief)
- Framing issues/language in game theory
- ...

"...[W]e cannot expect game and economic theory to be descriptive in the same sense that physics or astronomy are. Rationality is only one of several factors affecting human behavior; no theory based on this one factor alone can be expected to yield reliable predictions.

In fact, I find it somewhat surprising that our disciplines have any relation at all to real behavior. (I hope that most readers will agree that there is indeed such a relation, that we do gain some insight into the behavior of *Homo sapiens* by studying *Homo rationalis*.)"

R. Aumann. *What is game theory trying to accomplish?*. Frontiers of Economics, 1985.

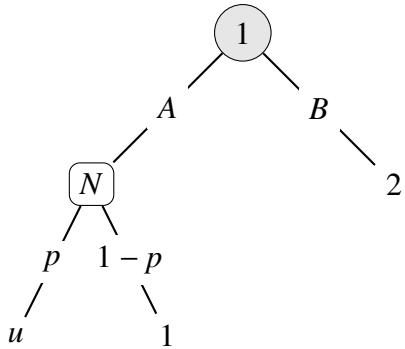
Can a player assign subjective probabilities to strategies under the control of other players who have their own objectives?

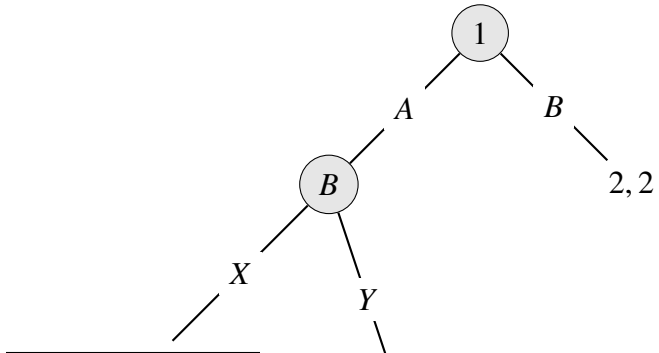
M. Mariotti. *Is Bayesian Rationality Compatible with Strategic Rationality?*. The Economic Journal, 105: 432, pgs. 1099 - 1109, 1995.

M. Mariotti. *Decisions in games: why there should be a special exemption from Bayesian rationality*. Journal of Economic Methodology, 4: 1, pgs. 43 - 60, 1997.

P. Hammond. *Expected Utility in Non-Cooperative Game Theory*. in *Handbook of Utility Theory*, 2004.

Games as consequences: “A decision maker prefers to be player i in game G_1 to being player j in game G_2”


1
A
B
N
2
p
1 − p
u
1



	L	R
U	$(1, 7)$	$(0, 0)$
D	$(0, 0)$	$(3, 3)$

Can the decision problem be *separated* from the game situation?

Can the decision problem be *separated* from the game situation?

Are strategies merely neutral access routes to consequences?

E. McClennen. *Rational choice in the context of ideal games*. in *Knowledge, Belief and Strategic Interaction*, pgs. 47-60, 1992.

utility must be measured *in the context of the game itself*.

I. Gilboa and D. Schmeidler. *A Derivation of Expected Utility Maximization in the Context of a Game*. Games and Economic Behavior, 44, pgs. 184 - 194, 2003.

The following two outcomes are not equivalent:

- “I get $90”
- “I get $90 and choose to leave $10 to my opponent”

The following two outcomes are not equivalent:

- “I get $10 and player one gets $90, and this was decided by Nature”
- “I get $10, player one gets $90 and this was decided by Player one”.

Players need two theories:

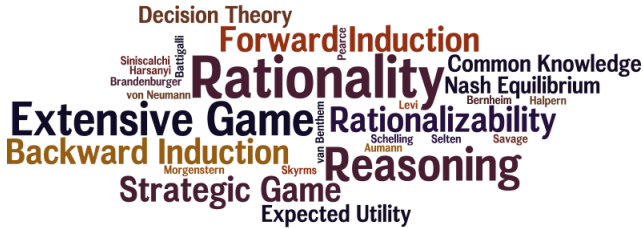
1. A theory to guide their decisions.
2. A theory to predict the behavior of their opponents.

“Game theory is decision theory about special decision makers, namely about decision makers who theorize decision-theoretically about the other persons figuring in their decision situations.” (Spohn, “How to make sense of Game Theory”)

"Rationality has a clear interpretation in individual decision making, but it does not transfer comfortably to interactive decisions, because interactive decision makers cannot maximize expected utility without strong assumptions about how the other participant(s) will behave. In game theory, common knowledge and rationality assumptions have therefore been introduced, but under these assumptions, rationality does not appear to be characteristic of social interaction in general." (pg. 152)

A. Colman. *Cooperation, psychological game theory, and limitations of rationality in social interaction*. Behavioral and Brain Sciences, 26, pgs. 139 - 198, 2003.

Thank you!


Decision Theory
Forward Induction
Siniscalchi
Harsanyi
Battigalli
Pearce
Rationality
Common Knowledge
Brandenburger
Nash Equilibrium
von Neumann
Bernheim
Halpern
Levi
Extensive Game
Rationalizability
Schelling
Selten
Savage
Aumann
van Benthem
Backward Induction
Morgenstern
Skyrms
Reasoning
Strategic Game
Expected Utility