Interpreting mistakes in games: From beliefs about mistakes to mistaken beliefs

Eric Pacuit

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

May 25, 2016

Plan

- Introductory remarks: Rationalizing *mistakes*
- Games and Game Models
- Backward and Forward Induction Reasoning
- Concluding remarks

(Joint with Aleks Knoks)

Just Enough Game Theory

A **game** is a mathematical model of a social interaction that includes

- the players (N);
- the actions (strategies) the players *can* take (for $i \in N$, $S_i \neq \emptyset$);
- the players' interests (for $i \in N$, $u_i : \Pi_{i\in N} S_i \rightarrow \mathbb{R}$); and
- the "structure" of the decision problem.

Just Enough Game Theory

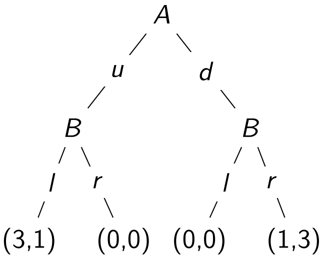
A **game** is a mathematical model of a social interaction that includes

- the players (N);
- the actions (strategies) the players *can* take (for $i \in N$, $S_i \neq \emptyset$);
- the players' interests (for $i \in N$, $u_i : \Pi_{i\in N} S_i \rightarrow \mathbb{R}$); and
- the "structure" of the decision problem.

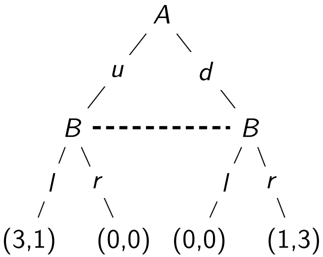
It does not specify the actions that the players do take.

		B	
		l	r
A	u	3, 1	0, 0
	d	0, 0	1, 3

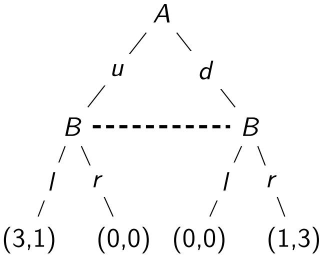
		B	
		l	r
A	u	3, 1	0, 0
	d	0, 0	1, 3

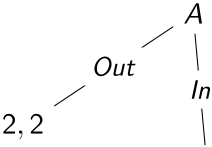


		B	
		l	r
A	u	3, 1	0, 0
	d	0, 0	1, 3



		B	
		l	r
A	u	3, 1	0, 0
	d	0, 0	1, 3





		B	
		l	r
A	u	3, 1	0, 0
	d	0, 0	1, 3

Just Enough Game Theory: Solution Concepts

A **solution concept** is a systematic description of the outcomes that may emerge in a family of games.

This is the starting point for most of game theory and includes many variants: Nash equilibrium (and refinements), backwards induction, or iterated dominance of various kinds.

They are usually thought of as the embodiment of “rational behavior” in some way and used to analyze game situations.

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**.

Models of Games

Suppose that G is a game.

- Outcomes of the game: $S = \Pi_{i \in N} S_i$
- 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)
- **s** is a function $\mathbf{s} : W \to \Pi_{i \in N} S_i$

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)
- **s** is a function $\mathbf{s} : W \to \Pi_{i \in N} S_i$
- For $s_i \in S_i$, $[s_i] = \{w \mid \mathbf{s}(w) = s_i\}$, 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 $\mathbf{s} : W \to \Pi_{i \in N} S_i$
- For $s_i \in S_i$, $[s_i] = \{w \mid \mathbf{s}(w) = s_i\}$, if $X \subseteq S$, $[X] = \bigcup_{s \in X}[s]$.
- **ex ante beliefs**: For each $i \in N$, $P_i \in \Delta(W)$. 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*: $P_{i,w}(\cdot) = P_i(\cdot \mid K_i)$, $\quad K_i \subseteq [\mathbf{s}_i(w)]$
- ...given all player i *fully believes*: $P_{i,w}(\cdot) = P_i(\cdot \mid B_i)$, $\quad B_i \subseteq [\mathbf{s}_i(w)]$

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

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

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*: $P_{i,w}(\cdot) = P_i(\cdot \mid B_i)$, $\quad B_i \subseteq [\mathbf{s}_i(w)]$

Expected utility: Given $w \in W$,

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

Belief Revision in Games

Exactly how a player revises her beliefs during the game depends, in part, on how she interprets the observed moves of her opponents.

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".

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 $\{\mathcal{M}_1^G, \mathcal{M}_2^G, \ldots\}$.

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.

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

To represent different *explanations* (i.e., beliefs) for the same strategy choice, we would need a set of models $\{\mathcal{M}_1^G, \mathcal{M}_2^G, \ldots\}$.

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.

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

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.

Richer models of games

1. A partition $\approx_i$ representing the different "**types**" of player i: $w \approx_i v$ means that w and v are subjectively indistinguishable to player i (i's beliefs, knowledge, and conditional beliefs are the same in both states).

2. A pseudo-partition R_i (serial, transitive and Euclidean relation) representing a player i's **working hypotheses** (full beliefs?, serious possibilities?,...).

3. Player i's **belief revision policy** described in terms of i's conditional beliefs.

Richer models of games

1. A partition $\approx_i$ representing the different "**types**" of player i: $w \approx_i v$ means that w and v are subjectively indistinguishable to player i (i's beliefs, knowledge, and conditional beliefs are the same in both states).

2. A pseudo-partition R_i (serial, transitive and Euclidean relation) representing a player i's **working hypotheses** (full beliefs?, serious possibilities?,...).

3. Player i's **belief revision policy** described in terms of i's conditional beliefs.

This can all be represented by a single relation $\preceq_i \subseteq W \times W$

Richer models of games

$\mathcal{M}^G = \langle W, \{\preceq_i, P_i\}_{i\in N}, \mathbf{s}\rangle$, where W, P_i and $\mathbf{s}$ are as before and $\preceq_i$ is a reflexive, transitive and locally-connected relation.

1. $w \approx_i v$ iff $w \preceq_i v$ or $v \preceq_i w$. Let $[w]_{\approx_i} = \{v \mid w \approx_i v\}$
2. $w\ R_i\ v$ iff $v \in Max_{\preceq_i}([w]_{\approx_i})$
3. $B_{i,w}(F) = Max_{\preceq_i}(F \cap [w]_{\approx_i})$

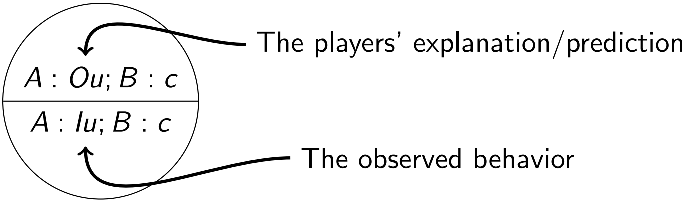
   $P_{i,w}(E \mid F) = P_i(E \mid B_{i,w}(F))$

Our Model

$$\mathcal{M}_G = \langle W, \{(\beta_i, \sigma_i)\}_{i \in N}, \{\succeq_i\}_{i \in N}, \{P_i\}_{i \in N} \rangle$$

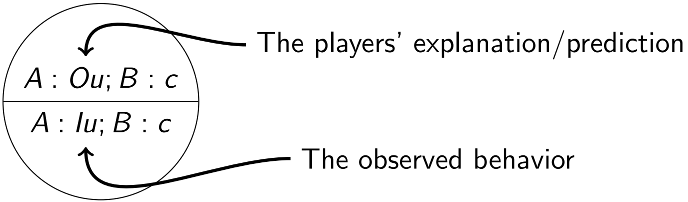
Our Model

$$\mathcal{M}_G = \langle W, \{(\beta_i, \sigma_i)\}_{i \in N}, \{\succeq_i\}_{i \in N}, \{P_i\}_{i \in N} \rangle$$



Our Model

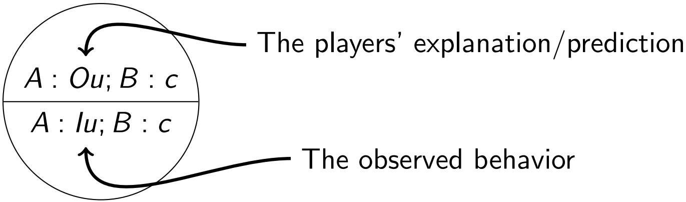
$$\mathcal{M}_G = \langle W, \{(\beta_i, \sigma_i)\}_{i \in N}, \{\succeq_i\}_{i \in N}, \{P_i\}_{i \in N} \rangle$$



- two functions: β and σ

Our Model

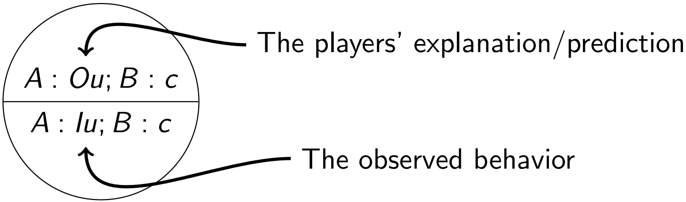
$$\mathcal{M}_G = \langle W, \{(\beta_i, \sigma_i)\}_{i \in N}, \{\succeq_i\}_{i \in N}, \{P_i\}_{i \in N} \rangle$$



- two functions: β and σ
- states represent *ex interim* stages of the game

Our Model

$$\mathcal{M}_G = \langle W, \{(\beta_i, \sigma_i)\}_{i \in N}, \{\succeq_i\}_{i \in N}, \{P_i\}_{i \in N} \rangle$$

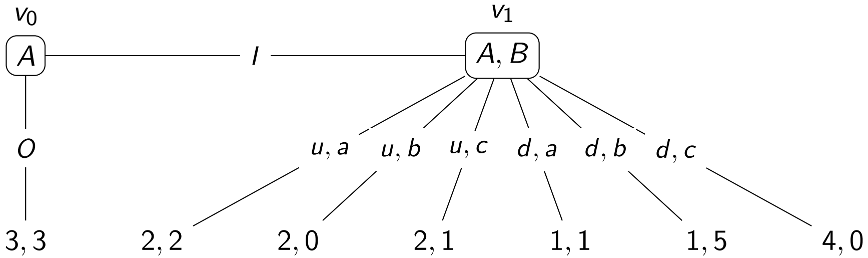


- two functions: β and σ
- states represent *ex interim* stages of the game
- what is a "mistake"?

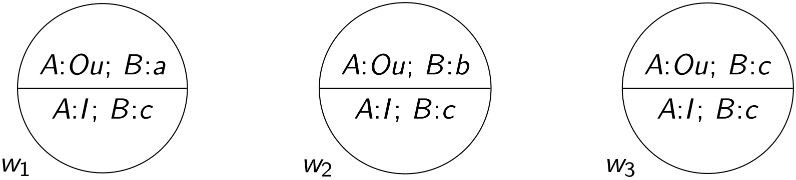
Suppose that W is a nonempty set of states. Each player i will be associated with two functions β_i and σ_i subject to the following constraints:

1. For each $i \in N$, $\beta_i(w)$ is a (possibly empty) i-history and $\sigma_i(w)$ is a strategy for player i.
2. The i-histories $\{\beta_i(w)\}_{i \in N}$ are **coherent**.

A player made a **mistake at a history** $h \in V_i$ in w provided that $\beta_i(w)_h \neq \sigma_i(w)(h)$ (if $\beta_i(w)_h$ is defined).



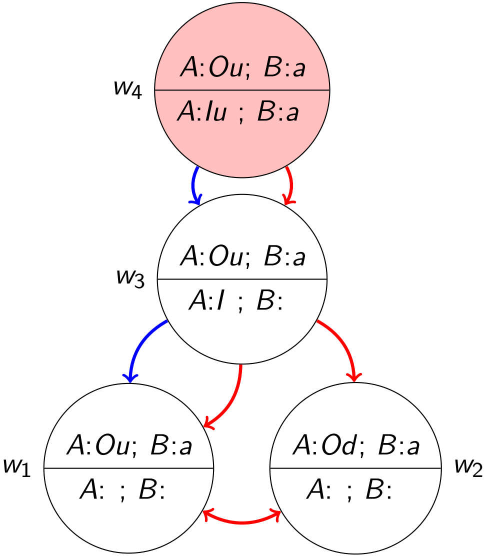
v_0
A
I
v_1
A, B
O
u, a
u, b
u, c
d, a
d, b
d, c
3, 3
2, 2
2, 0
2, 1
1, 1
1, 5
4, 0

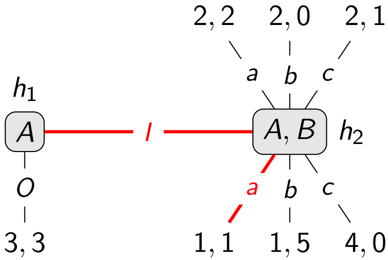

A:Ou; B:a
A:I; B:c
w1
A:Ou; B:b
A:I; B:c
w2
A:Ou; B:c
A:I; B:c
w3

For a game G, a game model is a tuple $\mathcal{M}_G = \langle W, \{(\beta_i, \sigma_i)\}_{i \in N}, \{\succeq_i\}_{i \in N}, \{P_i\}_{i \in N} \rangle$, where:

- For all $w \in W$ and $i \in N$, if $v \in [w]_i$, then $\sigma_i(w) = \sigma_i(v)$. That is, players *know* their own strategy.
- For all $w \in W$ and $i \in N$, for each initial segment $h' \subseteq h_w$ (including the empty history), there is a $w' \in [w]_i$ such that $h_w = h'$.

The Model: Example





Beliefs

$$[w]_i = \{v \mid w \preceq_i v \text{ or } v \preceq_i w\}$$

Beliefs

$$[w]_i = \{v \mid w \preceq_i v \text{ or } v \preceq_i w\}$$

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

Given *any* evidence $F \subseteq W$:

$$P_{i,w}(E \mid F) = P_i(E \mid \max_{\succeq_i}(F \cap [w]_i)).$$

Each state $w \in W$ is associated with a history h_w corresponding to the play of game associated with $\{\beta_i(w)\}_{i \in N}$.

Each state $w \in W$ is associated with a history h_w corresponding to the play of game associated with $\{\beta_i(w)\}_{i \in N}$.

Note that h_w need not be a maximal history, so h_w is the behavior that is observed at state w.

Each state $w \in W$ is associated with a history h_w corresponding to the play of game associated with $\{\beta_i(w)\}_{i \in N}$.

Note that h_w need not be a maximal history, so h_w is the behavior that is observed at state w.

For any $h \in H$, let $[h] = \{w \mid \beta_i(w) = beh_i(h) \text{ for all } i \in N\}$ be the event that the players behaved according to history h.

Each state $w \in W$ is associated with a history h_w corresponding to the play of game associated with $\{\beta_i(w)\}_{i \in N}$.

Note that h_w need not be a maximal history, so h_w is the behavior that is observed at state w.

For any $h \in H$, let $[h] = \{w \mid \beta_i(w) = beh_i(h) \text{ for all } i \in N\}$ be the event that the players behaved according to history h.

$P_{i,w}(E \mid [h_w])$ is i's probability of E given her most plausible explanation of the actions she observed at state w.

Optimal Choice - Induced Strategy

For each $w \in W$, the **strategy realized at** w **by player** i is $s_i(w) : V_i \to Act_i$ defined as follows:

$$s_i(w)(h) = \begin{cases} \beta_i(w)_h & \text{if } \beta_i(w)_h \text{ is defined} \\ \sigma_i(w)(h) & \text{otherwise} \end{cases}$$

Then, $\mathbf{s}(w) = (s_1(w), \ldots, s_n(w))$ is a profile of strategies, and let $Out(\mathbf{s})$ be the (unique) terminal history generated by $\mathbf{s}$.

Optimal Choice - Expected Utility

For any strategy $s_i \in S_i$ for player i, the **expected utility** of s_i at state w is:

$$EU_{i,w}(s_i) = \sum_{w' \in W} P_{i,w}(\{w'\} \mid [h_w]) u_i(Out(s_i, \mathbf{s}_{-i}(w))).$$

Optimal Choice

Let $S_i(w) \subseteq S_i$ be the set of strategies for player i that conform to player i's moves in state w.

$$Opt_i = \{w \mid \sigma_i(w) \text{ maximizes expected utility with respect to } P_{i,w} \text{ and } S_i(w)\}.$$

Rational Choice

We say that a state $w' \in [w]_i$ is an **earlier choice state** provided $\beta_i(w')$ is an initial segment of $\beta_i(w)$.

Player i is **rational-1** at state w provided $w' \in Opt_i$ for *all* earlier choice states w'. Let Rat_i^1 be the set of all states w such that i is rational-1 in w.

Taking Stock

- The focus is on dynamic games with simultaneous moves

Taking Stock

- The focus is on dynamic games with simultaneous moves
  - strategy choice is not an instantaneous commitment, but, rather, a representation of what the players will and would do in the course of playing the game
  - Epistemic models describe how the players will and would revise her beliefs during a play of the game

Taking Stock

- The focus is on dynamic games with simultaneous moves
  - strategy choice is not an instantaneous commitment, but, rather, a representation of what the players will and would do in the course of playing the game
  - Epistemic models describe how the players will and would revise her beliefs during a play of the game

- Game models can be used to *characterize* different solution concepts (e.g., iterated strict dominance, iterated weak dominance, Nash equilibrium, correlated equilibrium, backward induction, extensive-form rationalizability,...)

Backward and Forward Induction

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

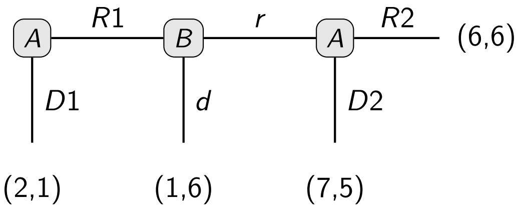
Backward and Forward Induction

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

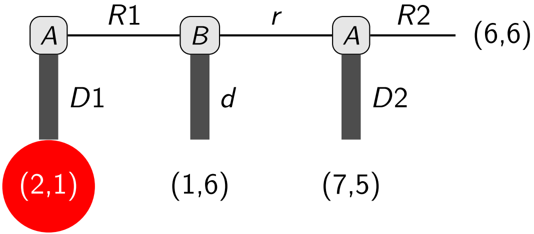
- How should we compare the two "styles of reasoning" about games? (Heifetz & Perea, Reny, Battigalli & Siniscalchi)
- How do (should) players choose between the two different styles of reasoning about games? (Perea, EP & Knoks)

  Aumann & Dreze: *"When all is said and done, how should we play and what should we expect"*.

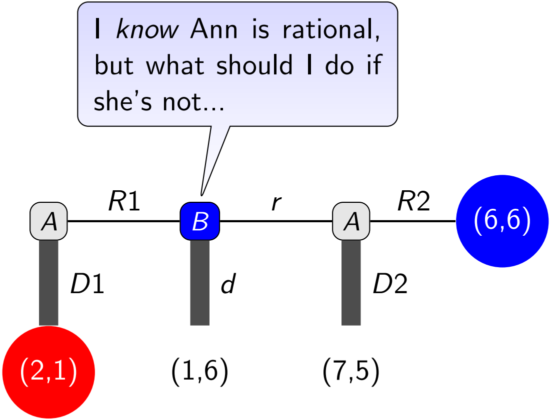
BI Puzzle?

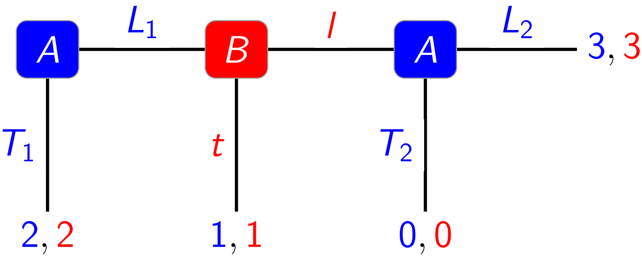


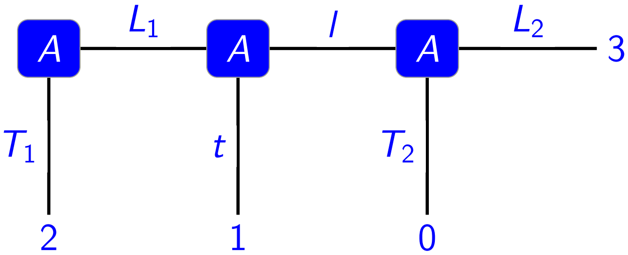
BI Puzzle?



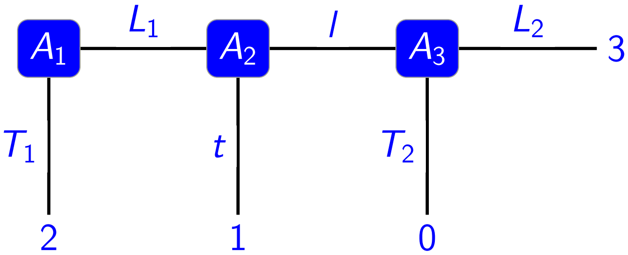
BI Puzzle?


I *know* Ann is rational, but what should I do if she's not...
A
R1
B
r
A
R2
(6,6)
D1
d
D2
(2,1)
(1,6)
(7,5)

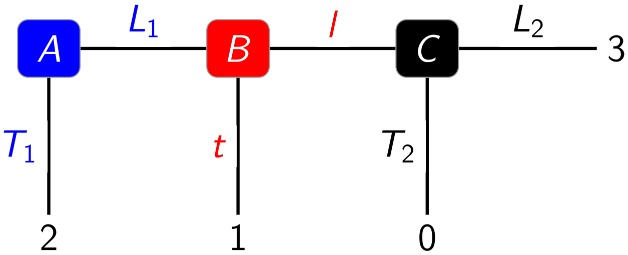

A
L_1
B
l
A
L_2
3, 3
T_1
t
T_2
2, 2
1, 1
0, 0

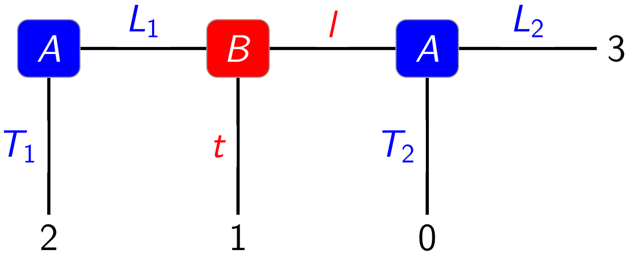


A
L1
A
l
A
L2
3
T1
t
T2
2
1
0



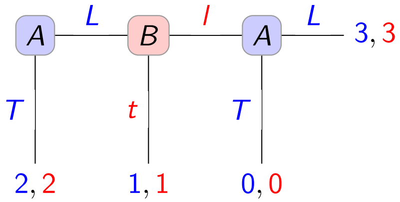
A_1
L_1
A_2
l
A_3
L_2
3
T_1
t
T_2
2
1
0


A
B
C
L1
l
L2
3
T1
t
T2
2
1
0



A
L_1
B
l
A
L_2
3
T_1
t
T_2
2
1
0

		Bob	
		t	l
Ann	T	2,2	2,2
	LT	1,1	0,0
	LL	1,1	3,3



Materially Rational: every choice actually made is optimal (i.e., maximizes subjective expected utility).

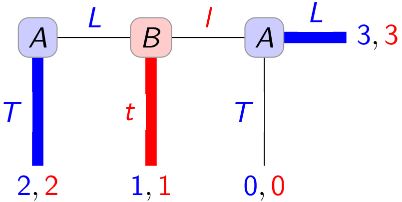
Substantively Rational: the player is materially rational and, in addition, for each *possible* choice, the player *would* have chosen rationally if she had had the opportunity to choose.

Materially Rational: every choice actually made is optimal (i.e., maximizes subjective expected utility).

Substantively Rational: the player is materially rational and, in addition, for each *possible* choice, the player *would* have chosen rationally if she had had the opportunity to choose.

E.g., Taking keys away from someone who is drunk.

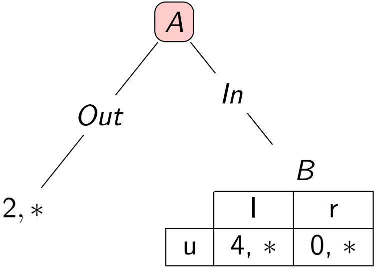
Ann \ Bob	t	l
T	2,2	2,2
LT	1,1	0,0
LL	1,1	3,3

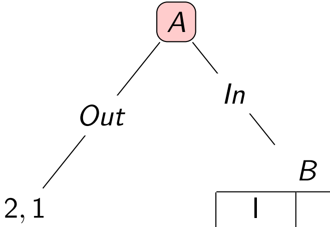


- Bob's belief in a causal counterfactual: Ann would choose L on her second move *if* she had a chance to move.
- But we need to ask what would Bob believe about Ann *if* he learned that he was wrong about her first choice. This is a question about Bob's belief revision policy.

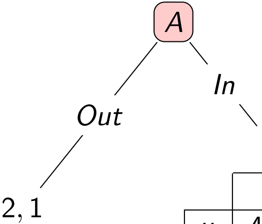
Informal characterizations of BI

- Future choices are *epistemically independent* of any observed behavior
- Any “off-equilibrium” choice is interpreted simply as a mistake (which will not be repeated)
- At each choice point in a game, the players only reason about future paths


A
Out
In
2, ∗
B
l
r
u
4, ∗
0, ∗

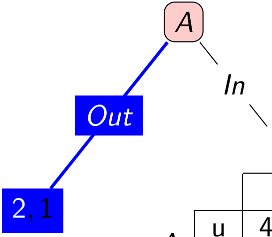


	l	r
u	4, 1	0, 0



B

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



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

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

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

Ann

B *N*

Bob (after *B*)

Ann		l	r
	u	2, 1	-2, 0
	d	-2, 0	-1, 4

Bob (after *N*)

Ann		l	r
	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

Ann		ll	lr	rl	rr
	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 \ Bob	ll	lr	rl	rr
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

Ann

	ll	lr	rl	rr
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

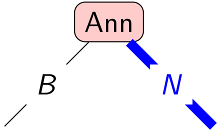
Bob

Ann

	ll	lr	rl	rr
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 (left subgame, after B)

Ann		l	r
	u	2, 1	-2, 0
	d	-2, 0	-1, 4

Bob (right subgame, after N)

Ann		l	r
	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.

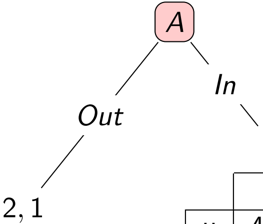
Four key issues

- Should the analysis take place on the tree or the matrix? (plans vs. strategies)
- 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?

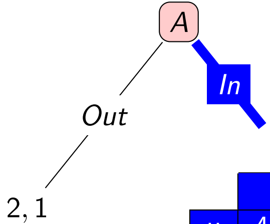
"...in general, a player's beliefs about what another player will do are based on an inference from two other kinds of beliefs: beliefs about the passive beliefs of that player, and beliefs about her rationality.

“...in general, a player’s beliefs about what another player will do are based on an inference from two other kinds of beliefs: beliefs about the passive beliefs of that player, and beliefs about her rationality. If one’s prediction based on these beliefs is defeated, one must choose whether to revise one’s belief about the other players’s beliefs or one’s belief that she is rational...

“...in general, a player’s beliefs about what another player will do are based on an inference from two other kinds of beliefs: beliefs about the passive beliefs of that player, and beliefs about her rationality. If one’s prediction based on these beliefs is defeated, one must choose whether to revise one’s belief about the other players’s beliefs or one’s belief that she is rational...But the assumption that the rationalization principle is common belief is itself an assumption about the passive beliefs of other players, and so it is itself something that (according to the principle) might have to be given up in the face of surprising behavioral information. So the rationalization principle undermines its own stability.” (pg. 51, Stalnaker)

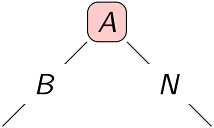


		B	
		l	r
A	u	4, 1	0, 0
	d	0, 0	1, 4



B

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

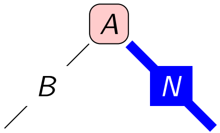


Bob

Ann		l	r
	u	2, 1	-2, 0
	d	-2, 0	-1, 4

Bob

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



Bob

Ann		l	r
	u	2, 1	-2, 0
	d	-2, 0	-1, 4

Bob

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

"...Only if one assumes a specific infinite hierarchy of belief revision priorities can one be sure that unlimited iteration of forward induction reasoning will work....But it seems to me that such detailed assumptions about belief revision policy....have no intuitive plausibility."

(Stalnaker, pg. 53)

Algorithm and a “Theorem”

Algorithm: Eliminate weakly dominated strategies for *just two* rounds, and then eliminate *strictly* dominated strategies iteratively.

Algorithm and a “Theorem”

Algorithm: Eliminate weakly dominated strategies for *just two* rounds, and then eliminate *strictly* dominated strategies iteratively.

“Theorem”: It can be proved that all and only strategies that survive this process are realizable in sufficiently rich models in which it is common belief that all players are rational, and that all revise their beliefs in conformity with the rationalization principle.

Algorithm and a "Theorem"

Algorithm: Eliminate weakly dominated strategies for *just two* rounds, and then eliminate *strictly* dominated strategies iteratively.

"Theorem": It can be proved that all and only strategies that survive this process are realizable in sufficiently rich models in which it is common belief that all players are rational, and that all revise their beliefs in conformity with the rationalization principle.

Algorithm and a "Theorem"

Algorithm: Eliminate weakly dominated strategies for *just two* rounds, and then eliminate *strictly* dominated strategies iteratively.

"Theorem": It can be proved that all and only strategies that survive this process are realizable in sufficiently rich models in which it is common belief that all players are rational, and that all revise their beliefs in conformity with the rationalization principle.

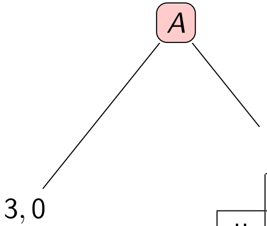
Algorithm and a “Theorem”

Algorithm: Eliminate weakly dominated strategies for *just two* rounds, and then eliminate *strictly* dominated strategies iteratively.

“Theorem”: It can be proved that all and only strategies that survive this process are realizable in sufficiently rich models in which it is common belief that all players are rational, and that all revise their beliefs in conformity with the rationalization principle.

Joint work with Aleks Knoks: “Theorem” $\hookrightarrow$ Theorem

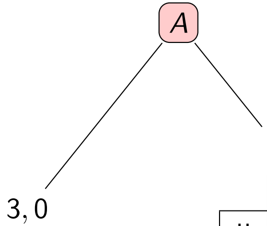
Backward *versus* Forward Induction



Ann \ Bob	l	c	r
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

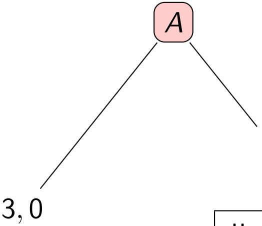


Bob

Ann		l	c	r
	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.

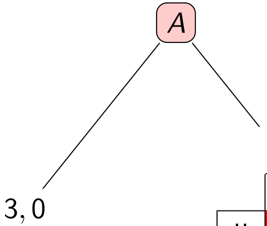
Backward *versus* Forward Induction



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

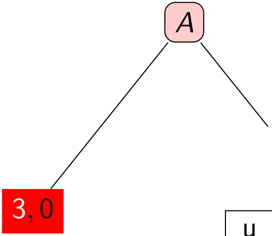


Bob

Ann	l	c	r
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.

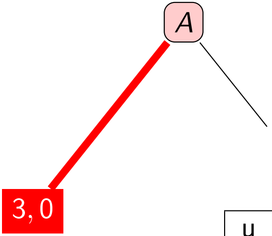
Backward *versus* Forward Induction



		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.

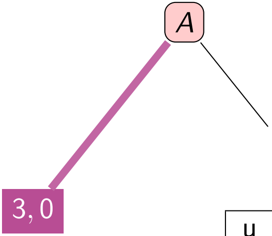
Backward *versus* Forward Induction



Ann		Bob: l	Bob: c	Bob: r
	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

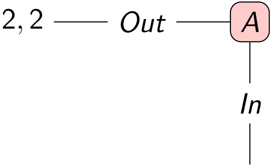


Bob

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

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

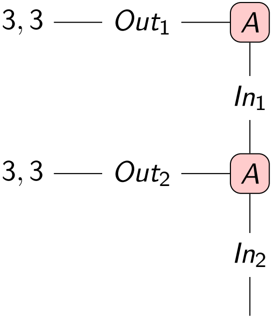
Rationalization *versus* Mistakes



B

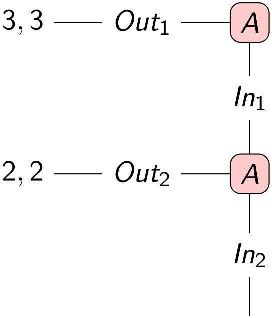
		l	r
A	u	4, 1	0, 0
	d	0, 0	1, 4

Rationalization *versus* Mistakes



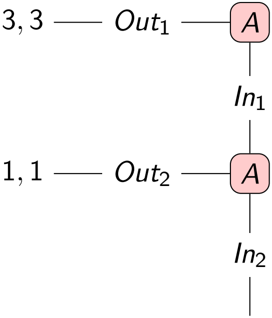
		B: l	B: r
A	u	4, 1	0, 0
A	d	0, 0	1, 4

Rationalization *versus* Mistakes



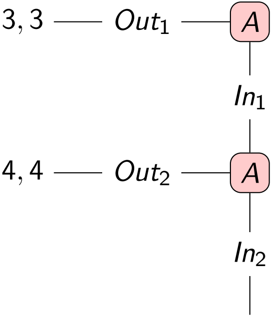
		B	
		l	r
A	u	4, 1	0, 0
	d	0, 0	1, 4

Rationalization *versus* Mistakes



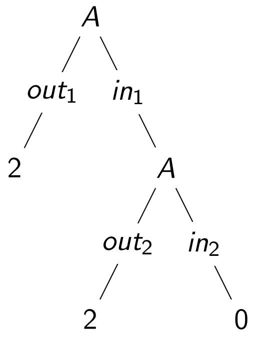
A \ B	l	r
u	4, 1	0, 0
d	0, 0	1, 4

Rationalization *versus* Mistakes

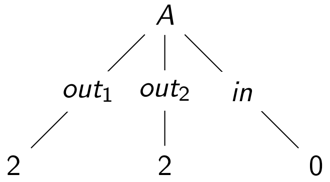


		B	
		l	r
A	u	4, 1	0, 0
	d	0, 0	1, 4

Allowing for mistakes



D_1



D_2

What can we do with our model?

What can we do with our model?

- Represent FI, BI, as well as "hybrid" belief revision policies

What can we do with our model?

- Represent FI, BI, as well as “hybrid” belief revision policies

- Condition on choices, “mistakes”, observed behavior, rationality, etc.

What can we do with our model?

- Represent FI, BI, as well as "hybrid" belief revision policies
- Condition on choices, "mistakes", observed behavior, rationality, etc.
- Prove (or re-prove) "characterization results"

What can we do with our model?

- Represent FI, BI, as well as “hybrid” belief revision policies
- Condition on choices, “mistakes”, observed behavior, rationality, etc.
- Prove (or re-prove) “characterization results”
- An implicit assumption in EGT literature: choice of strategy implies its execution: Our framework highlights the role that this assumption plays in (epistemic) game-theoretic analyses

Concluding Remarks: Characterizing BI

(1) For all $i \in N$ and $w, w' \in W$, if $w' \in [w]_i$, then for each $w'' \in \max_{\succeq_i}([w]_i \cap [h_{w'}])$ there is a $w''' \in \max_{\succeq_i}([w]_i)$ such that $\sigma_j(w'') = \sigma_j(w''')$ for all $j \in N$.

Players cannot learn anything about their opponents' strategies that they did not already know at the beginning of the game.

Concluding Remarks: Characterizing BI

(2) For all $i \in N$ and $w, w' \in W$, if $w' \in [w]_i$ and there is some $j \in N$ with $\beta_{v_j}(w') \neq \sigma_j(w')$, there is a $w'' \in [w]_i$ such that $w'' \succ_i w'$ and $\beta_v(w'') = \sigma_v(w'')$ and all the rest is the same.

Players never believe that a mistake is more likely to happen than not.

Finally, our the third constraint:

Concluding Remarks: Characterizing BI

(3) For all $w \in W$, $i \in N$ and $h \in H$, there is a $w' \in [w]_i$ such that h is generated by $\{\beta_i(w')\}_{i \in N}$.

Players do not *rule-out* any sequence of play.

Concluding Remarks: Characterizing BI

Theorem. Suppose that $\mathcal{G}$ is an extensive form game (without simultaneous moves) in "general position" and $\mathcal{M}_{\mathcal{G}}$ is a model for $\mathcal{G}$ satisfying the constraints (1)–(3). Then $CB(Rat^1) \subseteq CB(BI)$.

Concluding Remarks: Trembling Hand Mistakes

There cannot be any mistakes if the players are absolutely rational. Nevertheless, a satisfactory interpretation of equilibrium points in extensive games seems to require that the possibility of mistakes is not completely excluded. This can be achieved by a point of view which looks at complete rationality as the limiting case of incomplete rationality. *(Selten, pg. 35)*

Concluding Remarks: Irrationality in Games

[T]he rules of rational behavior must provide definitely for the possibility of irrational conduct on the part of others. Imagine that we have discovered a set of rules for all participants—to be termed as "optimal" or "rational"—each of which is indeed optimal provided that the other participants conform. Then the question remains as to what will happen if some of the participants do not conform.

(von Neumann & Morgenstern, pg. 32)

Concluding Remarks: 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.

Concluding Remarks: 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.

Concluding Remarks: 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. 307352., 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.

Concluding Remarks: 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 $s_i \in B_i$, 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
	a_1	0, 7	2, 5	7, 0	0, 1
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
	a_1	0, 7	2, 5	7, 0	0, 1
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

		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
- $(\{a_1, a_2, a_3\}, \{b_1, b_2, b_3\})$ 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 such that the *projection* of the the event where there is common belief that all players are rational is $B_1 \times \cdots \times B_n$.

Concluding Remarks: 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.

P. Battigalli and M. Siniscalchi. *Strong Belief and Forward Induction Reasoning*. Journal of Economic Theory 106(2), pgs. 356391, 2002.

R. Stalnaker. *Belief revision in games: forward and backward induction*. Mathematical Social Sciences, 36, pgs. 31 - 56, 1998.

Thank You!

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