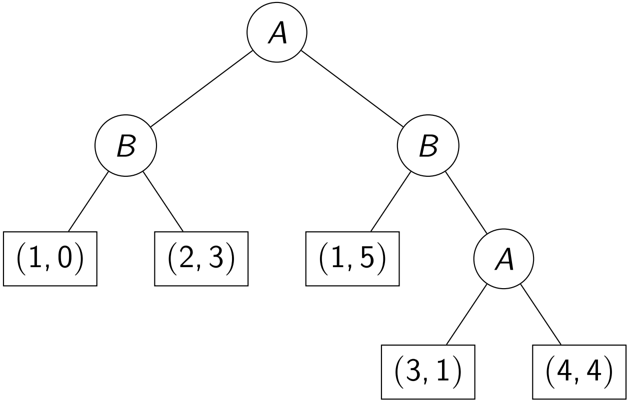
# Conditionals in Game Theory

Ilaria Canavotto, University of Maryland
Eric Pacuit, University of Maryland

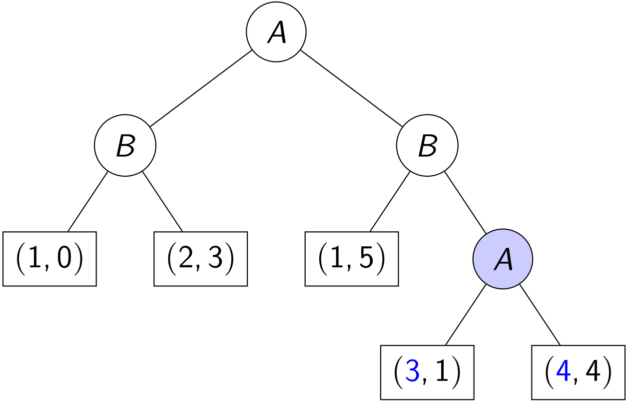
Lecture 4

ESSLLI 2022

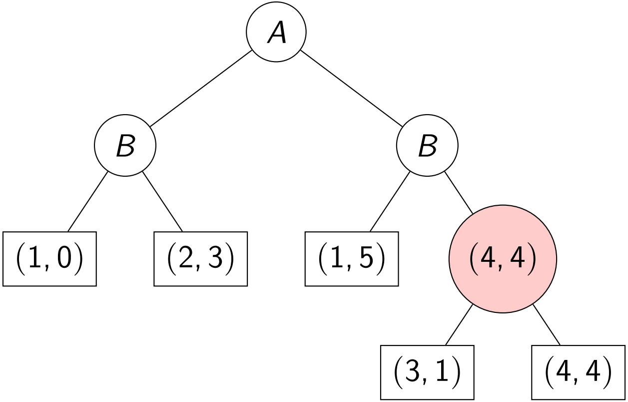
# Extensive Games and Backwards Induction



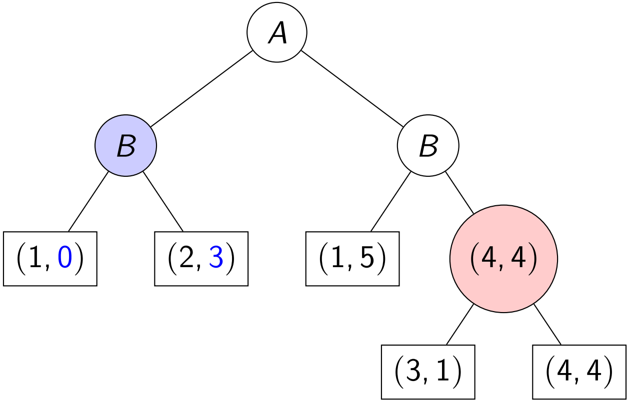
# Extensive Games and Backwards Induction



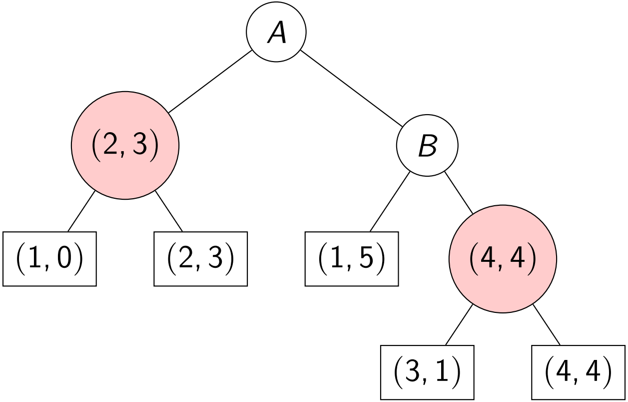
# Extensive Games and Backwards Induction



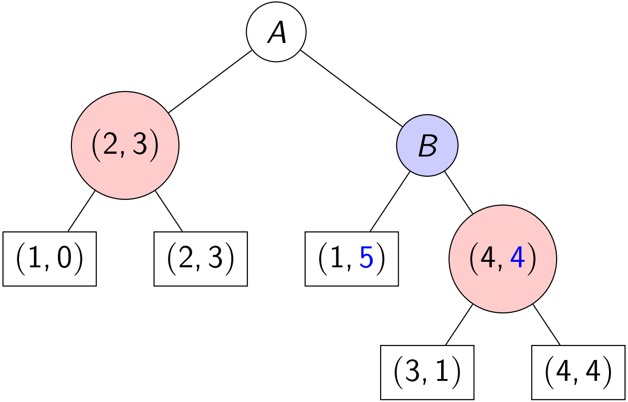
# Extensive Games and Backwards Induction



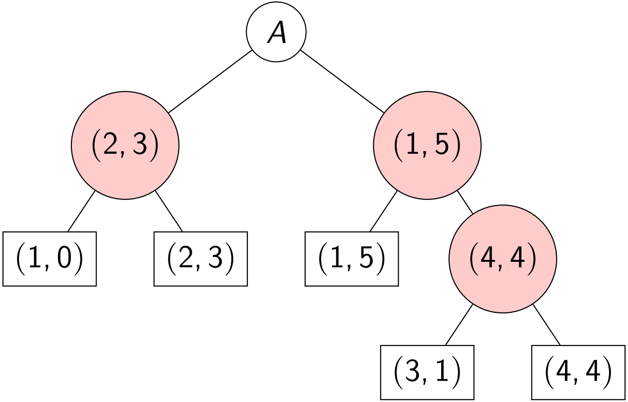
# Extensive Games and Backwards Induction



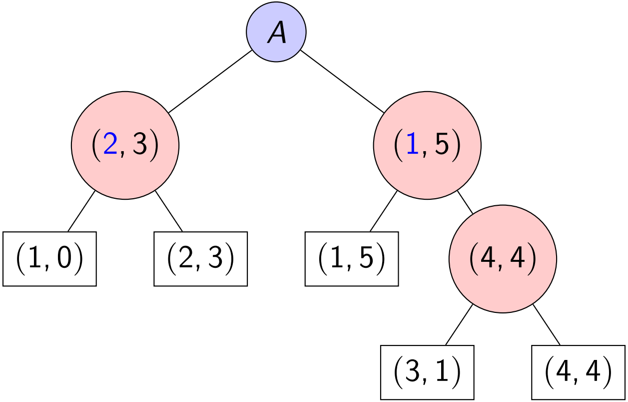
# Extensive Games and Backwards Induction



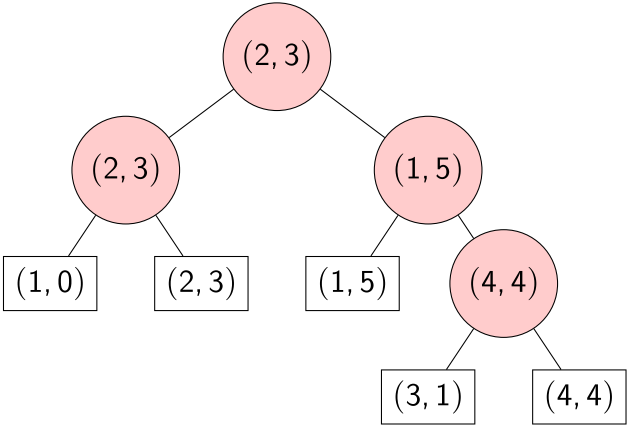
# Extensive Games and Backwards Induction



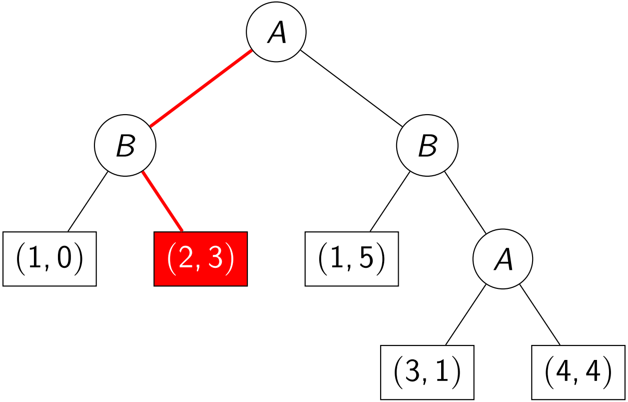
# Extensive Games and Backwards Induction



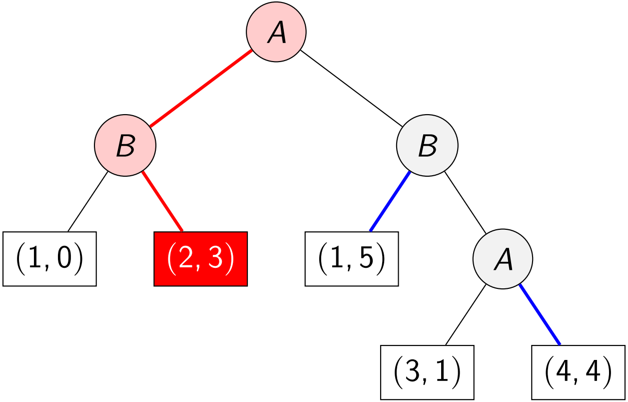
# Extensive Games and Backwards Induction



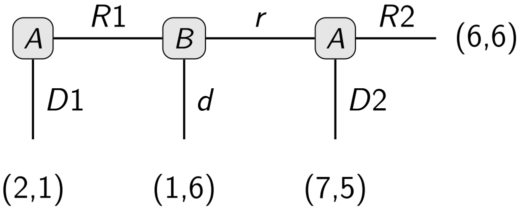
# Extensive Games and Backwards Induction



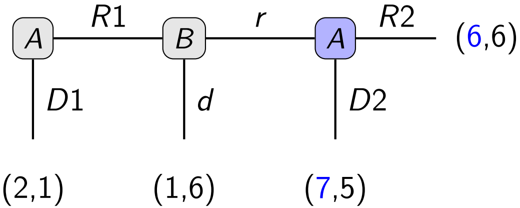
# Extensive Games and Backwards Induction



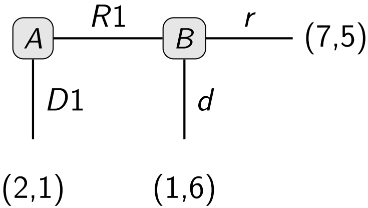
# BI Puzzle, I



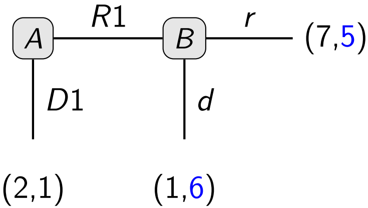
# BI Puzzle, I



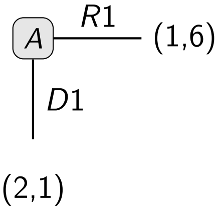
# BI Puzzle, I



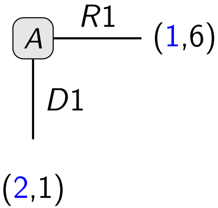
# BI Puzzle, I



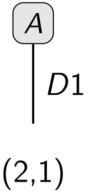
# BI Puzzle, I



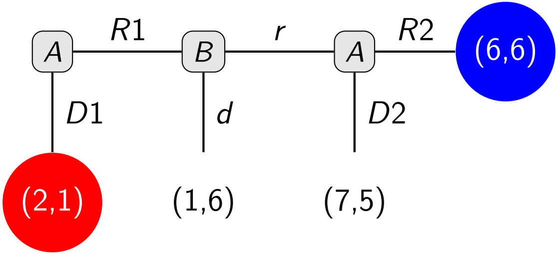
# BI Puzzle, I



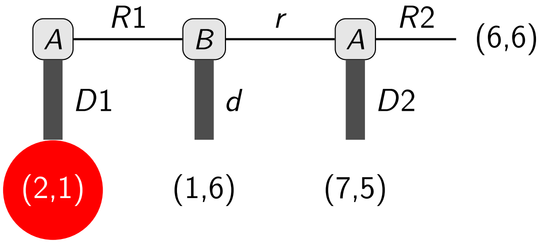
# BI Puzzle, I



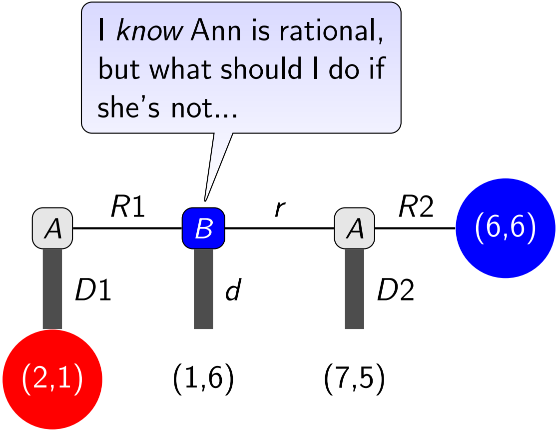
# BI Puzzle, I

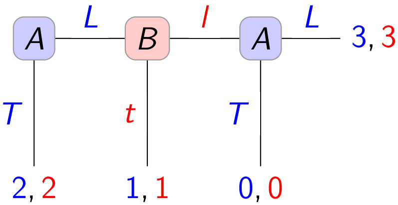


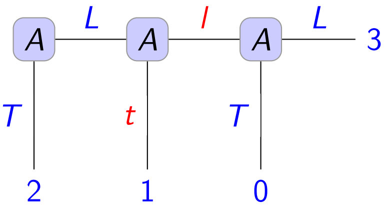
# BI Puzzle?

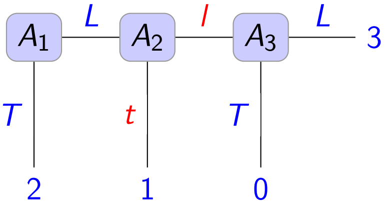


# BI Puzzle?

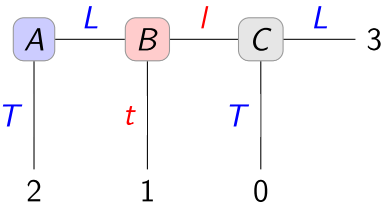

I *know* Ann is rational, but what should I do if she's not...
A
R1
B
r
A
R2
(6,6)
D1
d
D2
(2,1)
(1,6)
(7,5)

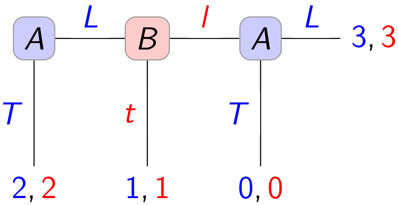

A
L
B
l
A
L
3, 3
T
t
T
2, 2
1, 1
0, 0


A
A
A
L
l
L
3
T
t
T
2
1
0

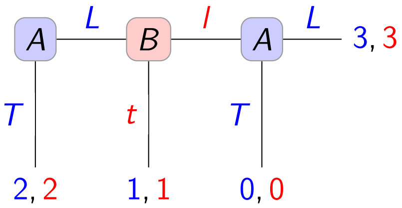


$A_1$
$A_2$
$A_3$
$L$
$l$
$L$
3
$T$
$t$
$T$
2
1
0


A
B
C
L
l
L
3
T
t
T
2
1
0


A
L
B
l
A
L
3, 3
T
t
T
2, 2
1, 1
0, 0

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

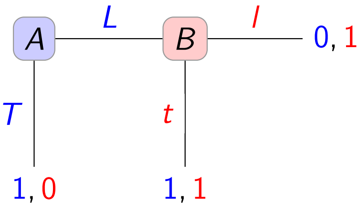


R. Aumann. *Backwards induction and common knowledge of rationality*. Games and Economic Behavior, 8, pgs. 6 - 19, 1995.

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

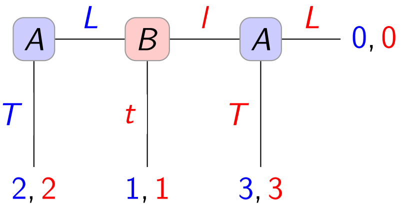
J. Halpern. *Substantive Rationality and Backward Induction*. Games and Economic Behavior, 37, pp. 425-435, 1998.

|  |  | Bob |  |
|---|---|---|---|
|  |  | $t$ | $l$ |
| Ann | $T$ | 1,0 | 1,0 |
|  | $L$ | 1,1 | 0,1 |




- The strategies of both players are rationalizable.
- Only $T$ is *perfectly rational* for Ann and $t$ is *perfectly rational* for Bob.

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



**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.

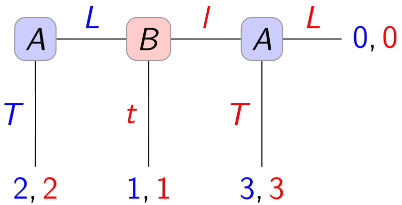
**Materially Rational**: every choice actually made is optimal (i.e., maximizes subjective expected utility).

**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.

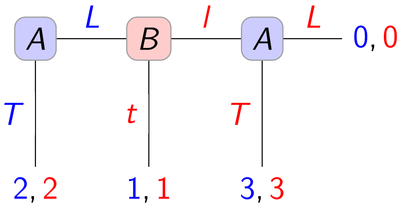
Bob

| Ann | $t$ | $l$ |
|---|---|---|
| $T$ | 2,2 | 2,2 |
| $LT$ | 1,1 | 3,3 |
| $LL$ | 1,1 | 0,0 |

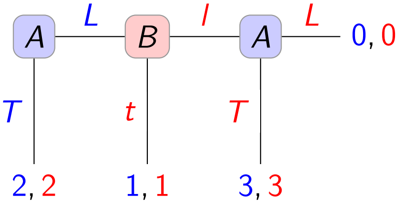



- Suppose that Bob believes that Ann will choose $T$ with probability 1; what should he do? This depends on what he thinks Ann would on the hypothesis that his belief about her is mistaken.
- Suppose that if Bob were surprised by her, then he concludes she is irrational, selecting $L$ on her second move. Bob's choice of $t$ is perfectly rational.

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




- Suppose Ann is sure that Bob will choose $t$, which is the only perfectly rational choice for Bob. Then, Ann's only rational choice is $T$.
- So, it might be that Ann and Bob both know each other's beliefs about each other, and are both perfectly rational, but they still fail to coordinate on the optimal outcome for both.




- Perhaps if Bob believed that Ann would choose $L$ are her second move then he wouldn't believe she was fully rational, *but it is not suggested that he believes this*.
- Divide Ann's strategy $T$ into two $TT$: $T$, and I *would* choose $T$ again on the second move if I were faced with that choice" and $TL$: "$T$, but I *would* choose $L$ on the second move..."
- Of these two only $TT$ is rational
- But if Bob **learned he was wrong**, he would conclude she is playing $LL$.

“To think there is something incoherent about this combination of beliefs and belief revision policy is to confuse epistemic with causal counterfactuals—it would be like thinking that because I believe that if Shakespeare hadn’t written Hamlet, it would have never been written by anyone, I must therefore be disposed to conclude that Hamlet was never written, were I to learn that Shakespeare was in fact not its author”

(pg. 152, Stalnaker)

1. General Smith is a shrewd judge of character—he knows (better than I) who is brave and who is not.
2. The general sends only brave men into battle.
3. Private Jones is cowardly.

I believe that (1) Jones would run away if he were sent into battle and (2) if Jones *is* sent into battle, then he won't run away.

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.

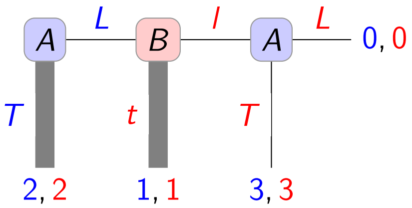
1. Ann cheats — she has seen her opponent's cards.
2. Ann has a losing hand, since I have seen both her hand and her opponent's.
3. Ann is rational.

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.

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



**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 $w$.

**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.

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

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.

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

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.

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

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.

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

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

# Models of Extensive Games

Let $\Gamma$ be a *non-degenerate* extensive game with perfect information. Let $\Gamma_i$ be the set of nodes controlled by player $i$.

# Models of Extensive Games

Let $\Gamma$ be a *non-degenerate* extensive game with perfect information. Let $\Gamma_i$ be the set of nodes controlled by player $i$.

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

# Models of Extensive Games

Let $\Gamma$ be a *non-degenerate* extensive game with perfect information. Let $\Gamma_i$ be the set of nodes controlled by player $i$.

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

Given a vertex $v$ in $\Gamma$ and strategy profile $s$, $s$ specifies a unique path from $v$ to an end-node.

# Models of Extensive Games

Let $\Gamma$ be a *non-degenerate* extensive game with perfect information. Let $\Gamma_i$ be the set of nodes controlled by player $i$.

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

Given a vertex $v$ in $\Gamma$ and strategy profile $s$, $s$ specifies a unique path from $v$ to an end-node.

$\mathcal{M}(\Gamma) = \langle W, (I_i)_{i \in N}, \sigma \rangle$ where $\sigma : W \to \mathit{Strat}(\Gamma)$ and $I_i$ is an equivalence relation on $W$.

# Models of Extensive Games

Let $\Gamma$ be a *non-degenerate* extensive game with perfect information. Let $\Gamma_i$ be the set of nodes controlled by player $i$.

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

Given a vertex $v$ in $\Gamma$ and strategy profile $s$, $s$ specifies a unique path from $v$ to an end-node.

$\mathcal{M}(\Gamma) = \langle W, (I_i)_{i \in N}, \sigma \rangle$ where $\sigma : W \rightarrow Strat(\Gamma)$ and $I_i$ is an equivalence relation on $W$.

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

# Models of Extensive Games

Let $\Gamma$ be a *non-degenerate* extensive game with perfect information. Let $\Gamma_i$ be the set of nodes controlled by player $i$.

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

Given a vertex $v$ in $\Gamma$ and strategy profile $s$, $s$ specifies a unique path from $v$ to an end-node.

$\mathcal{M}(\Gamma) = \langle W, (I_i)_{i\in N}, \sigma\rangle$ where $\sigma : W \rightarrow \mathit{Strat}(\Gamma)$ and $I_i$ is an equivalence relation on $W$.

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

(A1) For all $w \in W$, $I_i(w) \subseteq [\sigma_i(w)]$

# Rationality

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

# Rationality

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

$i$ **is rational at node** $v$ **in state** $w$ provided for all strategies $s' \neq \sigma_i(w)$, $h_i^v(\sigma(w')) \geqslant h_i^v((s', \sigma_{-i}(w')))$ for some $w' \in I_i(w)$.

# Substantive Rationality

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

# Stalnaker Rationality

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

# Stalnaker Rationality

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

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

# Stalnaker Rationality

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

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

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

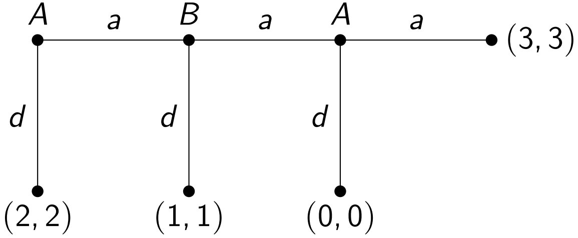
(F2) If $v$ is reached in $w$, then $f(w, v) = w$

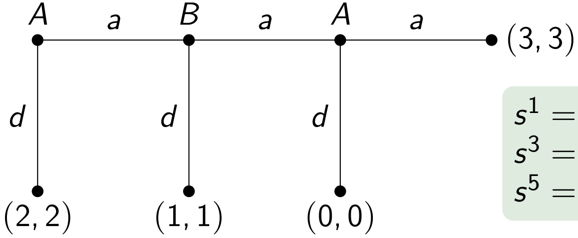
(F3) $\sigma(f(w, v))$ and $\sigma(w)$ agree on the subtree of $\Gamma$ below $v$

$S$-Rationality: $i$ is rational at vertex $v$ in $f(w, v)$ for every vertex in the game $\Gamma$.

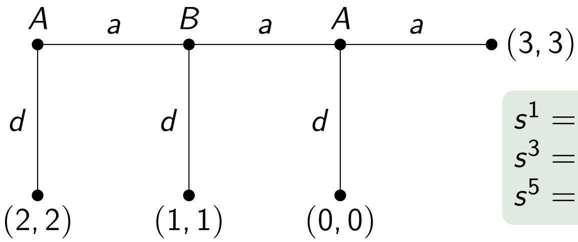
$A$-Rationality: $i$ is rational at vertex $v$ in $w$ for every vertex in the game $\Gamma$.
Player $i$ is $S$-rational at vertex $v$ in state $w$ if $i$ is rational at $v$ in $f(w, v)$:

- $i$'s action is the same at $v$ in $f(w, v)$ and $w$ (by F3)
- But $i$'s beliefs may be different at $w$ and $f(w, v)$
- Check whether $i$'s action is rational given $i$'s beliefs at $f(w, v)$.


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

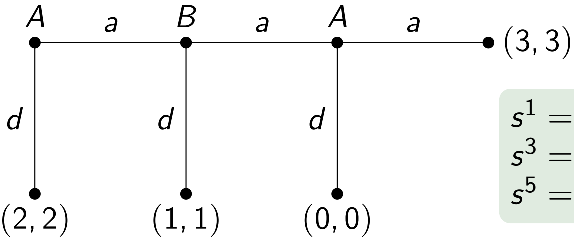



$s^1 = (da, d)$, $s^2 = (aa, d)$,
$s^3 = (ad, d)$, $s^4 = (aa, a)$,
$s^5 = (ad, a)$

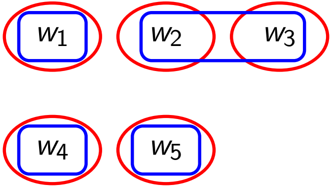


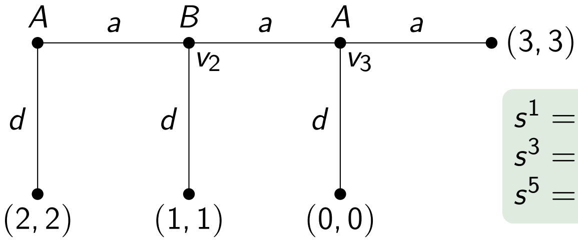

$s^1 = (da, d)$, $s^2 = (aa, d)$,
$s^3 = (ad, d)$, $s^4 = (aa, a)$,
$s^5 = (ad, a)$

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

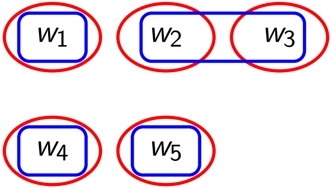



$$s^1 = (da, d),\ s^2 = (aa, d),$$
$$s^3 = (ad, d),\ s^4 = (aa, a),$$
$$s^5 = (ad, a)$$

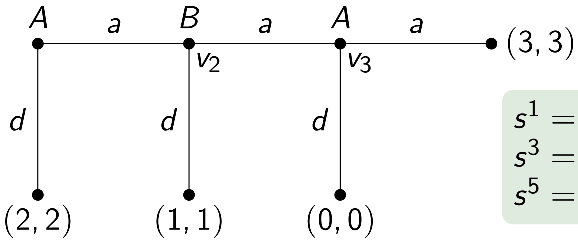




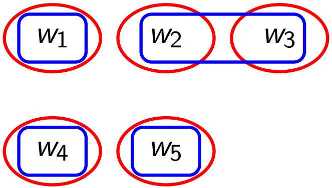

$s^1 = (da, d)$, $s^2 = (aa, d)$,
$s^3 = (ad, d)$, $s^4 = (aa, a)$,
$s^5 = (ad, a)$



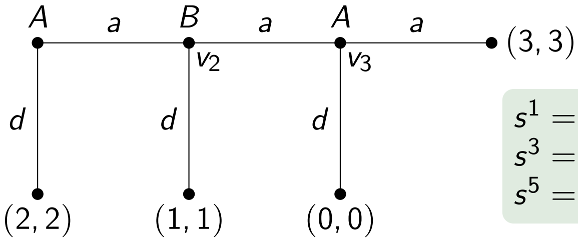

Bob is rational at $v_2$ in $w_2$—Bob considers it possible at $w_2$ that Ann may go down at $v_3$



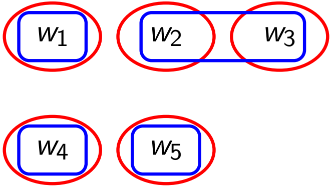

$s^1 = (da, d)$, $s^2 = (aa, d)$,
$s^3 = (ad, d)$, $s^4 = (aa, a)$,
$s^5 = (ad, a)$



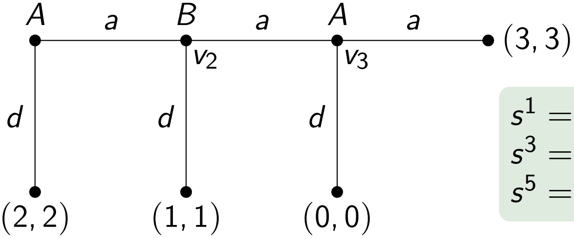

Ann is rational at $v_3$ in $w_4$



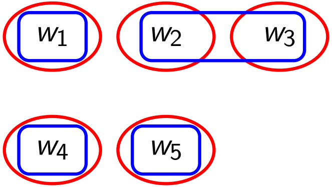

$s^1 = (da, d)$, $s^2 = (aa, d)$,
$s^3 = (ad, d)$, $s^4 = (aa, a)$,
$s^5 = (ad, a)$




Since $f(w_1, v_2) = w_2$ and $f(w_1, v_3) = w_4$ by F1-F3, we have $S$-rationality at $w_1$.




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




So we have common knowledge of $S$-rationality at $w_1$. But $\sigma(w_1)$ is not the backward induction strategy.

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

**Stalnaker's Theorem**: There exists a non-degenerate game $\Gamma$ of perfect information and an extended model of $\Gamma$ in which the selection function satisfies F1-F3 such that $C(S - Rat) \nsubseteq BI$.

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

**Stalnaker's Theorem**: There exists a non-degenerate game $\Gamma$ of perfect information and an extended model of $\Gamma$ in which the selection function satisfies F1-F3 such that $C(S - Rat) \nsubseteq BI$.

Revising beliefs during play:

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

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

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

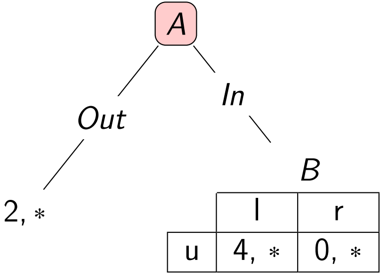
J. Halpern. *Substantive Rationality and Backward Induction*. Games and Economic Behavior, 37, pp. 425-435, 1998.

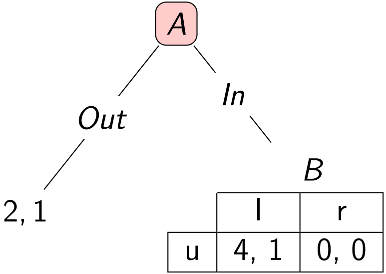
# Other characterizations of BI

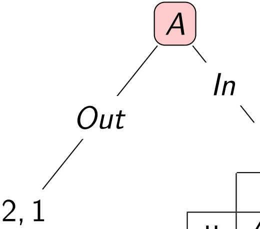
- Future choices are *epistemically independent* of any observed behavior
- Any “off-equilibrium” choice is interpreted simply as a mistake (which will not be repeated)
- At each choice point in a game, the players only reason about future paths

# Rationalizability Assumption

Instead of Bob changing his opinion about Ann's rationality, maintaining his belief about her passive beliefs, he might have maintained his belief in her rationality, changing his beliefs about her beliefs about him.

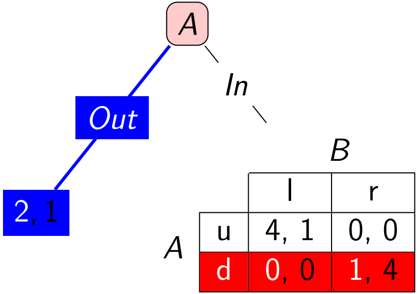

A
Out
In
2, *
B
l
r
u
4, *
0, *


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




*B*

| *A* | | l | r |
|---|---|---|---|
| | 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 |

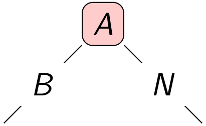
| | | $B$ | |
|---|---|---|---|
| | | l | r |
| | Out | 2, 1 | 2, 1 |
| $A$ | u | 4, 1 | 0, 0 |
| | d | 0, 0 | 1, 4 |

# 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.

# Burning Money



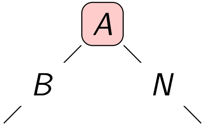

Bob

| Ann | l | r |
|---|---|---|
| u | 2, 1 | -1, 0 |
| d | -1, 0 | 0, 3 |

Bob

| Ann | l | r |
|---|---|---|
| u | 3, 1 | 0, 0 |
| d | 0, 0 | 1, 3 |

# Burning Money



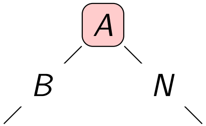

Bob (after $B$)

| Ann | | l | r |
|---|---|---|---|
| | u | 2, 1 | -1, 0 |
| | d | -1, 0 | 0, 3 |

Bob (after $N$)

| Ann | | l | r |
|---|---|---|---|
| | u | 3, 1 | 0, 0 |
| | d | 0, 0 | 1, 3 |

# Burning Money



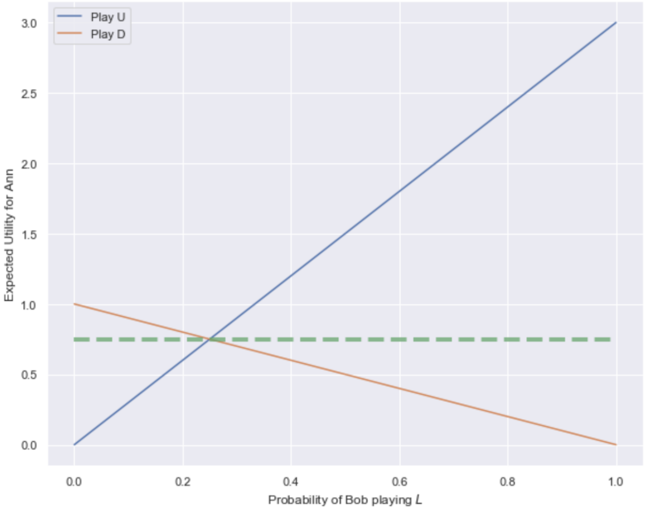

After $B$:

| | | Bob: l | Bob: r |
|---|---|---|---|
| Ann | u | 2, 1 | -1, 0 |
| | d | -1, 0 | 0, 3 |

After $N$:

| | | Bob: l | Bob: r |
|---|---|---|---|
| Ann | u | 3, 1 | 0, 0 |
| | d | 0, 0 | 1, 3 |

1. Ann can achieve at least 0.75 by choosing $N$.
2. 
3. 
4. 
5. 
6. 




Whatever her belief about Bob, Ann can achieve at least 0.75 by choosing $N$.

# Burning Money

$A$

$B$ $N$

After $B$:

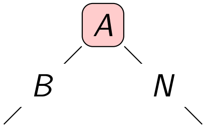
| | | Bob: l | Bob: r |
|---|---|---|---|
| Ann | u | 2, 1 | -1, 0 |
| | d | -1, 0 | 0, 3 |

After $N$:

| | | Bob: l | Bob: r |
|---|---|---|---|
| Ann | u | 3, 1 | 0, 0 |
| | d | 0, 0 | 1, 3 |

1. Ann can achieve at least 0.75 by choosing $N$.
2. If Ann plays $B$, then she will not subsequently choose $d$.
3. Since Bob knows 2., if Ann plays $B$, Bob will play $l$.
4. 
5. 
6. 

# Burning Money




Bob (after $B$)

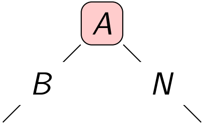
| Ann | | l | r |
|---|---|---|---|
| | u | 2, 1 | -1, 0 |
| | d | -1, 0 | 0, 3 |

Bob (after $N$)

| Ann | | l | r |
|---|---|---|---|
| | u | 3, 1 | 0, 0 |
| | d | 0, 0 | 1, 3 |

1. Ann can achieve at least 0.75 by choosing $N$.
2. If Ann plays $B$, then she will not subsequently choose $d$.
3. Since Bob knows 2., if Ann plays $B$, Bob will play $l$.
4. So, Ann expects a payout of 2 by playing $B$.
5. 
6. 

# Burning Money




Bob (after $B$)

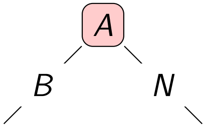
| | | l | r |
|---|---|---|---|
| Ann | u | 2, 1 | -1, 0 |
| | d | -1, 0 | 0, 3 |

Bob (after $N$)

| | | l | r |
|---|---|---|---|
| Ann | u | 3, 1 | 0, 0 |
| | d | 0, 0 | 1, 3 |

1. Ann can achieve at least 0.75 by choosing $N$.
2. If Ann plays $B$, then she will not subsequently choose $d$.
3. Since Bob knows 2., if Ann plays $B$, Bob will play $l$.
4. So, Ann expects a payout of 2 by playing $B$.
5. If Bob observes $N$, then he knows Ann will not play $d$ (since $B$ strictly dominances $Nd$).
6.

# Burning Money




Bob

| | | l | r |
|---|---|---|---|
| Ann | u | 2, 1 | -1, 0 |
| | d | -1, 0 | 0, 3 |

Bob

| | | l | r |
|---|---|---|---|
| Ann | u | 3, 1 | 0, 0 |
| | d | 0, 0 | 1, 3 |

1. Ann can achieve at least 0.75 by choosing *N*.
2. If Ann plays *B*, then she will not subsequently choose *d*.
3. Since Bob knows 2., if Ann plays *B*, Bob will play *l*.
4. So, Ann expects a payout of 2 by playing *B*.
5. If Bob observes *N*, then he knows Ann will not play *d* (since *B* strictly dominances *Nd*).
6. Knowing all of the above, Ann will play *Nu* and Bob will play *ll*.

Ben-Porath and Dekel (1992) generalized this result as follows: in games in which a player has a strict preference for a equilibrium point, and if this player can self-sacrifice (burning utility), then, based on the forward induction rationality and iterative elimination of weakly dominated strategies, such player will achieve her most preferred outcome.

E. Ben-Porath and E. Dekel. *Signaling future actions and the potential for sacrifice*. Journal of Economic Theory, 57, 36-51, 1992.

"If disposing of the dollar were a relevant consideration in the players' perception of the situation, then the result would (probably) make sense. however, I cannot believe that any reasonable person would consider a pre-game disposal of a dollar to be relevant in the analysis of the battle of the sexes. It is my opinion that a formal description of the situation should exclude the choice of disposal even in cases where a description of the game is given by a referee who specifies the possibility of disposing of the dollar (recall that there is rarely a referee)." (Rubinstein, pg. 920)

A. Rubinstein. *Comments on the Interpretation of Game Theory*. Econometrica, 59(4), pp. 909 - 924, 1991.

1. Ann's act of burning utility can be interpreted by Bob as irrational or as an error and, for this reason, should not be considered in the prediction of Ann's future behavior.
2. Cheap talk: "If instead of being able to throw a dollar out of the window, Ann is allowed to throw a bill worth nothing out of the window (or nod her head), or even stat that "I am going to play $u$", then the process of successive elimination of weakly dominated strategies is not powerful an all equilibria of the battle of the sexes would survive." (Rubinstein, pg. 921)

# Rationalization Principle

A player should believe that all players are perfectly rational, and this belief should be robust relative to any compatible information about the behavior of any player....

# Rationalization Principle

A player should believe that all players are perfectly rational, and this belief should be robust relative to any compatible information about the behavior of any player....If you are surprised by the actions of some player, you should change your beliefs about that player's passive beliefs, rather than about her rationality. If possible, find an alternative hypothesis about her beliefs about other players that will make what she does perfectly rational.

“...in general, a payer’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 payer'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 payer'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)

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.

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.

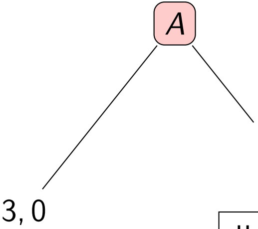
P. Battigalli and M. Siniscalchi. *Strong Belief and Forward Induction Reasoning*. Journal of Economic Theory, 106, pp. 356 - 391, 2002.

# 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)

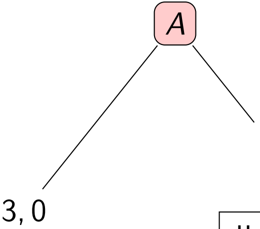
# Backward *versus* Forward Induction




| | | 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.

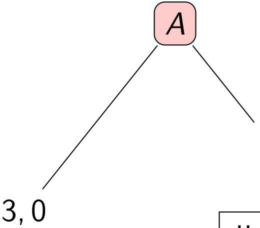
# Backward *versus* Forward Induction




| | | Bob | | |
|---|---|---|---|---|
| | | 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.

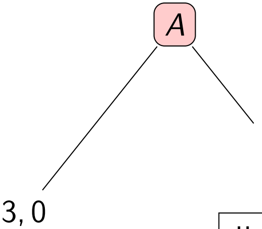
# Backward *versus* Forward Induction




| | | 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.

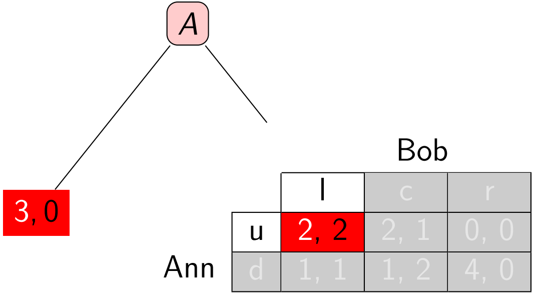
# Backward *versus* Forward Induction




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

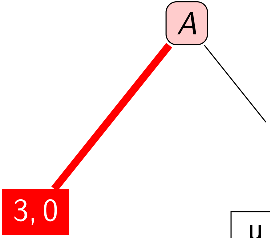
A. Perea. *Backward Induction* versus *Forward Induction Reasoning*. Games, 1, pgs. 168 - 188, 2010.

# Backward *versus* Forward Induction




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

# Backward *versus* Forward Induction

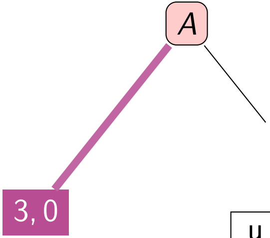



Bob

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

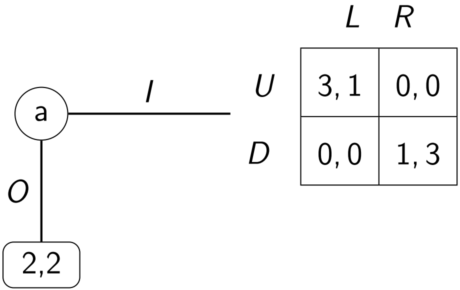
A. Perea. *Backward Induction* versus *Forward Induction Reasoning*. Games, 1, pgs. 168 - 188, 2010.

# Forward induction reasoning

| | $L$ | $R$ |
|---|---|---|
| $U$ | $3, 1$ | $0, 0$ |
| $D$ | $0, 0$ | $1, 3$ |

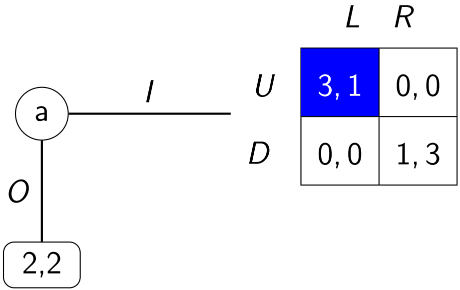
A. Knoks and EP. *Deliberating between Backward and Forward Induction Reasoning: First Steps*. Proceedings of Theoretical Aspects of Rationality and Knowledge (TARK 2015).

# Forward induction reasoning



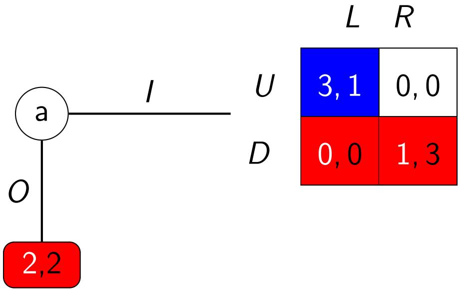

A. Knoks and EP. *Deliberating between Backward and Forward Induction Reasoning: First Steps*. Proceedings of Theoretical Aspects of Rationality and Knowledge (TARK 2015).

# Forward induction reasoning



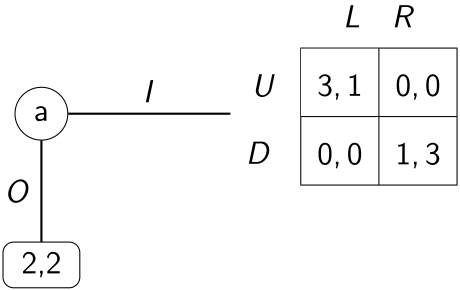

A. Knoks and EP. *Deliberating between Backward and Forward Induction Reasoning: First Steps*. Proceedings of Theoretical Aspects of Rationality and Knowledge (TARK 2015).

# Forward induction reasoning



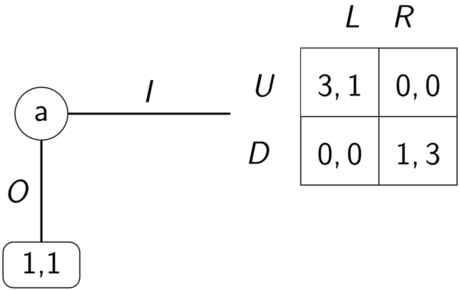

A. Knoks and EP. *Deliberating between Backward and Forward Induction Reasoning: First Steps*. Proceedings of Theoretical Aspects of Rationality and Knowledge (TARK 2015).

# Forward induction reasoning



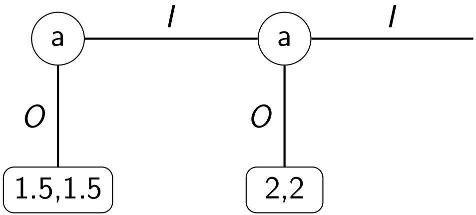

A. Knoks and EP. *Deliberating between Backward and Forward Induction Reasoning: First Steps*. Proceedings of Theoretical Aspects of Rationality and Knowledge (TARK 2015).

# Forward induction reasoning




A. Knoks and EP. *Deliberating between Backward and Forward Induction Reasoning: First Steps*. Proceedings of Theoretical Aspects of Rationality and Knowledge (TARK 2015).

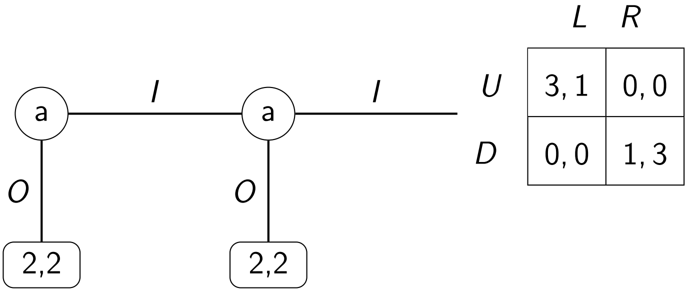
# Forward induction reasoning




| | $L$ | $R$ |
|---|---|---|
| $U$ | 3, 1 | 0, 0 |
| $D$ | 0, 0 | 1, 3 |

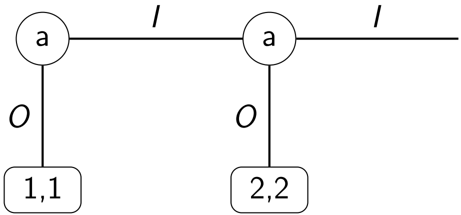
A. Knoks and EP. *Deliberating between Backward and Forward Induction Reasoning: First Steps*. Proceedings of Theoretical Aspects of Rationality and Knowledge (TARK 2015).

# Forward induction reasoning




A. Knoks and EP. *Deliberating between Backward and Forward Induction Reasoning: First Steps*. Proceedings of Theoretical Aspects of Rationality and Knowledge (TARK 2015).

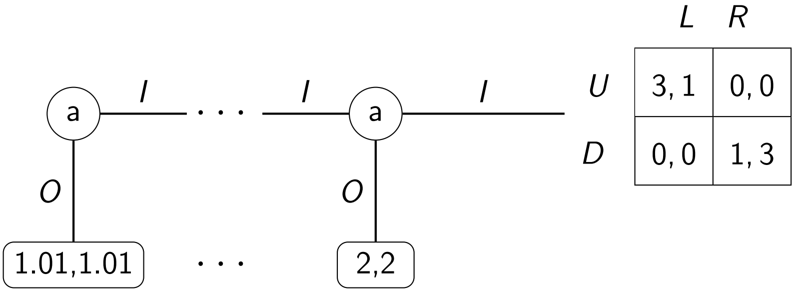
# Forward induction reasoning




| | $L$ | $R$ |
|---|---|---|
| $U$ | 3, 1 | 0, 0 |
| $D$ | 0, 0 | 1, 3 |

A. Knoks and EP. *Deliberating between Backward and Forward Induction Reasoning: First Steps*. Proceedings of Theoretical Aspects of Rationality and Knowledge (TARK 2015).

# Forward induction reasoning




A. Knoks and EP. *Deliberating between Backward and Forward Induction Reasoning: First Steps*. Proceedings of Theoretical Aspects of Rationality and Knowledge (TARK 2015).