Reasoning in Games

Lecture 5

Eric Pacuit

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

August 14, 2015

Plan

- ✓ Day 1: Decision Theory
- ✓ Day 2: From Decisions to Games
- ✓ Day 3: Game Models
- ✓ Day 4: Modeling Deliberation (in Games)
- ▸ Day 5: Backward and Forward Induction, Concluding Remarks

Why go beyond the basic Bayesian model?

Why go beyond the basic Bayesian model?

- **Counterfactual beliefs/choices** are important for assessing the rationality of a strategy.

Why go beyond the basic Bayesian model?

- **Counterfactual beliefs/choices** are important for assessing the rationality of a strategy.

- **Static models of dynamic games**: A game model represents how a player will and would change her beliefs if her opponents take the game in various directions.

Why go beyond the basic Bayesian model?

- **Counterfactual beliefs/choices** are important for assessing the rationality of a strategy.
- **Static models of dynamic games**: A game model represents how a player will and would change her beliefs if her opponents take the game in various directions.
- A **conditional choice** (do a if E) is rational iff doing a would be rational if the player were to learn E.

Why go beyond the basic Bayesian model?

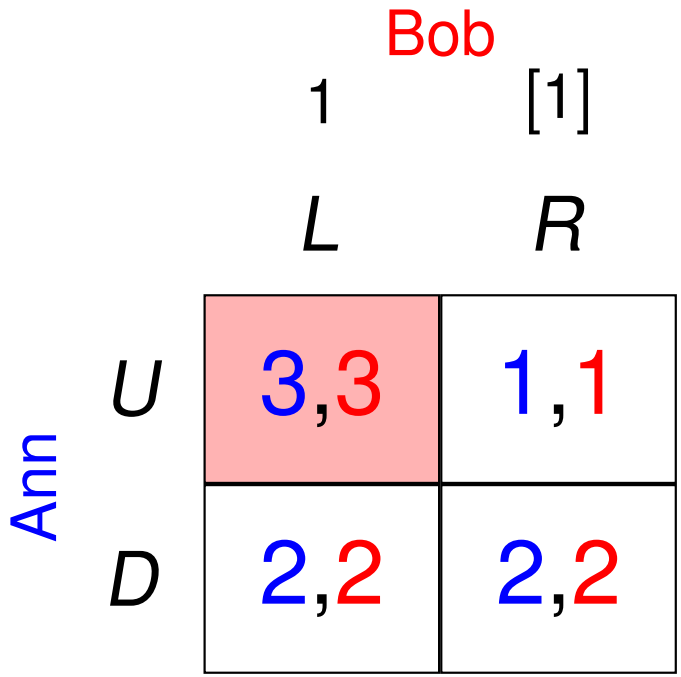
- **Counterfactual beliefs/choices** are important for assessing the rationality of a strategy.
- **Static models of dynamic games**: A game model represents how a player will and would change her beliefs if her opponents take the game in various directions.
- A **conditional choice** (do a if E) is rational iff doing a would be rational if the player were to learn E.
- Strategy choices should be robust against **mistaken beliefs** (weak dominance).

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

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 $[w]_{\approx_i} = \{v \mid w \approx_i v\}$

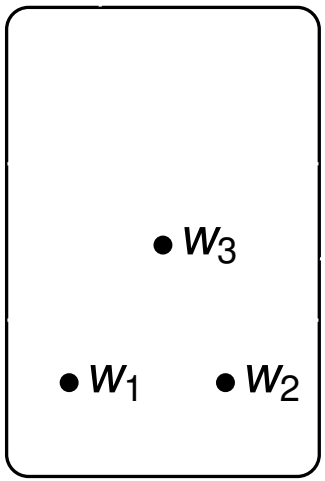
2. $w \; R_i \; v$ iff $v \in Max_{\preceq_i}([w]_{\approx_i})$

3. $B_{i,w}(F) = Max_{\preceq_i}(F \cap [w]_{\approx_i})$

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

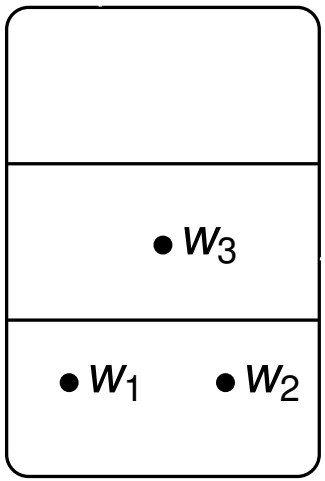
Belief Revision via Plausibility

- $W = \{w_1, w_2, w_3\}$



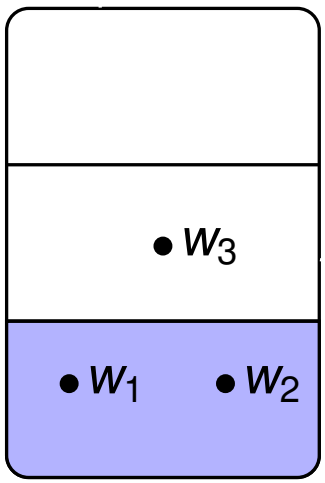
Belief Revision via Plausibility

- $W = \{w_1, w_2, w_3\}$
- $w_1 \preceq w_2$ and $w_2 \preceq w_1$ (w_1 and w_2 are equi-plausbile)
- $w_1 \prec w_3$ ($w_1 \preceq w_3$ and $w_3 \not\preceq w_1$)
- $w_2 \prec w_3$ ($w_2 \preceq w_3$ and $w_3 \not\preceq w_2$)

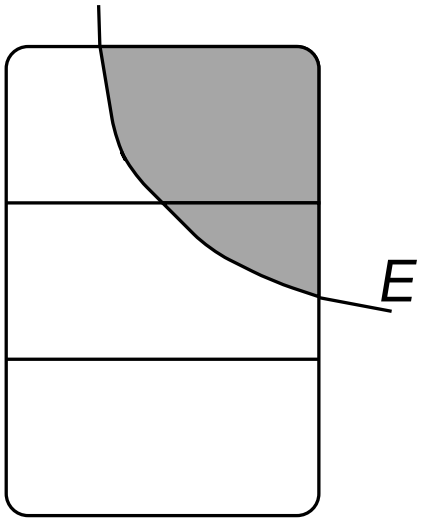


Belief Revision via Plausibility

- $W = \{w_1, w_2, w_3\}$
- $w_1 \preceq w_2$ and $w_2 \preceq w_1$ (w_1 and w_2 are equi-plausbile)
- $w_1 \prec w_3$ ($w_1 \preceq w_3$ and $w_3 \not\preceq w_1$)
- $w_2 \prec w_3$ ($w_2 \preceq w_3$ and $w_3 \not\preceq w_2$)
- $\{w_1, w_2\} \subseteq Max_{\preceq}([w_i])$

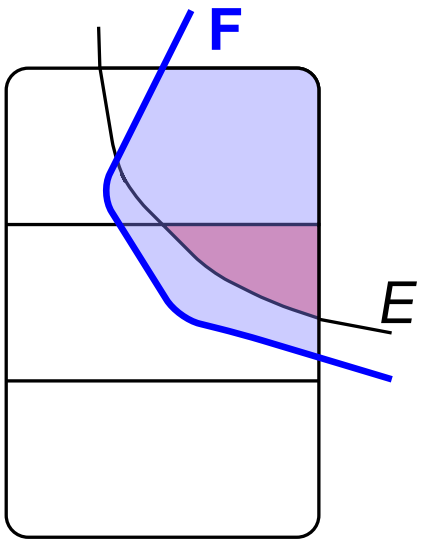


Belief Revision via Plausibility



Conditional Belief: $B^E F$

Belief Revision via Plausibility



Conditional Belief: $B^E F$

$$Max_{\leq}(E) \subseteq F$$

Resiliency, Robust Belief, Stable Belief

B. Skyrms. *Resiliency, propensities, and causal necessity*. Journal of Philosophy, 74:11, pgs. 704 - 713, 1977.

A. Baltag and S. Smets. *Probabilistic Belief Revision*. Synthese, 2008.

H. Leitgeb. *Reducing belief simpliciter to degrees of belief*. Annals of Pure and Applied Logic, 16:4, pgs. 1338 - 1380, 2013.

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

Certainty: $P(H) = 1$

Absolute Certainty: for all E: $P(H \mid E) = 1$

Certainty: $P(H) = 1$

Absolute Certainty: for all E: $P(H \mid E) = 1$

Strong Belief: for all $E \in \mathfrak{A}$ with $H \cap E \neq \emptyset$ and $P(E) \neq 0$: $P(H \mid E) = 1$

Certainty: $P(H) = 1$

Absolute Certainty: for all E: $P(H \mid E) = 1$

Strong Belief: for all $E \in \mathfrak{U}$ with $H \cap E \neq \emptyset$ and $P(E) \neq 0$: $P(H \mid E) = 1$

The set of possible evidence/observations

Certainty: $P(H) = 1$

Absolute Certainty: for all E: $P(H \mid E) = 1$

Strong Belief: for all $E \in \mathfrak{A}$ with $H \cap E \neq \emptyset$ and $P(E) \neq 0$: $P(H \mid E) = 1$

The set of possible evidence/observations

The evidence does not contradict the hypothesis

Certainty: $P(H) = 1$

Absolute Certainty: for all E: $P(H \mid E) = 1$

Strong Belief: for all $E \in \mathfrak{A}$ with $H \cap E \neq \emptyset$ and $P(E) \neq 0$: $P(H \mid E) \geq t$

The set of possible evidence/observations

The evidence does not contradict the hypothesis

Contextually defined threshold

CPS (Popper Space)

A **conditional probability space** (CPS) over $(W, \mathfrak{U})$ is a tuple $(W, \mathfrak{U}, \mathfrak{B}, \mu)$ such that $\mathfrak{U}$ is an algebra over W, $\mathfrak{B}$ is a set of subsets of W (not necessarily an algebra) that does not contain $\emptyset$ and $\mu : \mathfrak{U} \times \mathfrak{B} \rightarrow [0, 1]$ satisfying the following conditions:

1. $\mu(U \mid U) = 1$ if $U \in \mathfrak{U}'$
2. $\mu(E_1 \cup E_1 \mid U) = \mu(E_1 \mid U) + \mu(E_2 \mid U)$ if $E_1 \cap E_2 = \emptyset$, $U \in \mathfrak{B}$ and $E_1, E_2 \in \mathfrak{U}$
3. $\mu(E \mid U) = \mu(E \mid X) * \mu(X \mid U)$ if $E \subseteq X \subseteq U$, $U, X \in \mathfrak{B}$ and $E \in \mathfrak{U}$.

LPS (Lexicographic Probability Space)

A **lexicographic probability space** (LPS) (of length α) is a tuple $(W, \mathcal{F}, \vec{\mu})$ where W is a set of possible worlds, $\mathcal{F}$ is an algebra over W and $\vec{\mu}$ is a sequence of (finitely/countable additive) probability measures on $(W, \mathcal{F})$ indexed by ordinals $< \alpha$.

Fix an *LPS* $\vec{\mu} = (\mu_0, \ldots, \mu_n)$

- E is certain: $\mu_0(E) = 1$

Fix an *LPS* $\vec{\mu} = (\mu_0, \ldots, \mu_n)$

- E is certain: $\mu_0(E) = 1$
- E is absolutely certain: $\mu_i(E) = 1$ for all $i = 1, \ldots, n$

Fix an *LPS* $\vec{\mu} = (\mu_0, \ldots, \mu_n)$

- E is certain: $\mu_0(E) = 1$
- E is absolutely certain: $\mu_i(E) = 1$ for all $i = 1, \ldots, n$
- E is *assumed*: there exists k such that $\mu_i(E) = 1$ for all $i \leq k$ and $\mu_i(E) = 0$ for all $k < i < n$.

Fix an *LPS* $\vec{\mu} = (\mu_0, \ldots, \mu_n)$

- E is certain: $\mu_0(E) = 1$
- E is absolutely certain: $\mu_i(E) = 1$ for all $i = 1, \ldots, n$
- E is *assumed*: there exists k such that $\mu_i(E) = 1$ for all $i \leq k$ and $\mu_i(E) = 0$ for all $k < i < n$.

The key notion is **rationality and common assumption of rationality** (RCAR).

NPS (non-standard probability measures)

$\mathbb{R}^*$ is a *non-Archimedean* field that includes the real numbers as a subfield but also has *infinitesimals*.

For all $b \in \mathbb{R}^*$ such that $-r < b < r$ for some $r \in \mathbb{R}$, there is a unique closest real number a such that $|a - b|$ is an infinitesimal. Let $st(b)$ denote the closest standard real to b.

NPS (non-standard probability measures)

$\mathbb{R}^*$ is a *non-Archimedean* field that includes the real numbers as a subfield but also has *infinitesimals*.

For all $b \in \mathbb{R}^*$ such that $-r < b < r$ for some $r \in \mathbb{R}$, there is a unique closest real number a such that $|a - b|$ is an infinitesimal. Let $st(b)$ denote the closest standard real to b.

A **nonstandard probability space** (NPS) is a tuple $(W, \mathcal{F}, \mu)$ where W is a set of possible worlds, $\mathcal{F}$ is an algebra over W and μ assigns to elements of $\mathcal{F}$, nonnegative elements of $\mathbb{R}^*$ such that $\mu(W) = 1$, $\mu(E \cup F) = \mu(E) + \mu(F)$ if E and F are disjoint.

J. Halpern. *Lexicographic probability, conditional probability, and nonstandard probability*. Games and Economic Behavior, 68:1, pgs. 155 - 179, 2010.

Given an *extensive game* G, let $\mathcal{H}$ be the set of **histories** in G (i.e., finite paths in G), and $[h]$ be the set of states in which the history h is realized.

Given an *extensive game* G, let $\mathcal{H}$ be the set of **histories** in G (i.e., finite paths in G), and $[h]$ be the set of states in which the history h is realized.

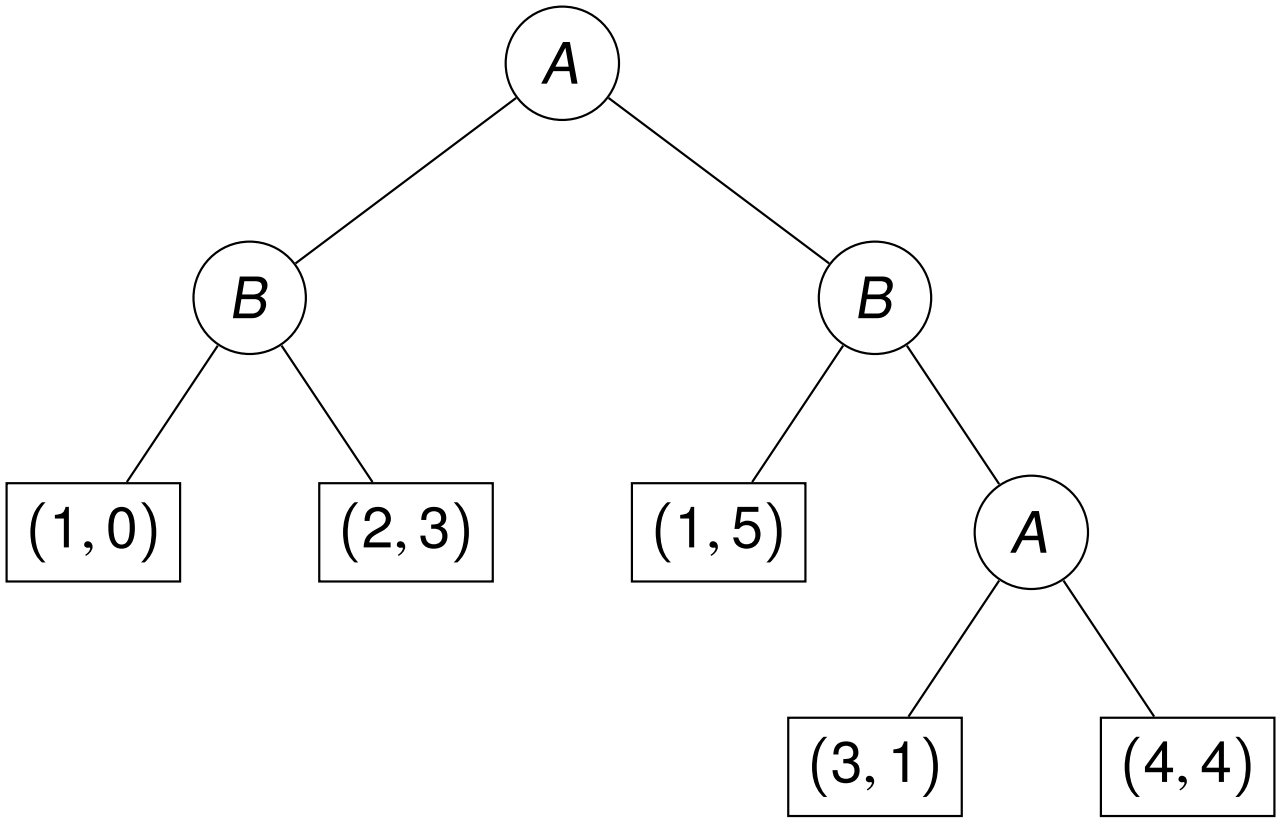
$$SB_{i,w}(E) = \bigcap_{h \,:\, E \cap [h] \neq \emptyset} P_{i,w}(E \mid [h]) = 1$$

Given an *extensive game* G, let $\mathcal{H}$ be the set of **histories** in G (i.e., finite paths in G), and $[h]$ be the set of states in which the history h is realized.

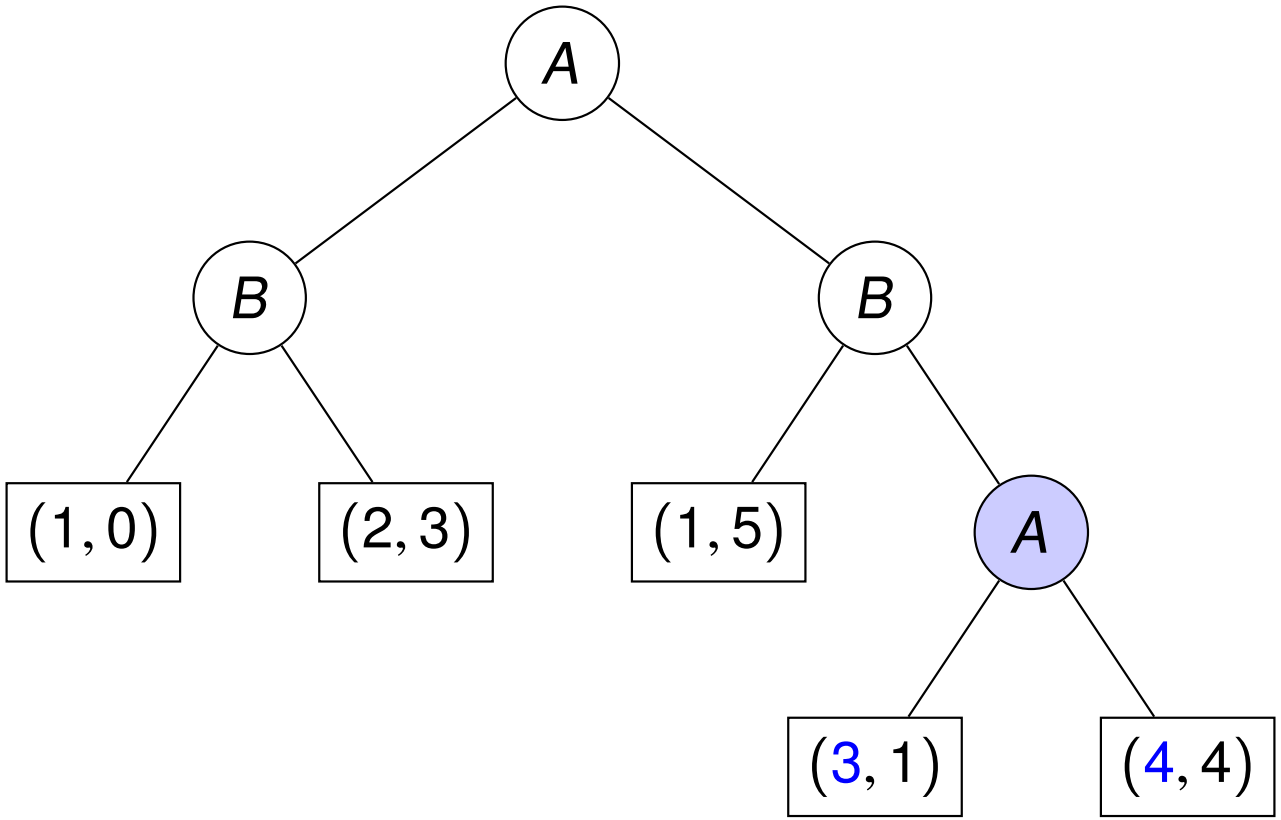
$$SB_{i,w}(E) = \bigcap_{h \,:\, E \cap [h] \neq \emptyset} P_{i,w}(E \mid [h]) = 1$$

The "working hypothesis" E is maintained given any observation that does not rule-out E.

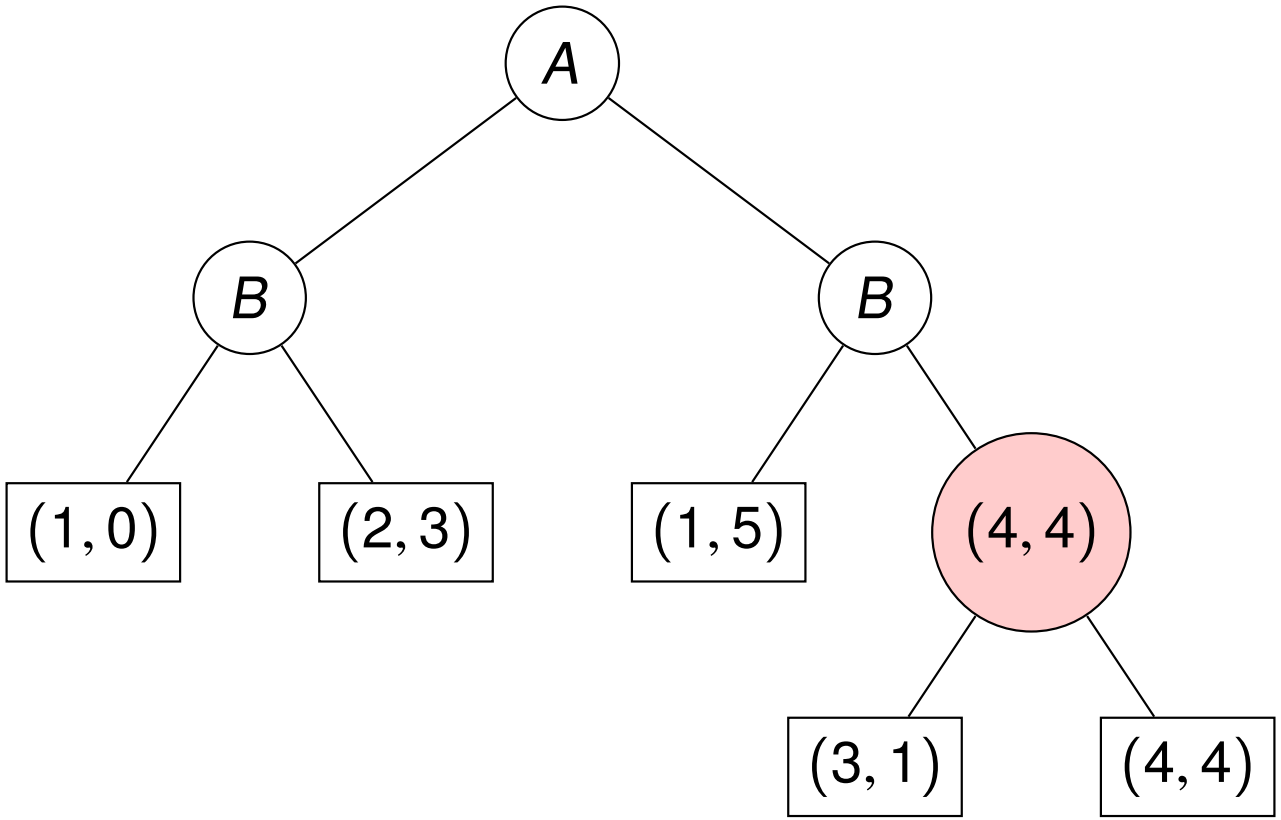
Backward Induction



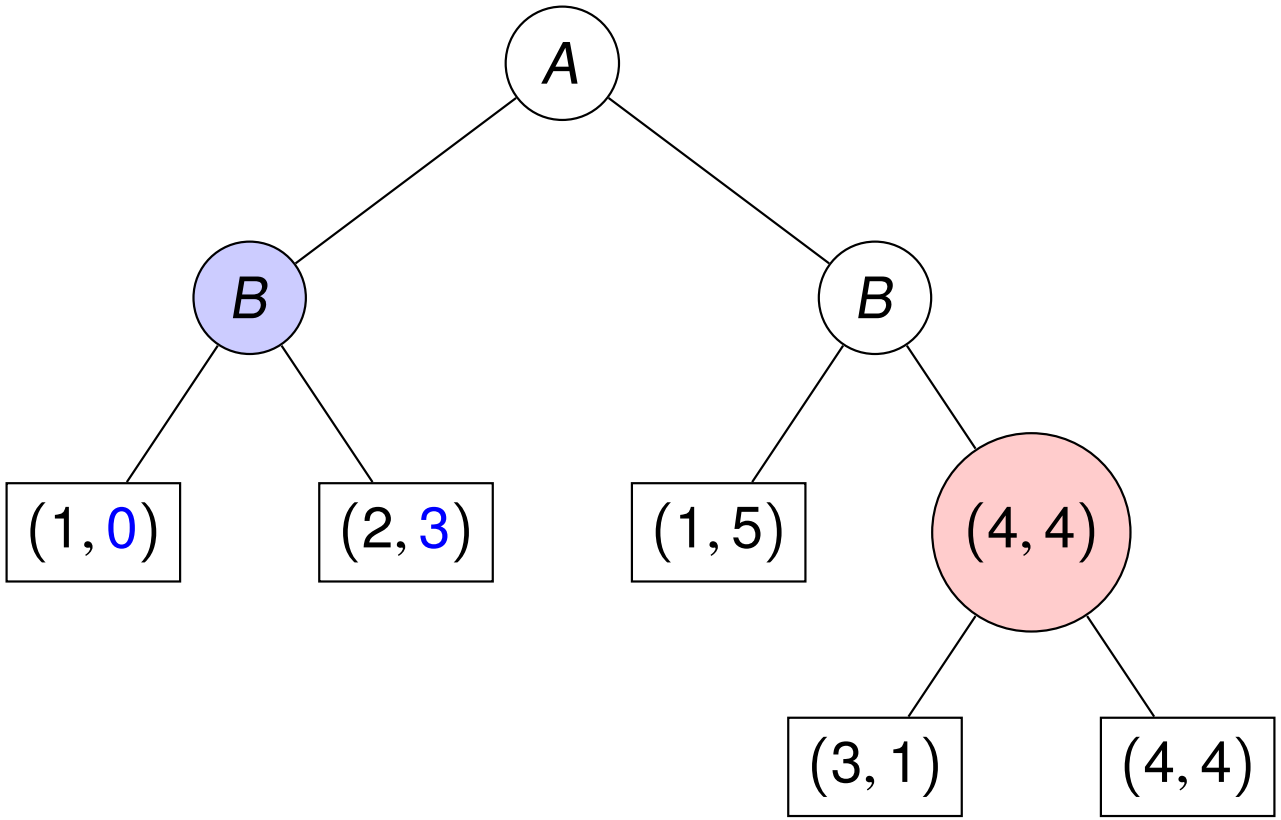
Backward Induction



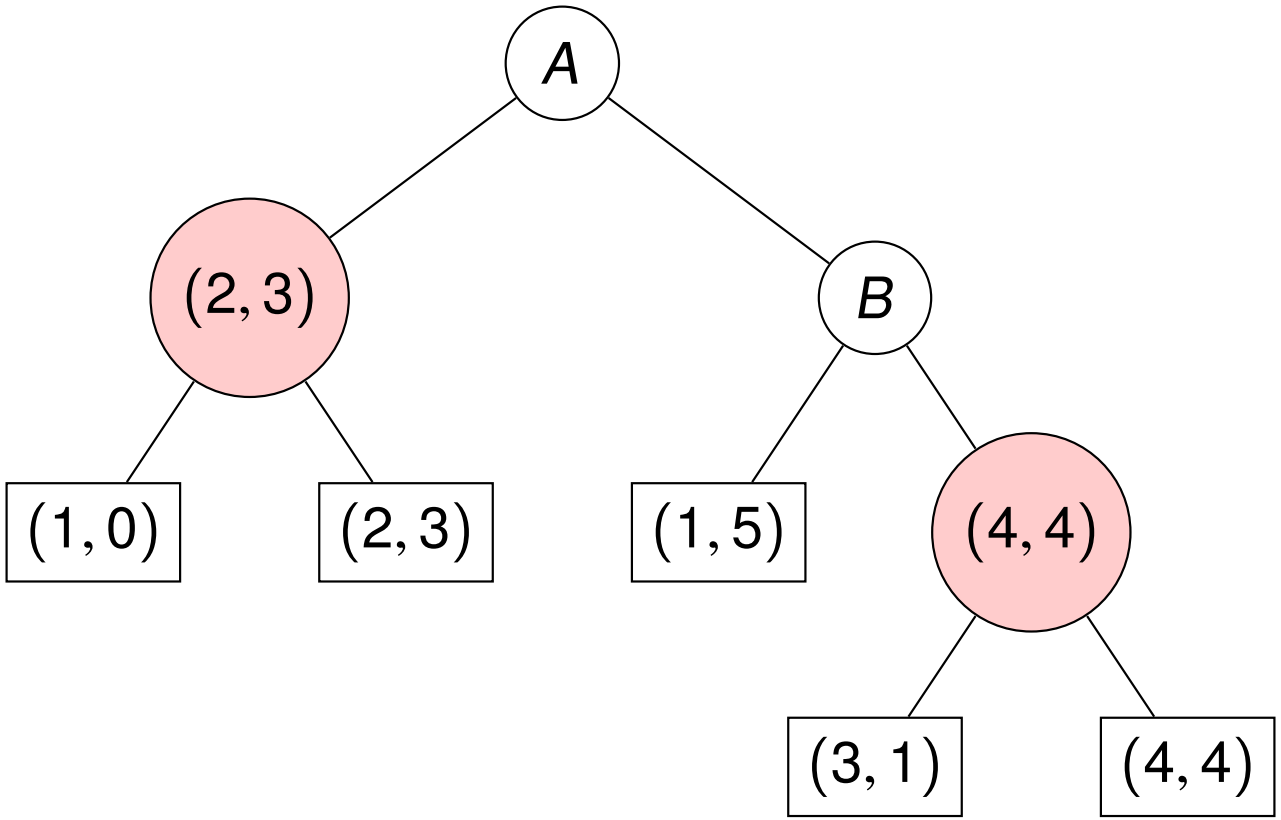
Backward Induction



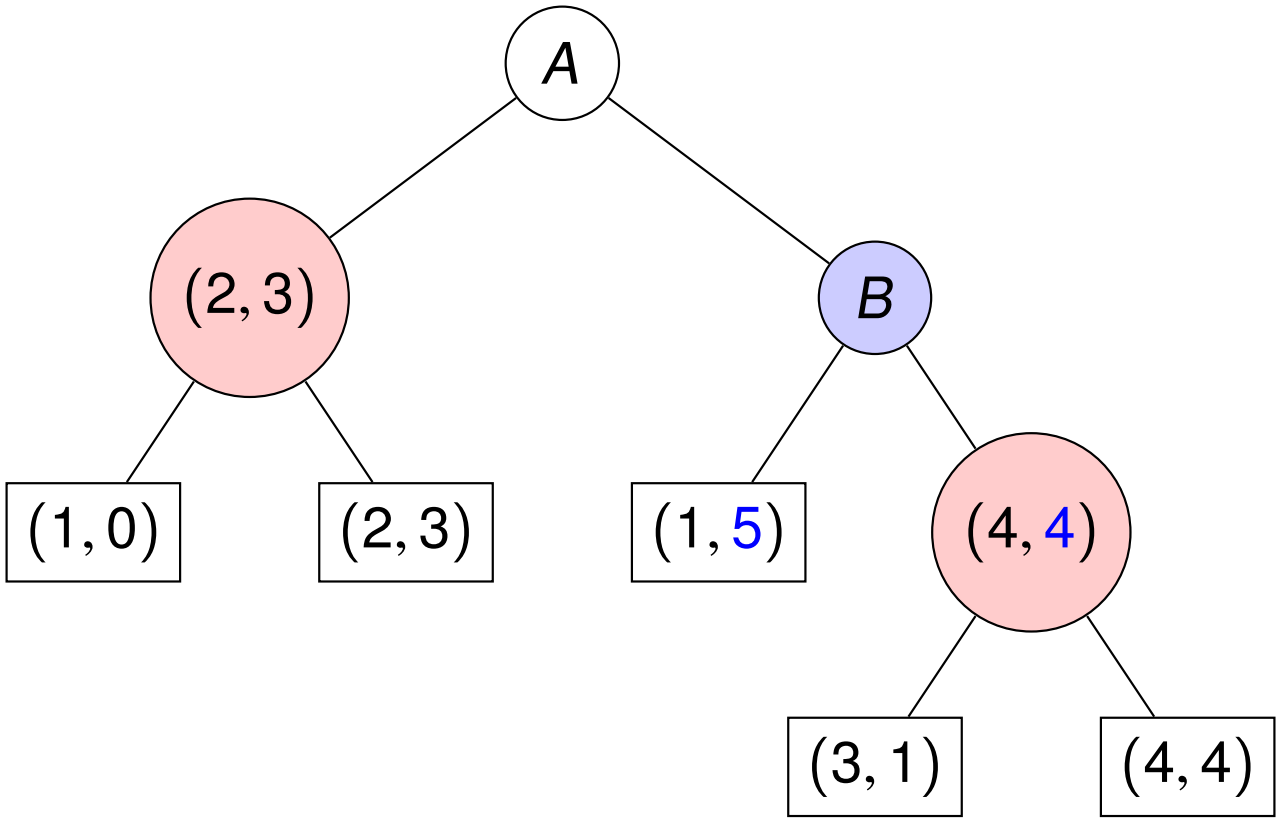
Backward Induction



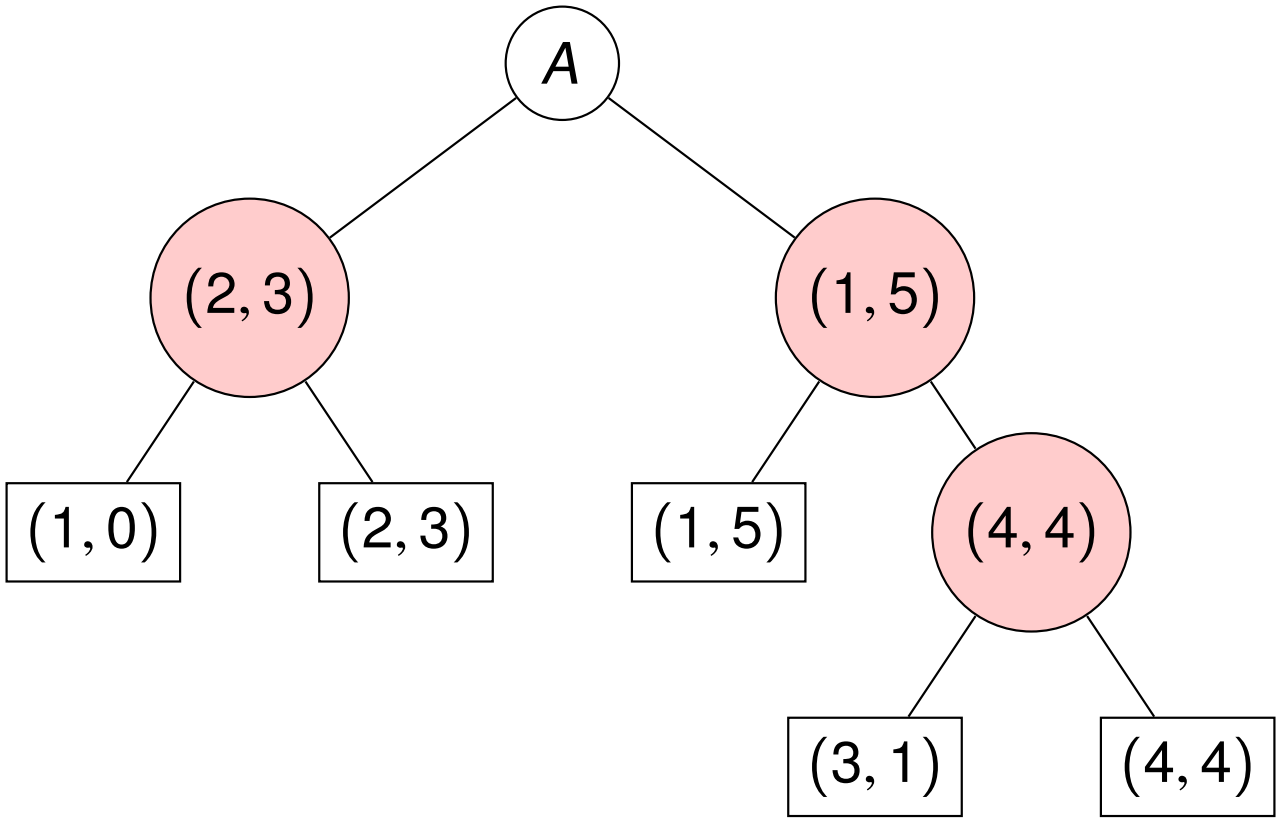
Backward Induction



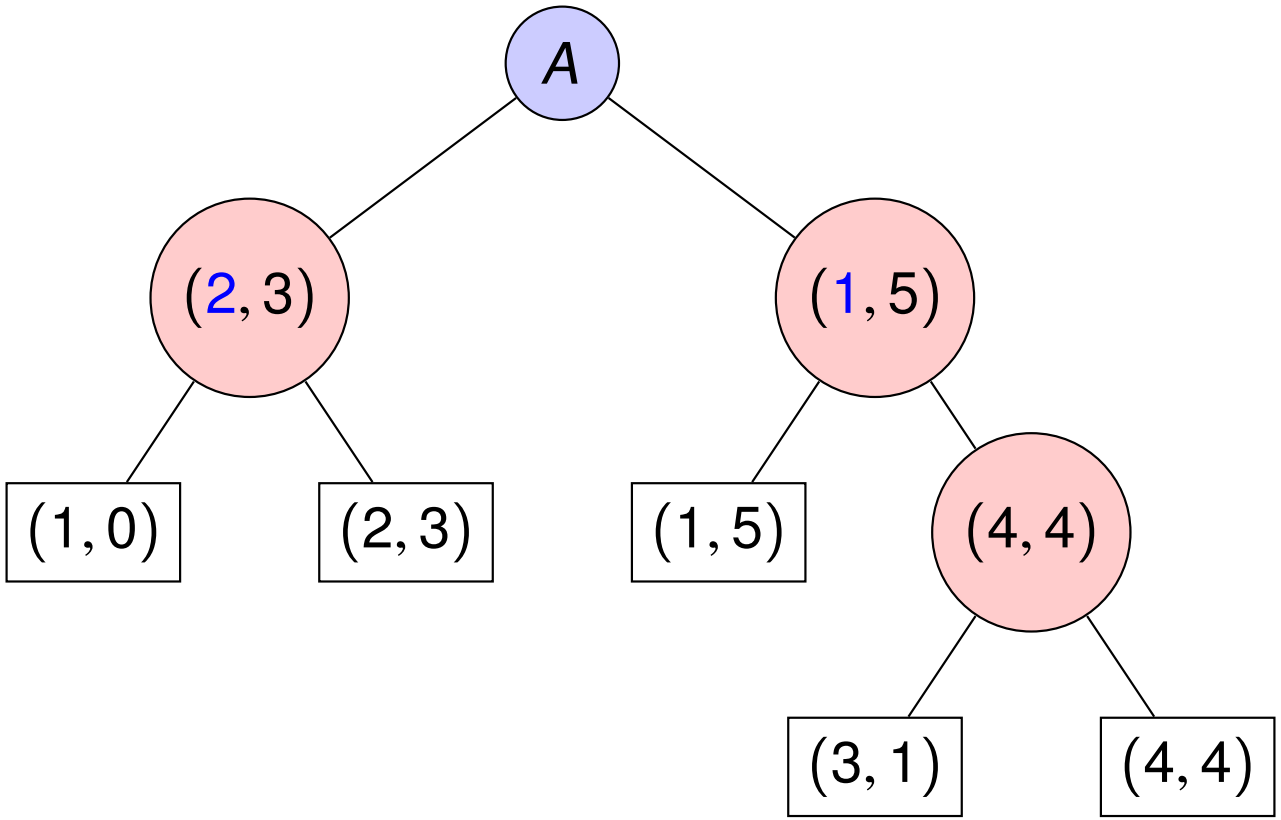
Backward Induction



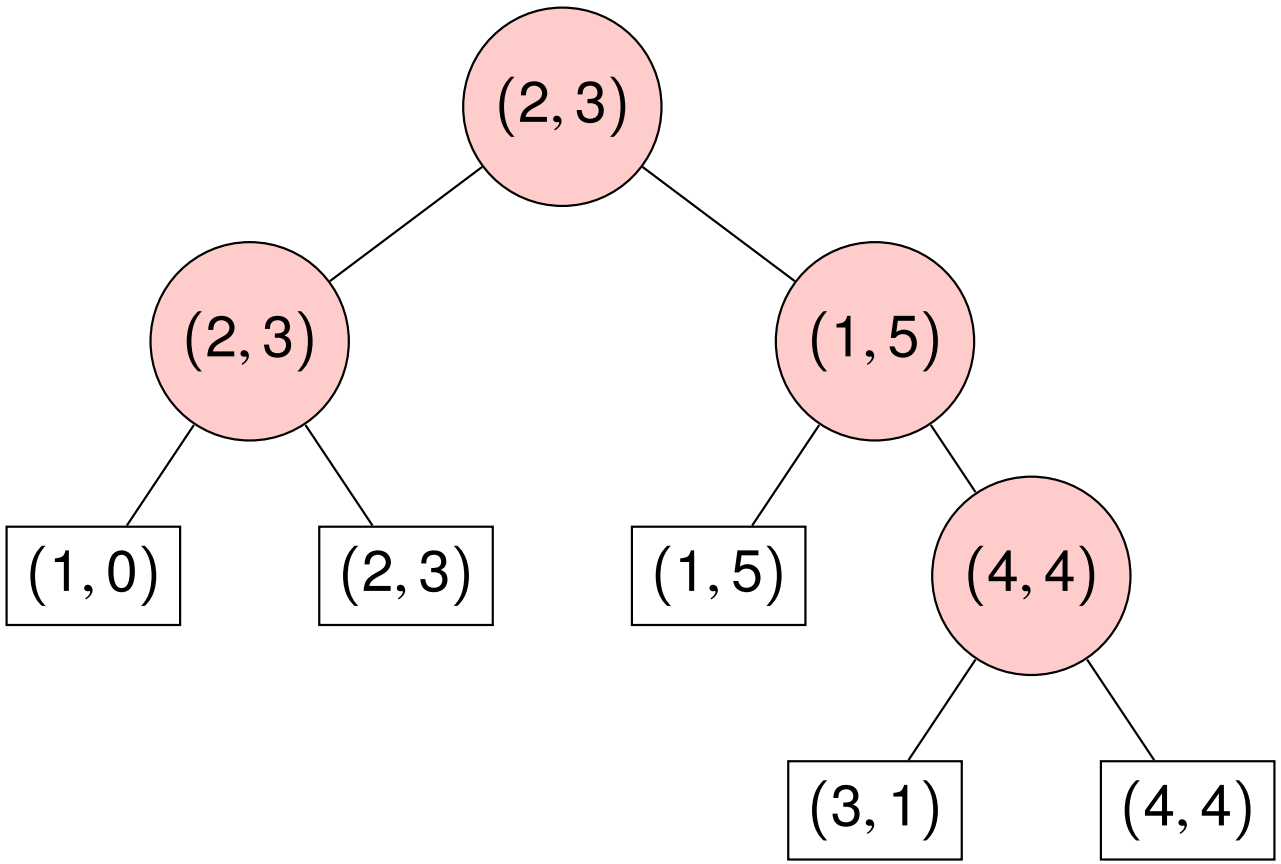
Backward Induction



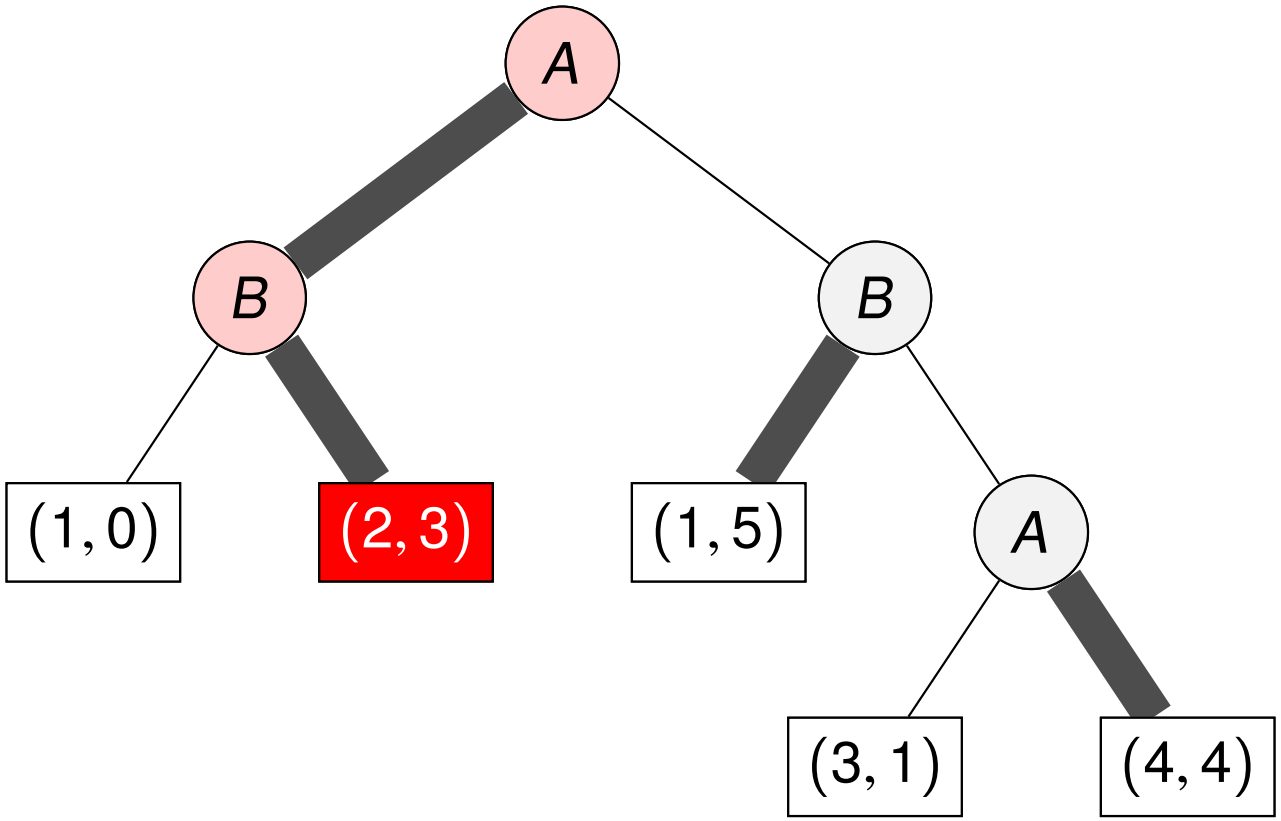
Backward Induction



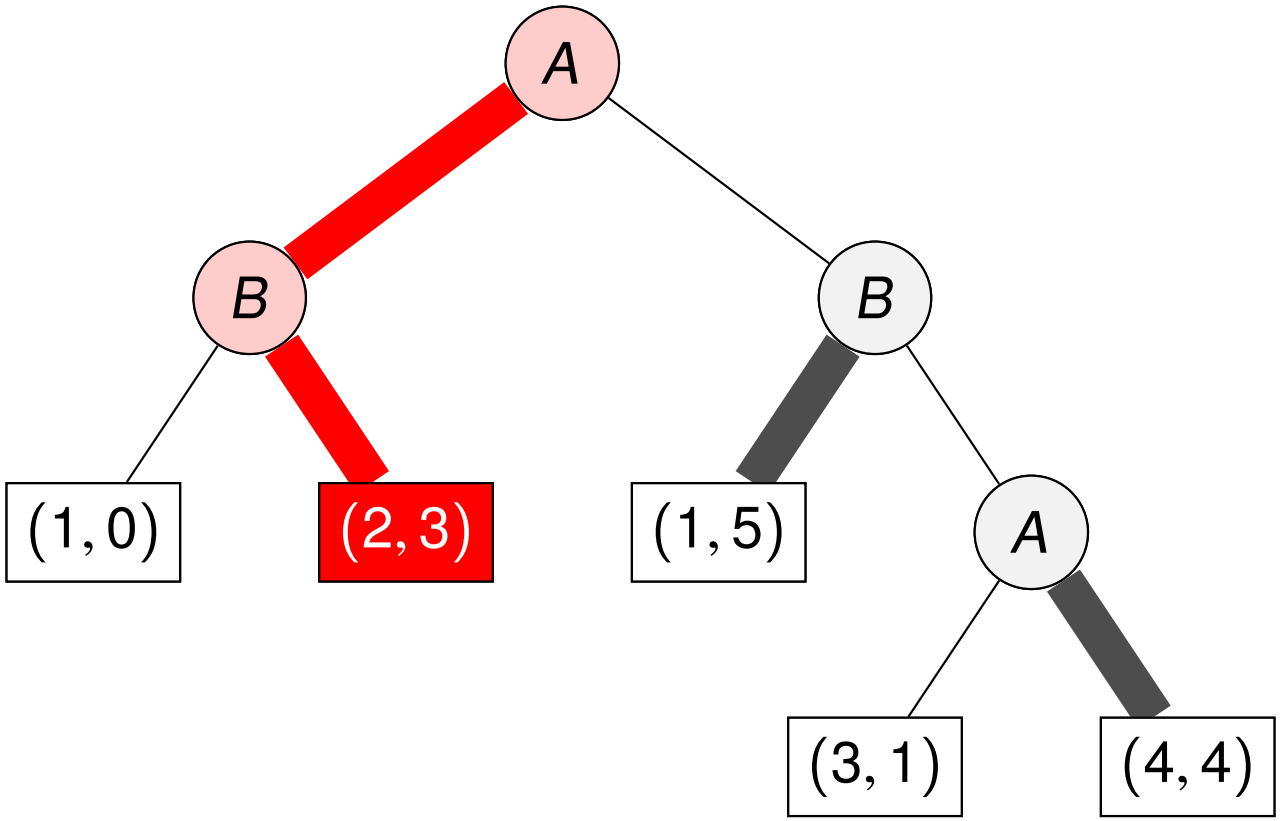
Backward Induction



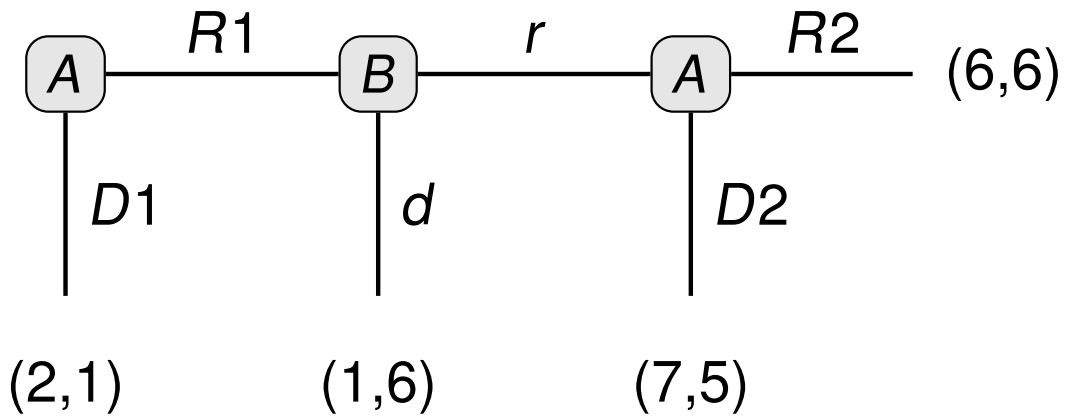
Backward Induction



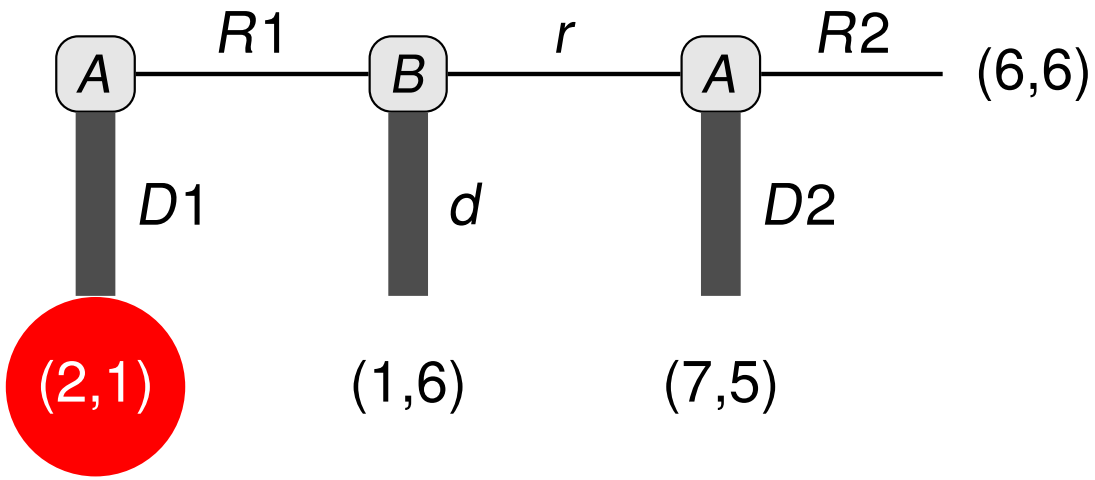
Backward Induction



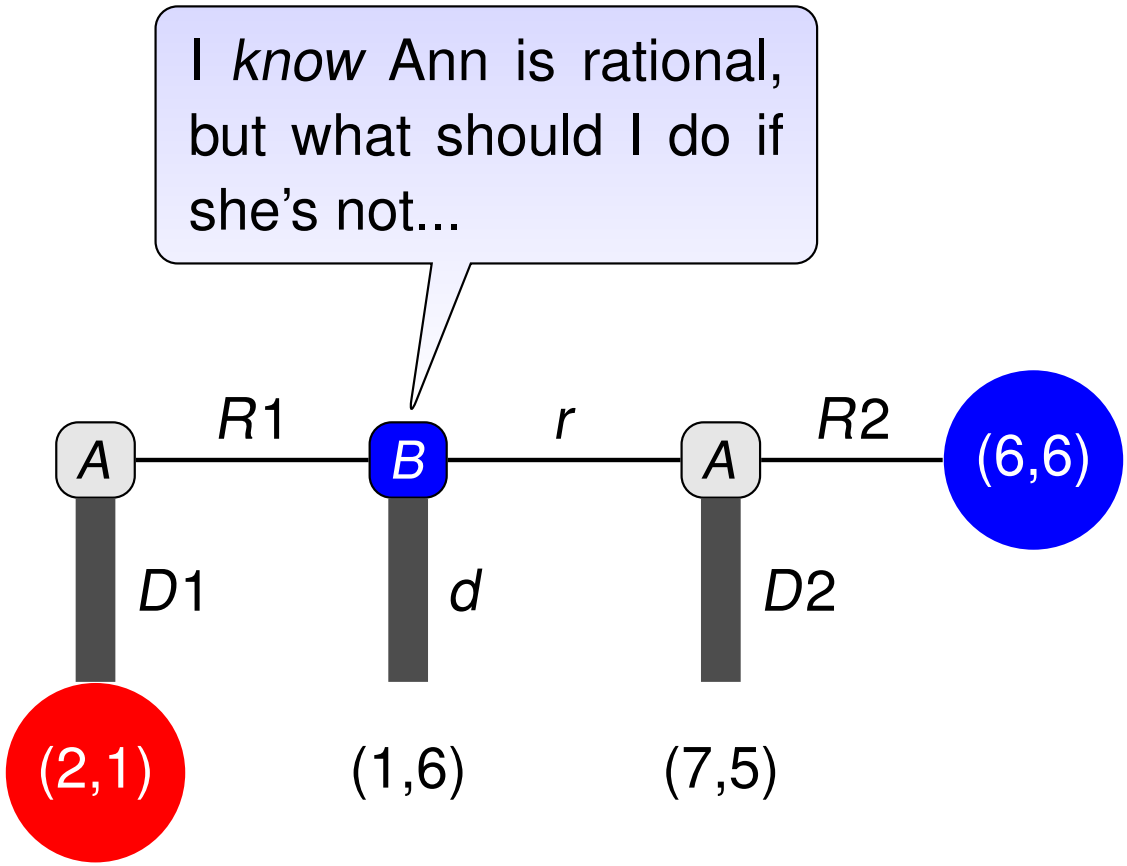
BI Puzzle?

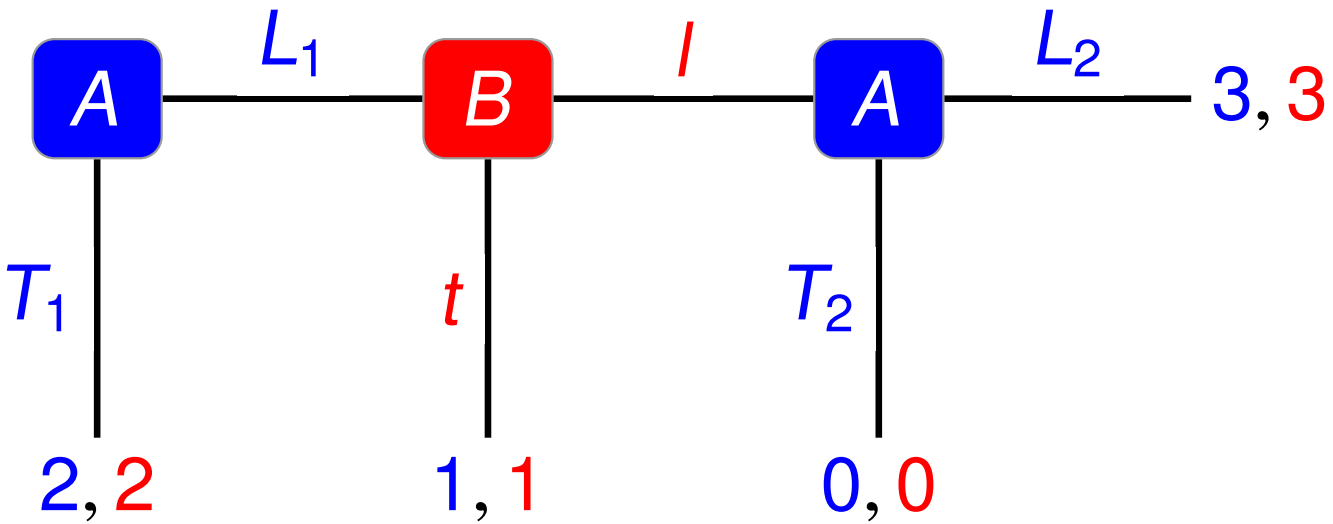


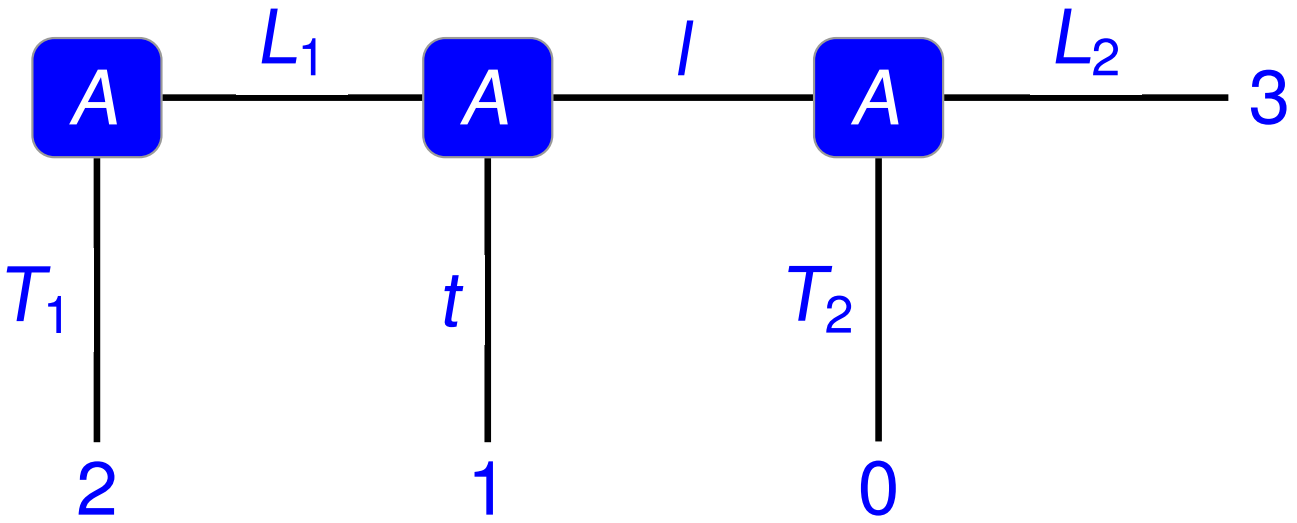
BI Puzzle?

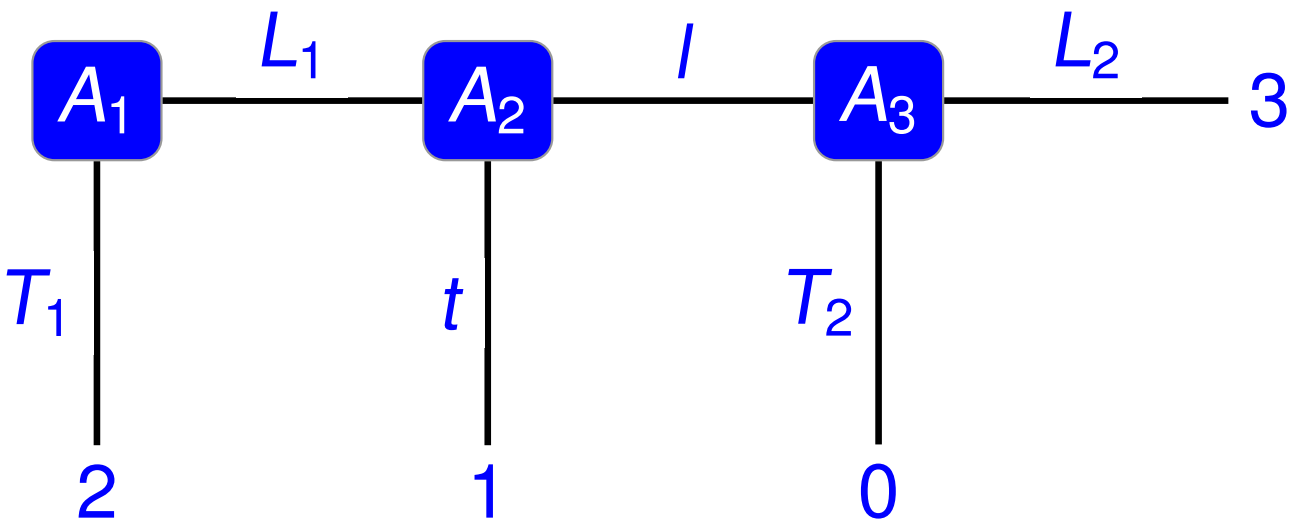


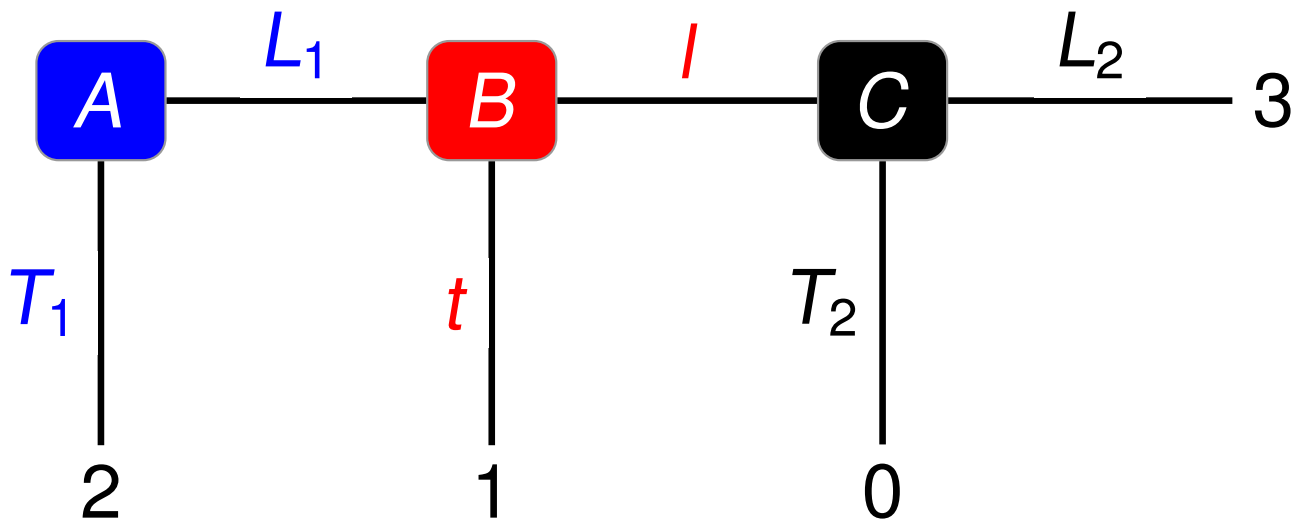
BI Puzzle?

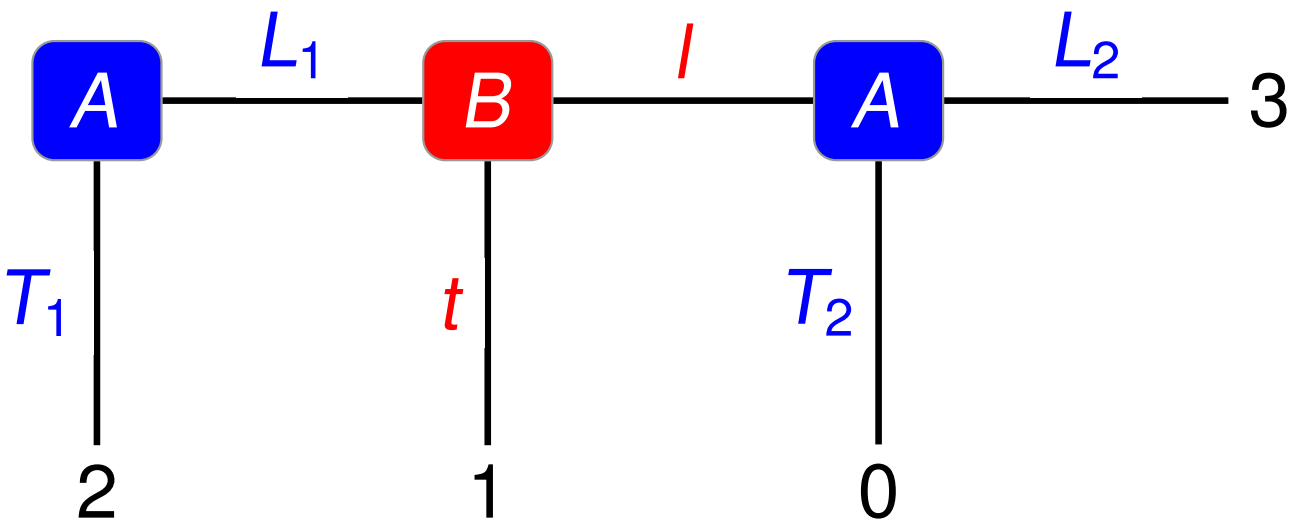

I *know* Ann is rational, but what should I do if she's not...
A
R1
B
r
A
R2
(6,6)
D1
d
D2
(2,1)
(1,6)
(7,5)


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


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


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

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

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

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

Materially Rational: A player *i* is materially rational at a state *w* if every choice actually made is rational.

Substantively Rational: A player *i* is substantively rational at a state *w* if 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: A player *i* is materially rational at a state *w* if every choice actually made is rational.

Substantively Rational: A player *i* is substantively rational at a state *w* if 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.

Theorem (Aumann) In any model, if there is common knowledge that the players are substantively rational at state w, the the backward induction solution is played at w.

Two propositions φ and ψ are epistemically independent for player i in world w iff $P_{i,w}(\varphi \mid \psi) = P_{i,w}(\varphi \mid \neg\psi)$ and $P_{i,w}(\psi \mid \varphi) = P_{i,w}(\psi \mid \neg\varphi)$

A possible belief revision policy: Information about different players should be epistemically independent.

Theorem (Stalnaker's interpretation of Aumann's theorem) Let G be a game of perfect information in agent form (i.e., players only move once) in which for each player different outcomes have different payoffs. Let $\mathcal{M}$ be a model for G in which it is common belief that all agents are perfectly rational, and that all agents adopt belief revision policies that treat information about different agents as epistemically independent. Then in $\mathcal{M}$, the subgame perfect equilibrium strategy profile is realized.

Theorem (Stalnaker's interpretation of Aumann's theorem) Let G be a game of perfect information in agent form (i.e., players only move once) in which for each player different outcomes have different payoffs. Let $\mathcal{M}$ be a model for G in which it is common belief that all agents are perfectly rational, and that all agents adopt belief revision policies that treat information about different agents as epistemically independent. Then in $\mathcal{M}$, the subgame perfect equilibrium strategy profile is realized.

Theorem (Stalnaker's interpretation of Aumann's theorem) Let G be a game of perfect information in agent form (i.e., players only move once) in which for each player different outcomes have different payoffs. Let $\mathcal{M}$ be a model for G in which it is common belief that all agents are perfectly rational, and that all agents adopt belief revision policies that treat information about different agents as epistemically independent. Then in $\mathcal{M}$, the subgame perfect equilibrium strategy profile is realized.

Theorem (Stalnaker's interpretation of Aumann's theorem) Let G be a game of perfect information in agent form (i.e., players only move once) in which for each player different outcomes have different payoffs. Let $\mathcal{M}$ be a model for G in which it is common belief that all agents are perfectly rational, and that all agents adopt belief revision policies that treat information about different agents as epistemically independent. Then in $\mathcal{M}$, the subgame perfect equilibrium strategy profile is realized.

1. Ann cheats — she has seen her opponent's cards.
2. Ann has a losing hand, since I have seen both her hand and her opponent's.
3. Ann is rational.

So, I conclude that she will not bet. But how should I revise my beliefs if I learn that Ann did bet?

1. Ann cheats — she has seen her opponent's cards.
2. Ann has a losing hand, since I have seen both her hand and her opponent's.
3. Ann is rational.

So, I conclude that she will not bet. But how should I revise my beliefs if I learn that Ann did bet?

It may be perfectly reasonable for me to be disposed to give up 2.

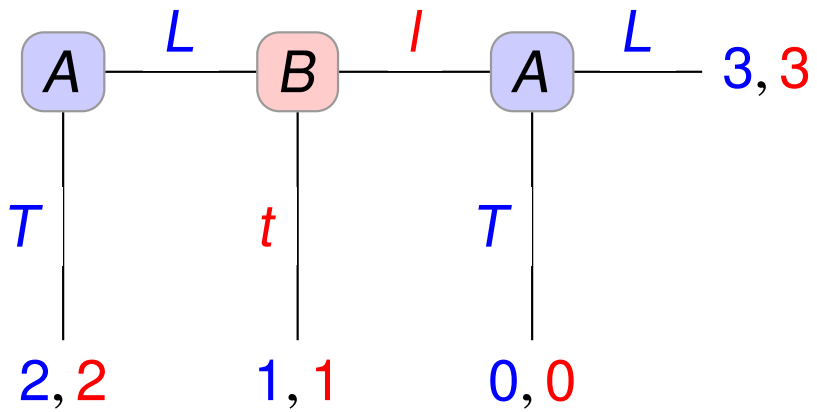
1. Ann cheats — she has seen her opponent's cards.
2. Ann has a losing hand, since I have seen both her hand and her opponent's.
3. Ann is rational.

So, I conclude that she will not bet. But how should I revise my beliefs if I learn that Ann did bet?

It may be perfectly reasonable for me to be disposed to give up 2.

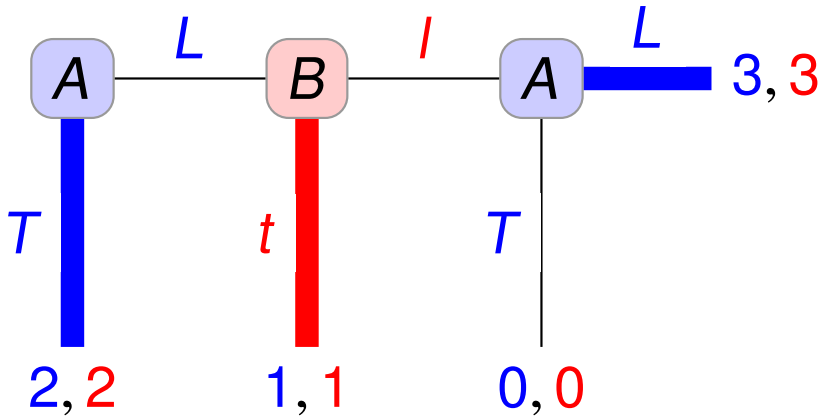
I believe that (1) I Ann *were* to bet, she would lose (since she has a losing hand) and (2) If I were to *learn* that she *did* bet, I would conclude she will win.

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



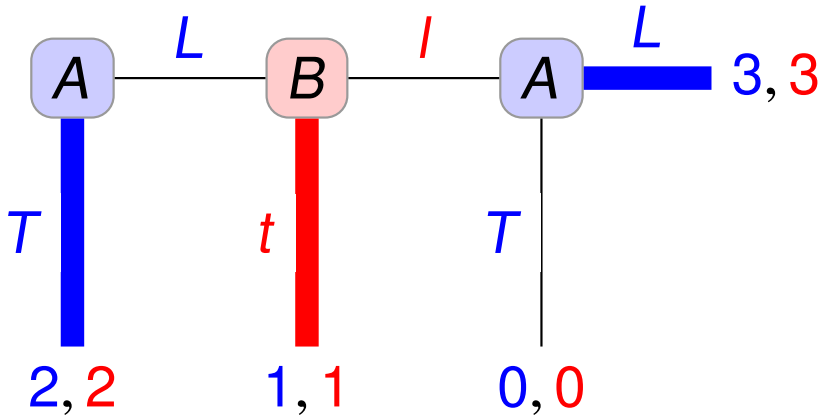
- The backward induction solution is (LL, l)
- Consider a model with a single possible world assigned the profile (TL, t).

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

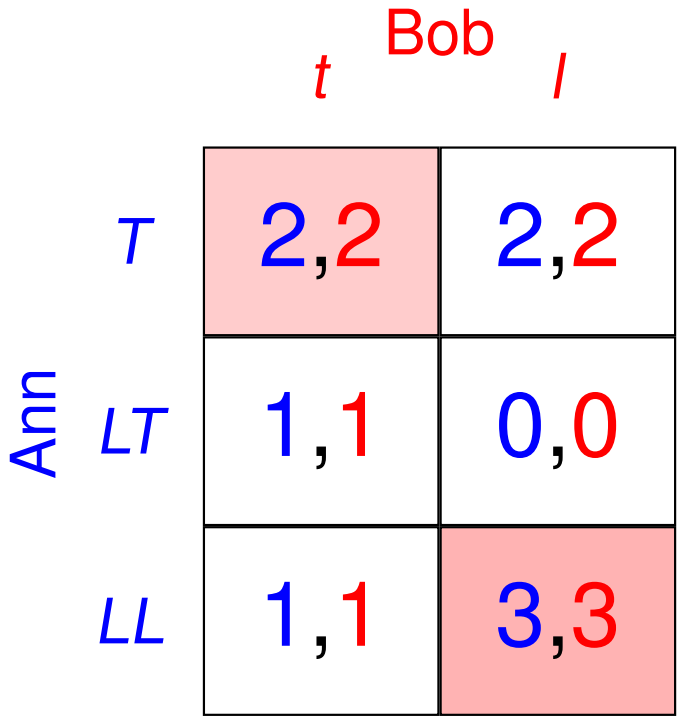


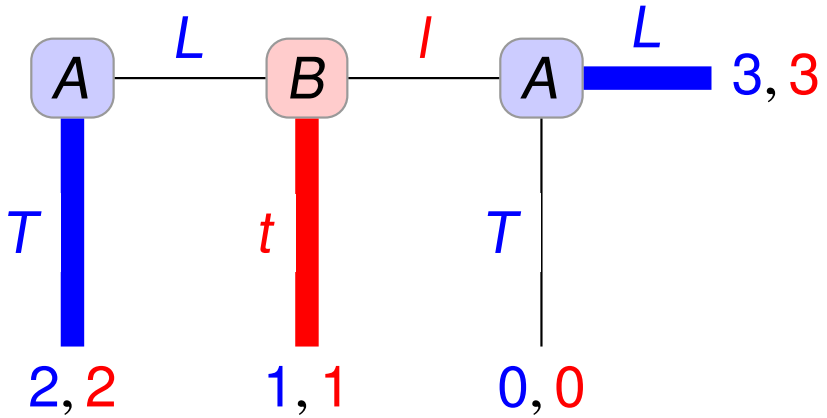
- T is a best response to t, so Ann is materially rational. She is also substantively rational. (Why?)
- Bob doesn't move, so Bob is materially rational. Is he substantively rational?

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



- Is Bob substantively rational? Would t be rational, *if* he had a chance to act?
- Suppose that Bob is disposed to revise his beliefs in such a way that if Ann acted irrationally once, she will act irrationally later in the game.





- 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.

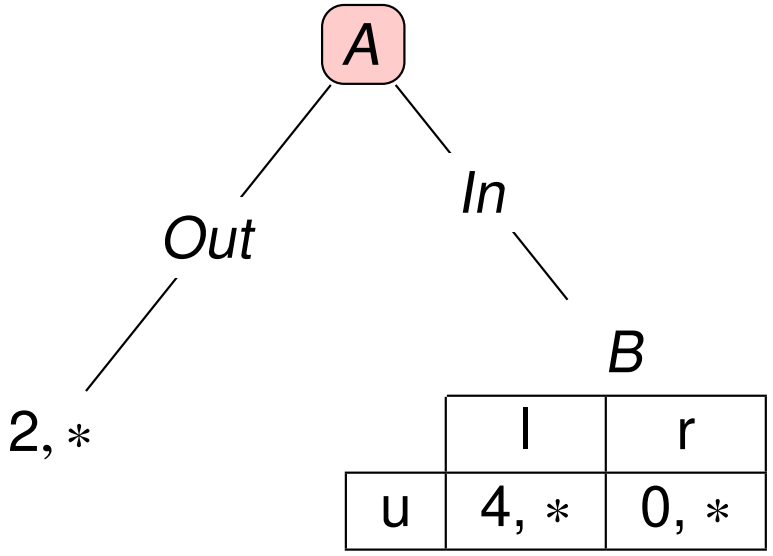
Informal characterizations of BI

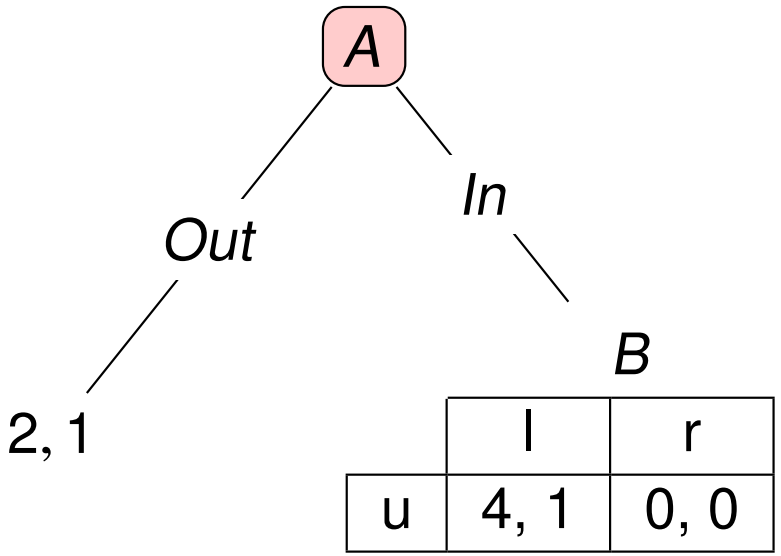
- Future choices are *epistemically independent* of any observed behavior
- Any “off-equilibrium” choice is interpreted simply as a mistake (which will not be repeated)
- At each choice point in a game, the players only reason about future paths

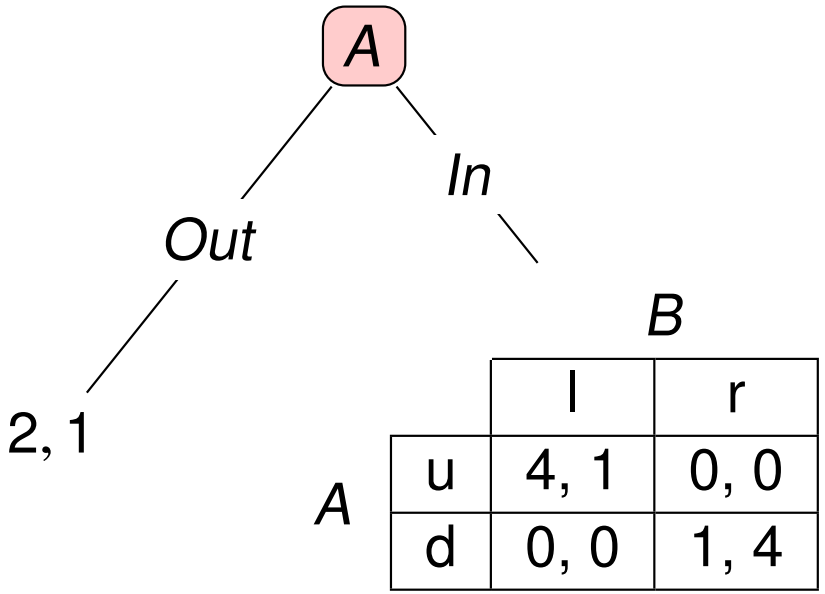
Rationalizing Observed Actions

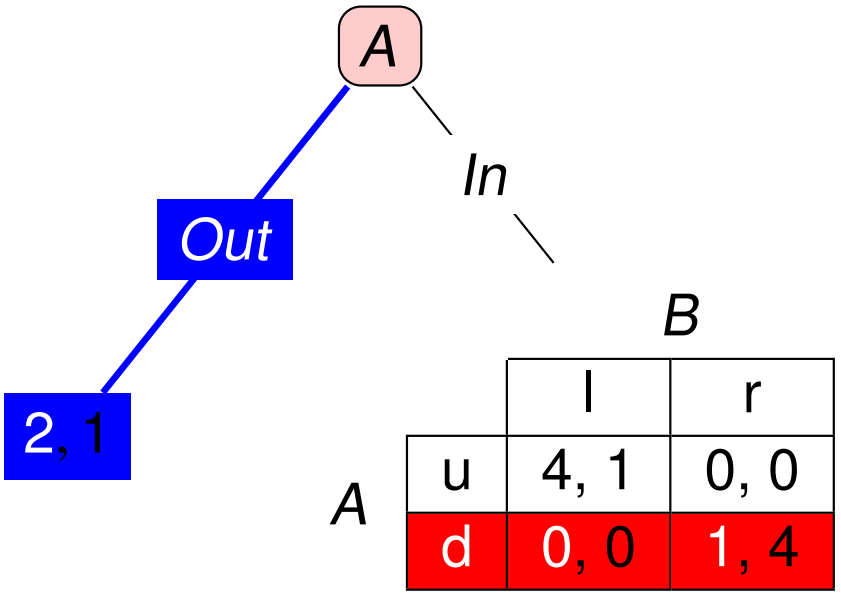
After observing an (unexpected) move by some player, you could:

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

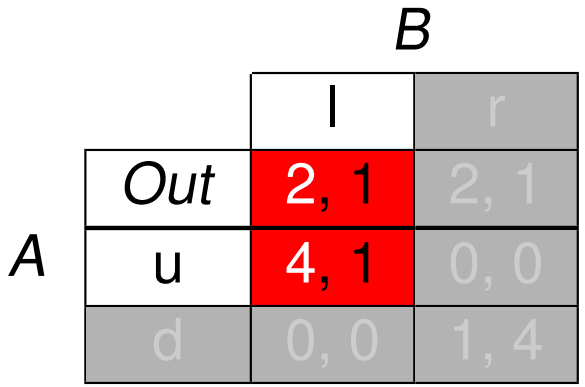

A
Out
In
2, 1
B
l
r
A
u
4, 1
0, 0
d
0, 0
1, 4

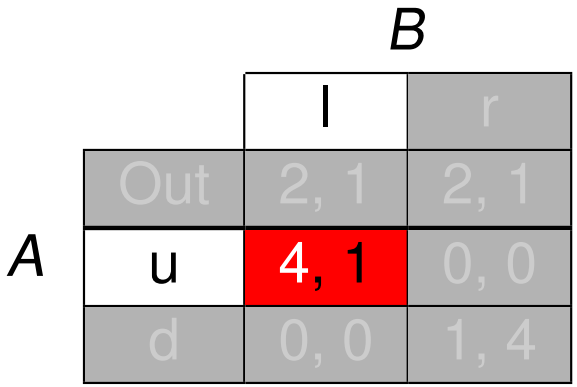

A
Out
In
2, 1
B
l
r
A
u
4, 1
0, 0
d
0, 0
1, 4

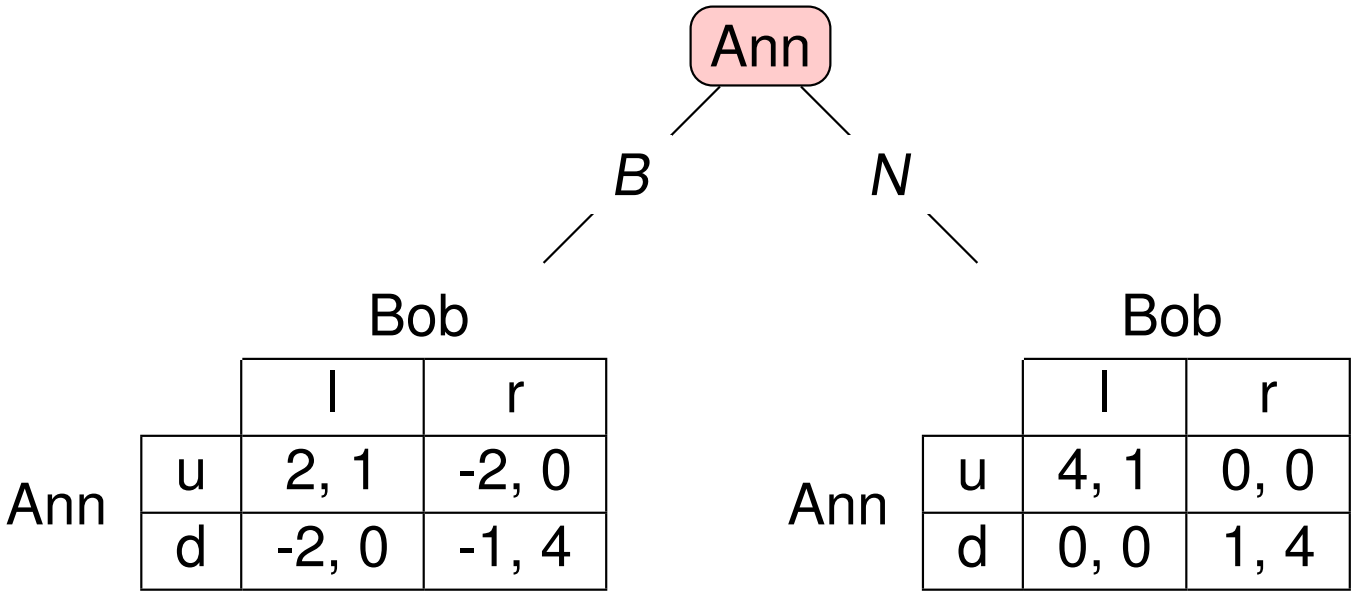
		B	
		l	r
A	*Out*	2, 1	2, 1
	u	4, 1	0, 0
	d	0, 0	1, 4

		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
Out
2, 1
2, 1
A
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

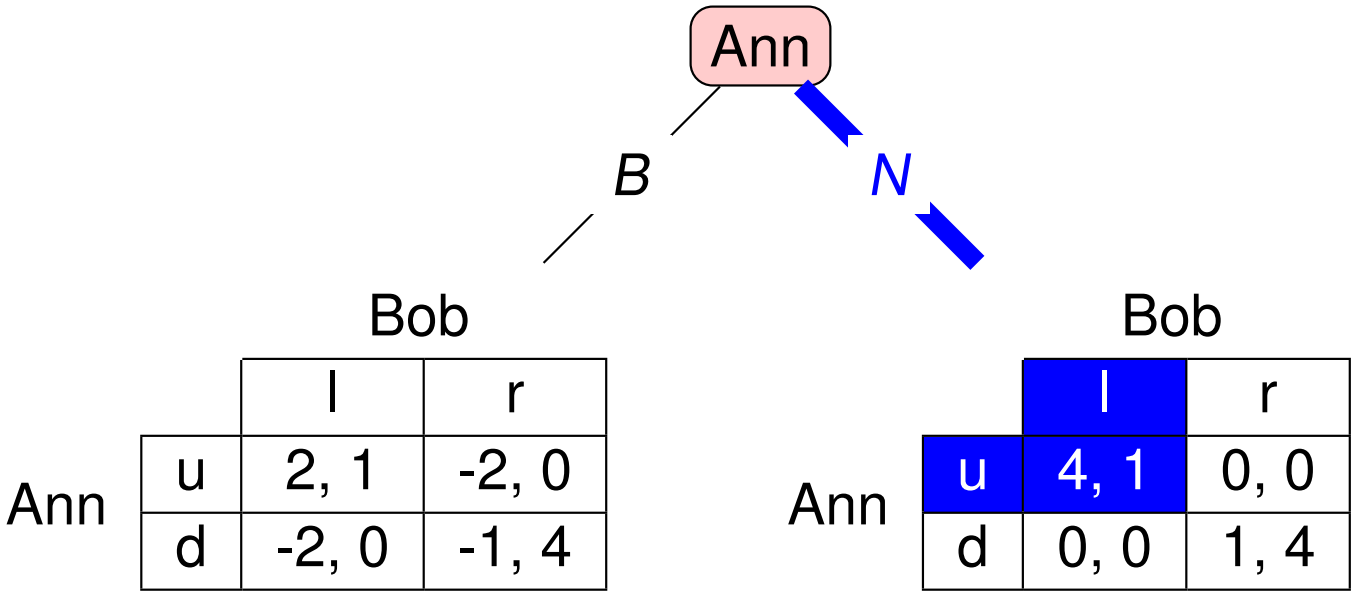
		Bob			
		ll	lr	rl	rr
Ann	Bu	2, 1	2, 1	-2, 0	-2, 0
	Bd	-2, 0	-2, 0	-1, 4	-1, 4
	Nu	4, 1	0, 0	4, 1	0, 0
	Nd	0, 0	1, 4	0, 0	1, 4

		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.

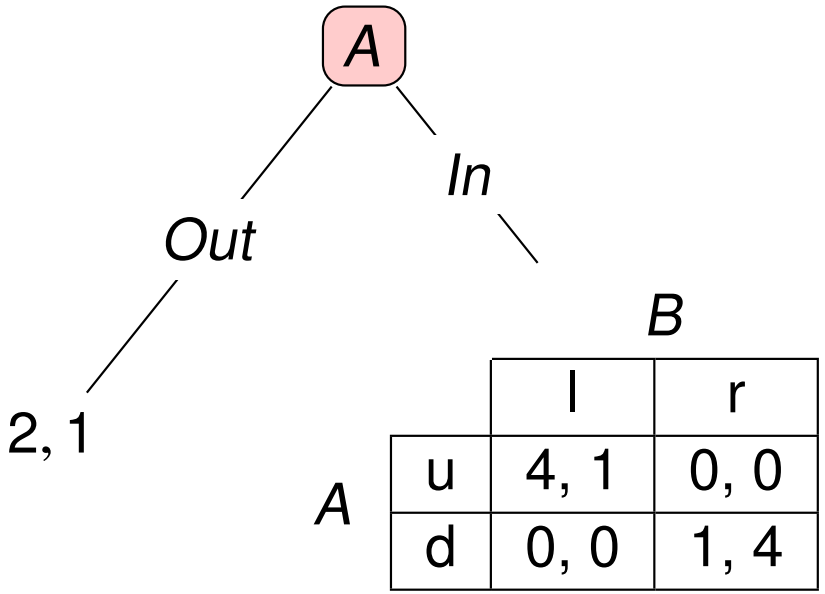
Issues

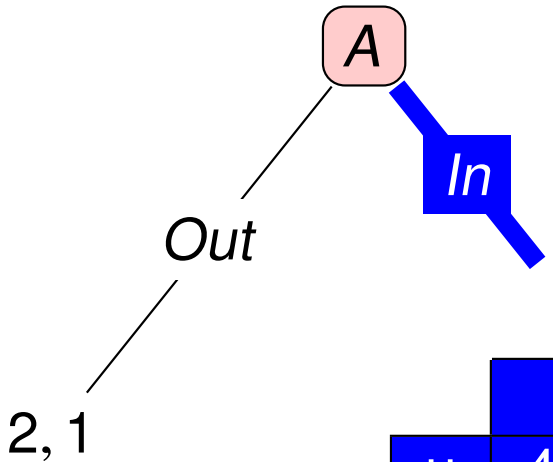
- The players' conditional beliefs must be *rich enough* to employ the forward induction principle.
- Do the players robustly believe the forward induction principle?
- Can players become more/less confident in the forward induction principle?

"...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.

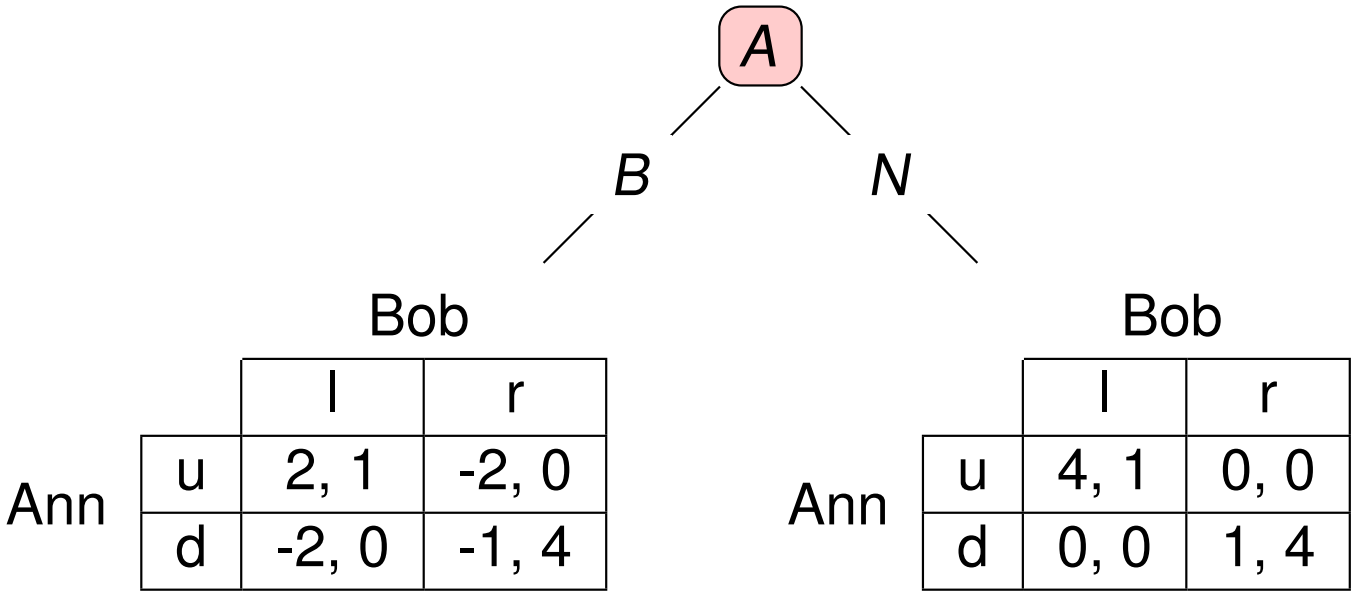
"...in general, a player's beliefs about what another player will do are based on an inference from two other kinds of beliefs: beliefs about the passive beliefs of that player, and beliefs about her rationality. If one's prediction based on these beliefs is defeated, one must choose whether to revise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

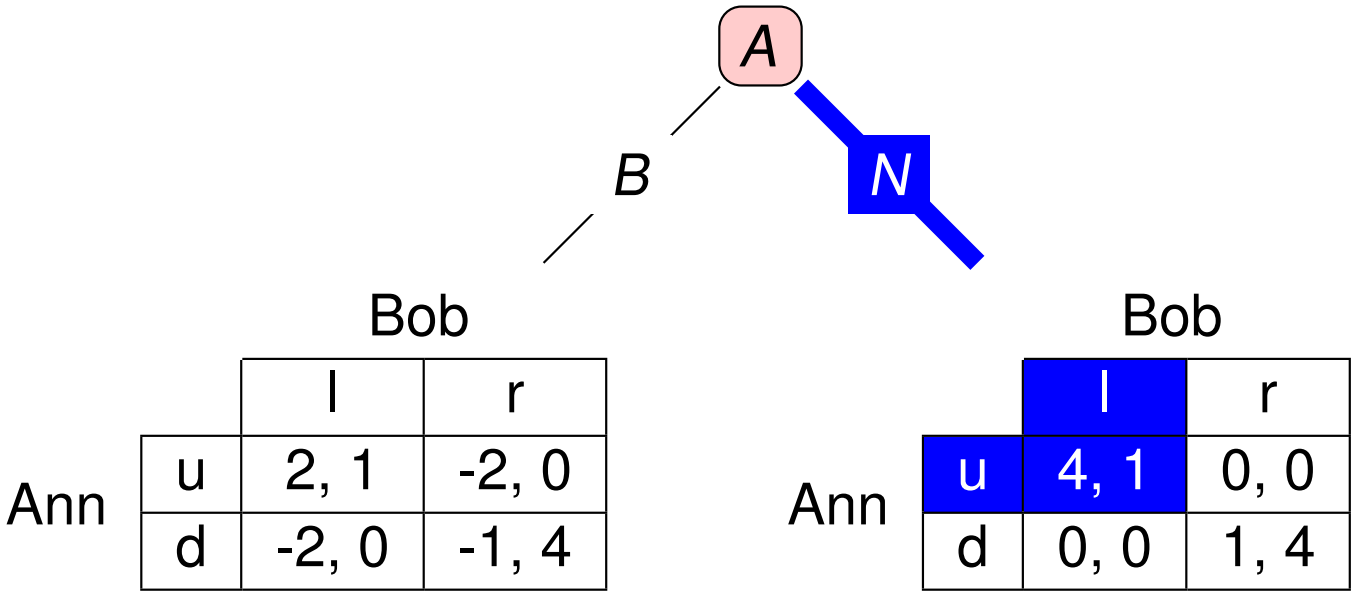
"...in general, a player's beliefs about what another player will do are based on an inference from two other kinds of beliefs: beliefs about the passive beliefs of that player, and beliefs about her rationality. If one's prediction based on these beliefs is defeated, one must choose whether to revise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


A
Out
In
2, 1
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


A
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


A
B
N
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.

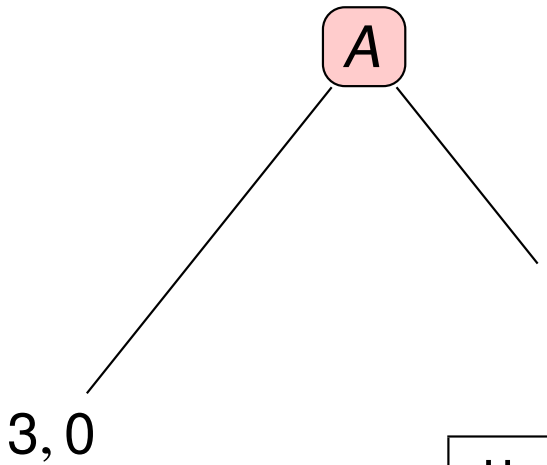
Algorithm and a "Theorem"

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

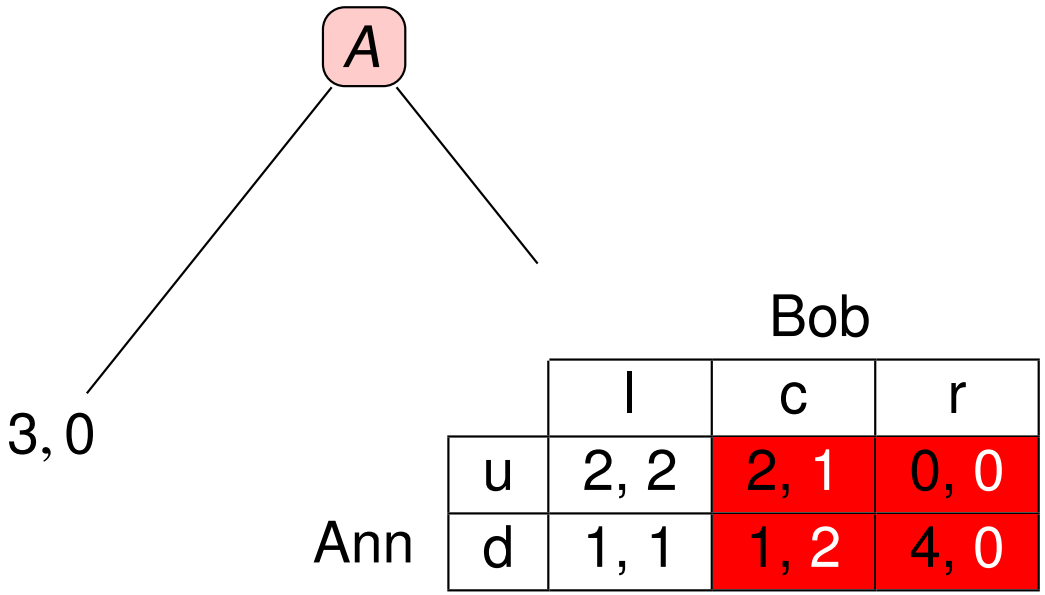
Backward *versus* Forward Induction



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

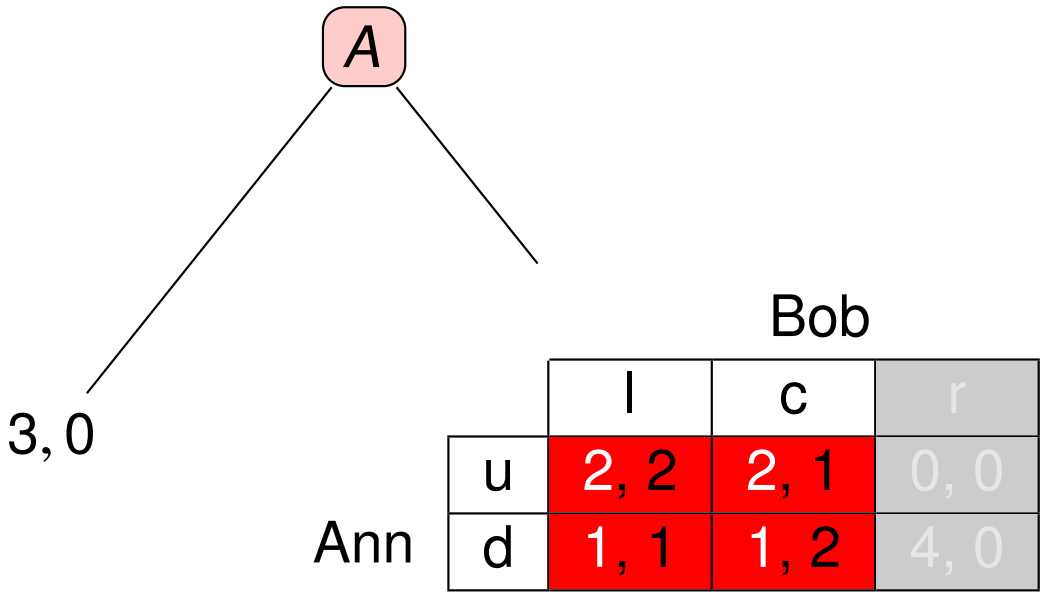
A. Perea. *Backward Induction* versus *Forward Induction Reasoning*. Games, 1, pgs. 168 - 188, 2010.

Backward *versus* Forward Induction



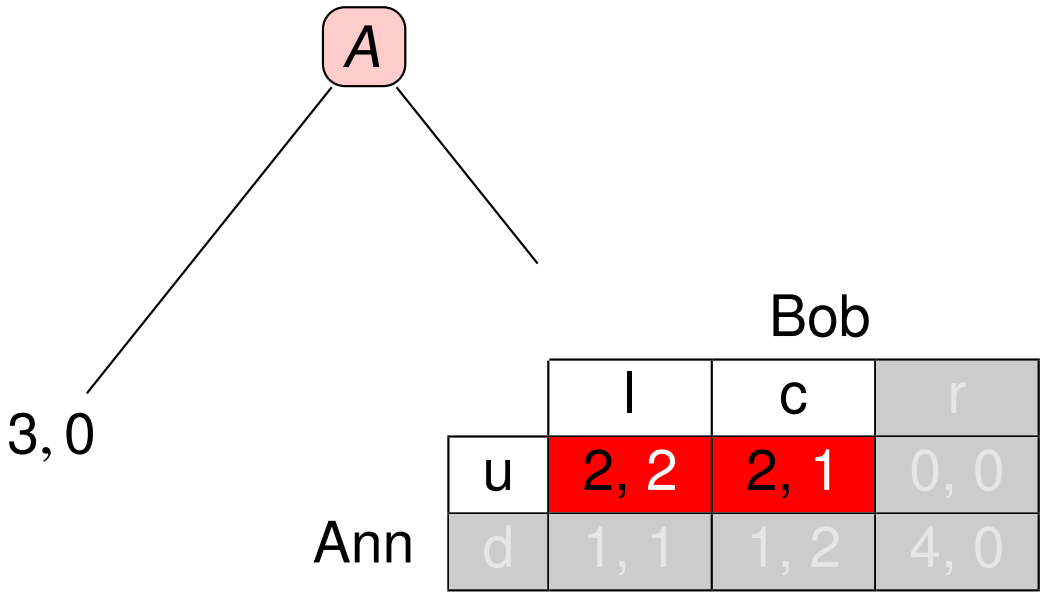
A. Perea. *Backward Induction* versus *Forward Induction Reasoning*. Games, 1, pgs. 168 - 188, 2010.

Backward *versus* Forward Induction



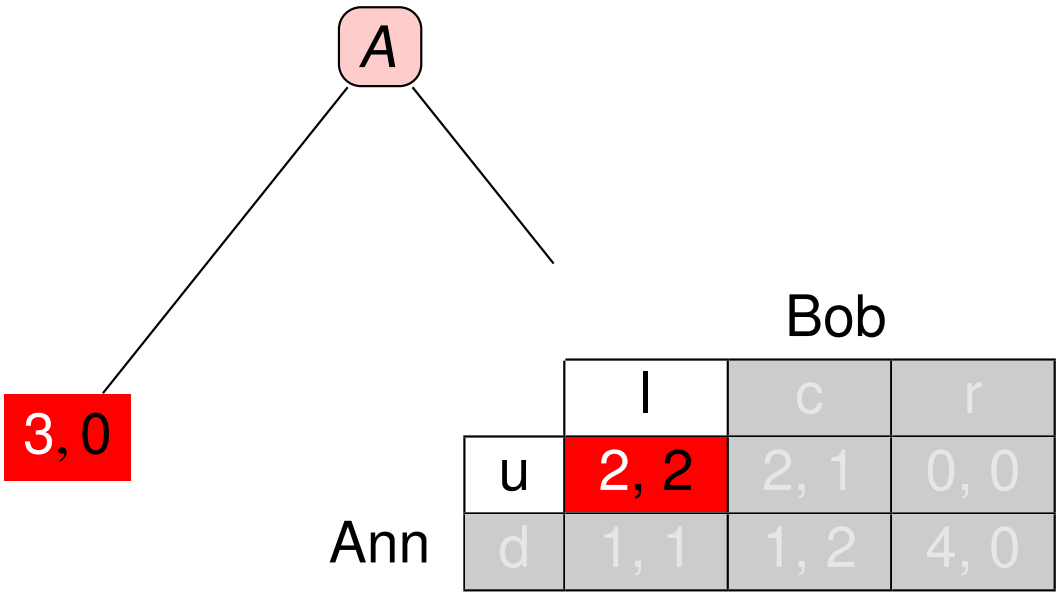
A. Perea. *Backward Induction* versus *Forward Induction Reasoning*. Games, 1, pgs. 168 - 188, 2010.

Backward *versus* Forward Induction



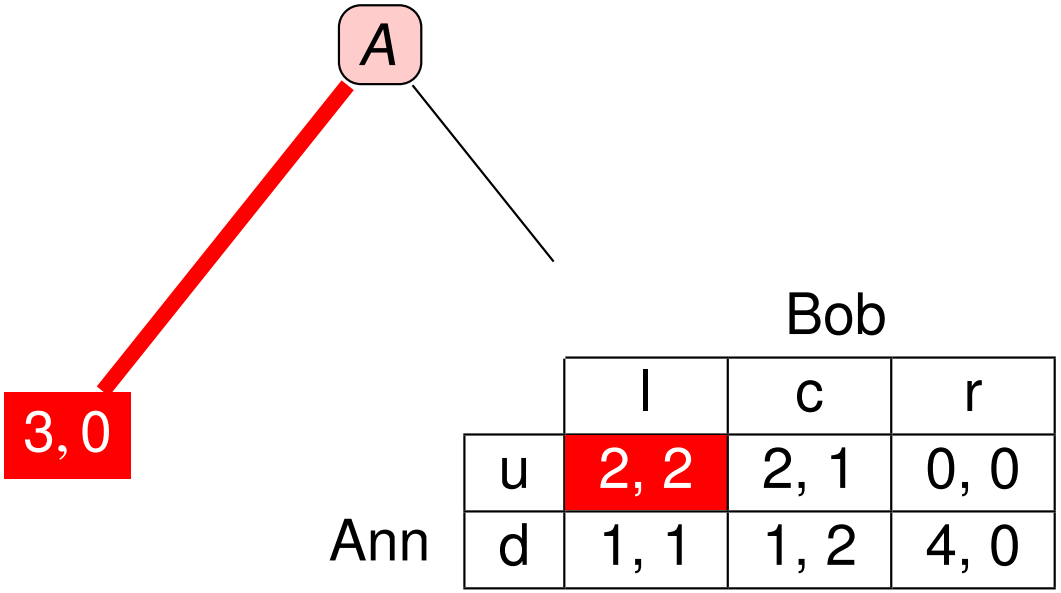
A. Perea. *Backward Induction* versus *Forward Induction Reasoning*. Games, 1, pgs. 168 - 188, 2010.

Backward *versus* Forward Induction



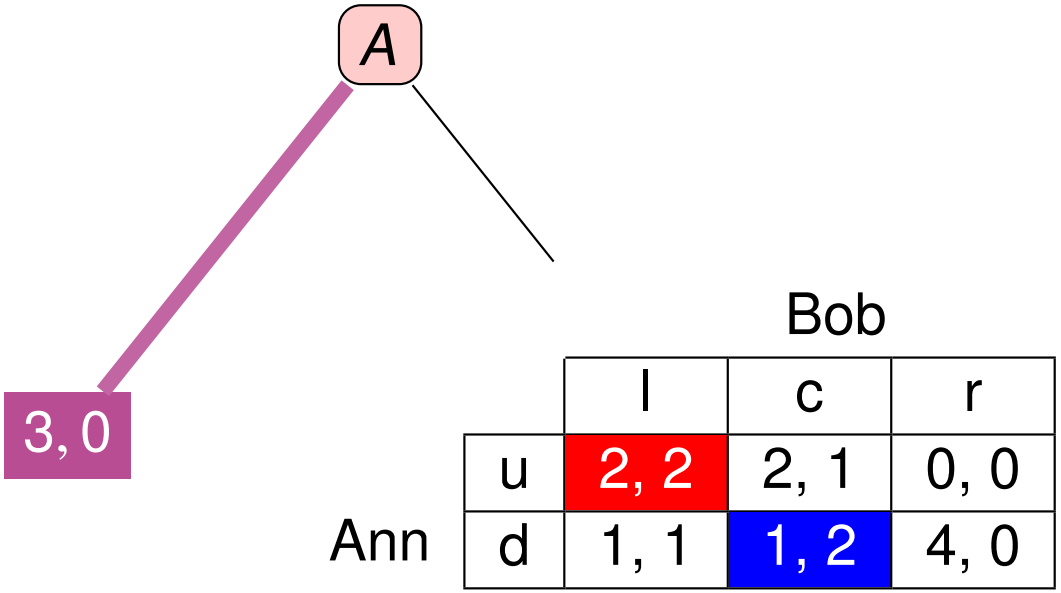
A. Perea. *Backward Induction* versus *Forward Induction Reasoning*. Games, 1, pgs. 168 - 188, 2010.

Backward *versus* Forward Induction



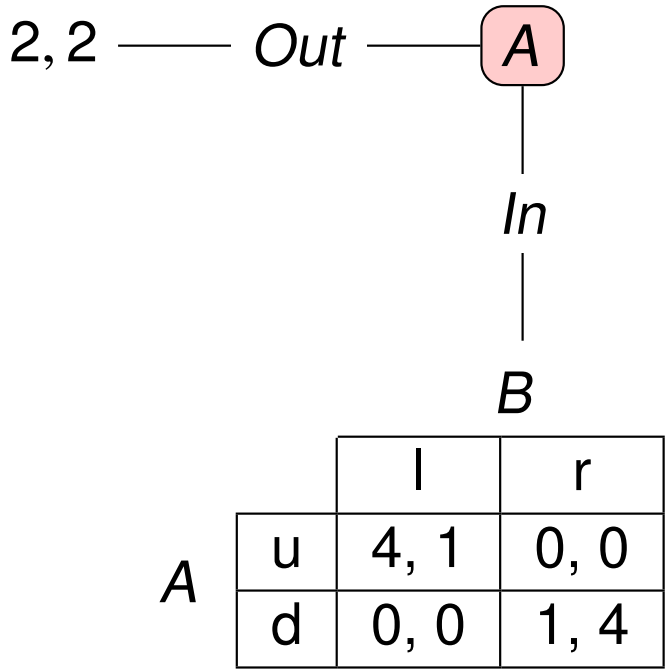
A. Perea. *Backward Induction* versus *Forward Induction Reasoning*. Games, 1, pgs. 168 - 188, 2010.

Backward *versus* Forward Induction



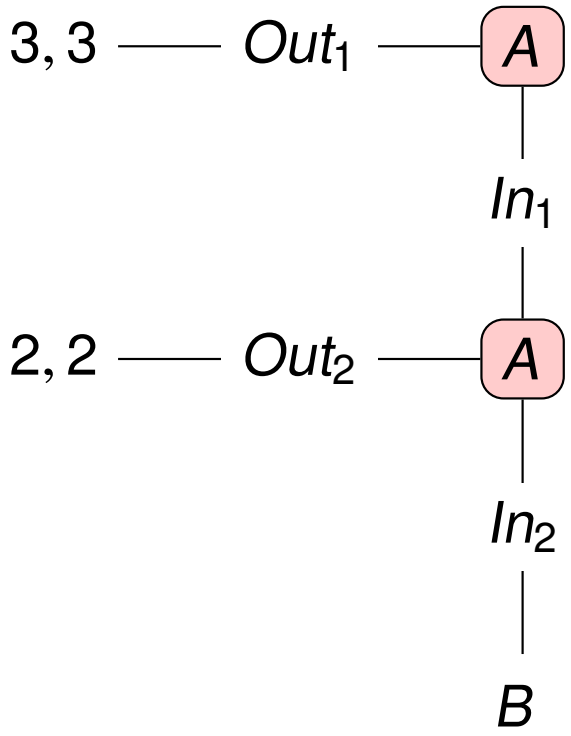
A. Perea. *Backward Induction* versus *Forward Induction Reasoning*. Games, 1, pgs. 168 - 188, 2010.

Rationalization *versus* Mistakes



A. Knoks and EP. *Deliberating between Backward and Forward Induction: First Steps*. TARK, 2015.

Rationalization *versus* Mistakes



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

A. Knoks and EP. *Deliberating between Backward and Forward Induction: First Steps*. TARK, 2015.

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, Knoks & EP)
- How do (should) players choose between the two different styles of reasoning about games? (Perea)

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

Concluding Remarks

Have we captured *strategic reasoning*?

Strategic reasoning

- 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
- . . .

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.

"...[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.

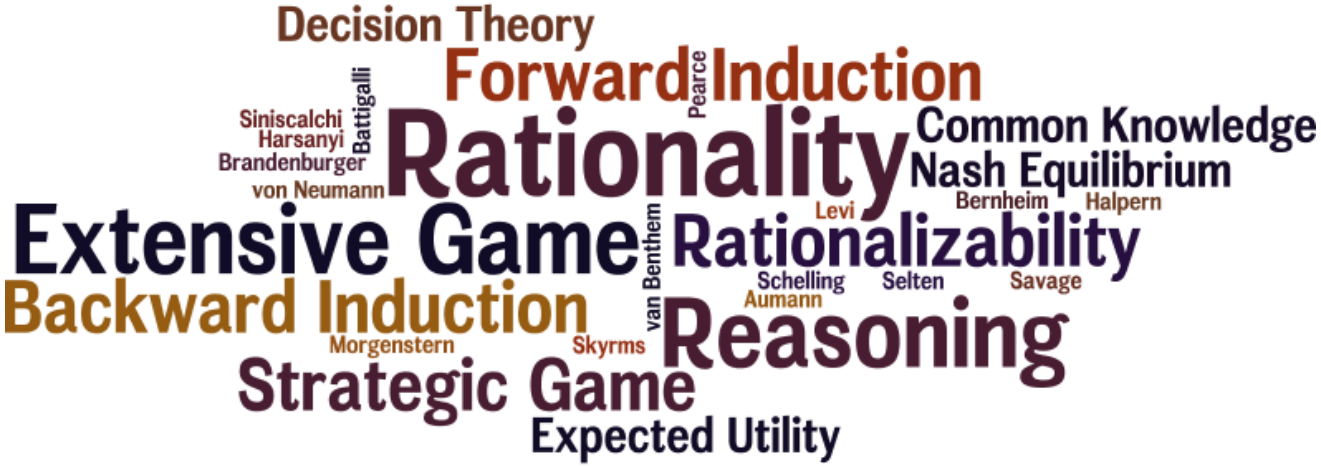
In fact, I find it somewhat surprising that our disciplines have any relation at all to real behavior. (I hope that most readers will agree that there is indeed such a relation, that we do gain some insight into the behavior of *Homo sapiens* by studying *Homo rationalis*.)"

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

Plan

- ✓ Day 1: Decision Theory
- ✓ Day 2: From Decisions to Games
- ✓ Day 3: Game Models
- ✓ Day 4: Modeling Deliberation (in Games)
- ✓ Day 5: Backward and Forward Induction, Concluding Remarks

Thank you!


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