Logics of Action, Ability, Knowledge and Obligation

John F. Horty Eric Pacuit

Department of Philosophy
University of Maryland

pacuit.org/esslli2019/epstit

August 9, 2019

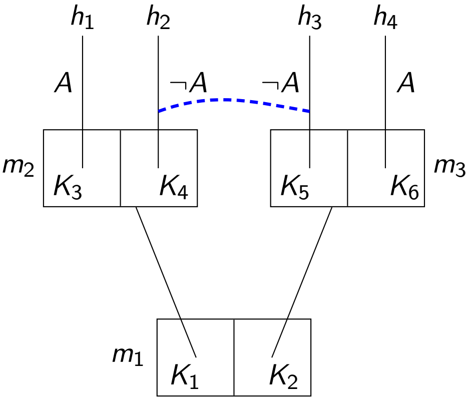
A. Herzig. *Logics of knowledge and action: critical analysis and challenges*. Autonomous Agent and Multi-Agent Systems, 2014.

V. Goranko and EP. *Temporal aspects of the dynamics of knowledge*. in Johan van Benthem on Logic and Information Dynamics, Outstanding Contributions to Logic, (eds. Alexandru Baltag and Sonja Smets), pp. 235 - 266, 2014.

J. Broeresen, A. Herzig and N. Troquard. *What groups do, can do and know they can do: An analysis in normal modal logics*. Journal of Applied and Non-Classical Logics, 19:3, pgs. 261 - 289, 2009.

W. van der Hoek and M. Wooldridge. *Cooperation, knowledge and time: Alternating-time temporal epistemic logic and its applications*. Studia Logica, 75, pgs. 125 - 157, 2003.

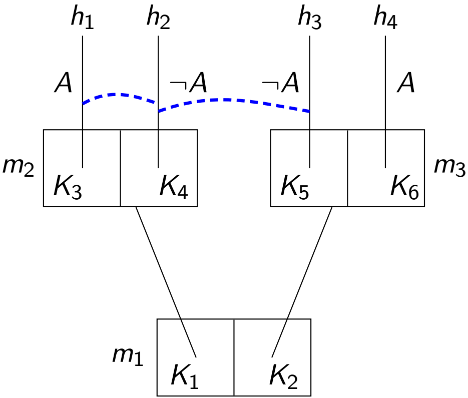
$$\langle Tree, <, Agent, Choice, \{\sim_\alpha\}_{\alpha \in Agent}, V \rangle$$



$\sim_\alpha$ is an equivalence relation on indices

$m/h \sim_\alpha m'/h'$: nothing α knows distinguishes m/h from m'/h', or m/h and m'/h' are indistinguishable

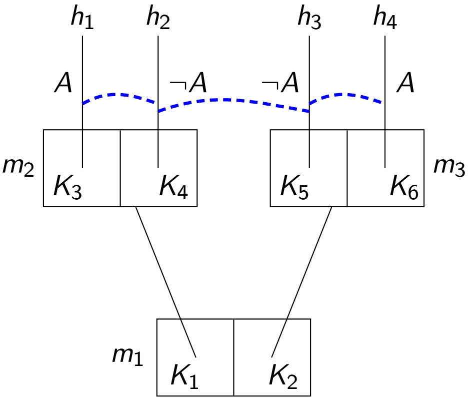
$$\langle \mathit{Tree}, <, \mathit{Agent}, \mathit{Choice}, \{\sim_\alpha\}_{\alpha \in \mathit{Agent}}, V \rangle$$



$\sim_\alpha$ is an equivalence relation on indices

$m/h \sim_\alpha m'/h'$: nothing α knows distinguishes m/h from m'/h', or m/h and m'/h' are indistinguishable

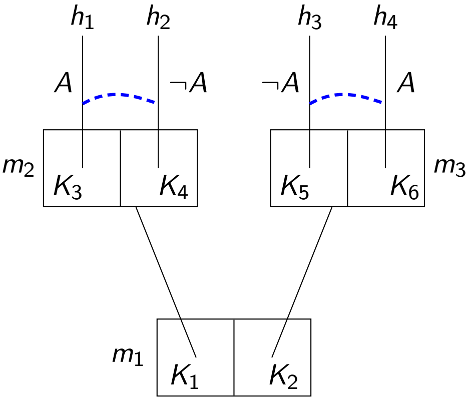
$\langle Tree, <, Agent, Choice, \{\sim_\alpha\}_{\alpha \in Agent}, V\rangle$



$\sim_\alpha$ is an equivalence relation on indices

$m/h \sim_\alpha m'/h'$: nothing α knows distinguishes m/h from m'/h', or m/h and m'/h' are indistinguishable

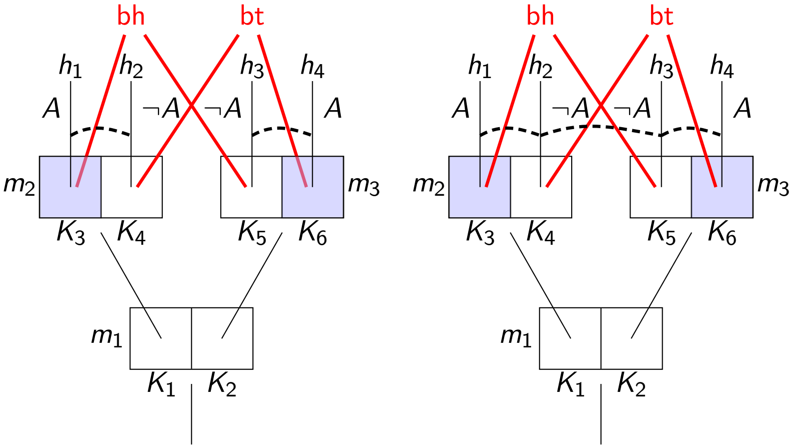
$$\langle Tree, <, Agent, Choice, \{\sim_\alpha\}_{\alpha \in Agent}, V\rangle$$



$\sim_\alpha$ is an equivalence relation on indices

$m/h \sim_\alpha m'/h'$: nothing α knows distinguishes m/h from m'/h', or m/h and m'/h' are indistinguishable

Ability



bh
bt
h_1 h_2 h_3 h_4
A $\neg A$ $\neg A$ A
m_2 m_3
K_3 K_4 K_5 K_6
m_1
K_1 K_2
bh
bt
h_1 h_2 h_3 h_4
A $\neg A$ $\neg A$ A
m_2 m_3
K_3 K_4 K_5 K_6
m_1
K_1 K_2

Deliberative perspective

$(C5)$ If $m/h \sim_\alpha m'/h'$, then $m/h'' \sim_\alpha m'/h'''$ for all $h'' \in H^m$ and $h''' \in H^{m'}$

Indistinguishability between moments: $m \sim_\alpha m'$ iff $m/h \sim_\alpha m'/h'$ for all $h \in H^m$ and $h' \in H^{m'}$.

Game Theory

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

- the actions the players *can* take
- the players' interests (i.e., preferences),
- the "structure" of the decision problem

Game Theory

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

- the actions the players *can* take
- the players' interests (i.e., preferences),
- the "structure" of the decision problem

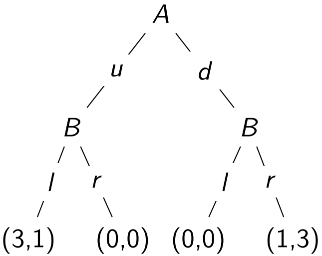
It does not specify the actions that the players do take.

Games

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

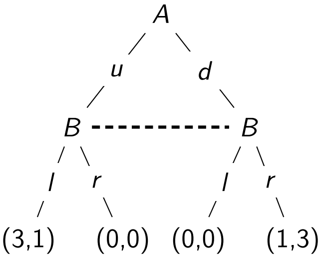
Games

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



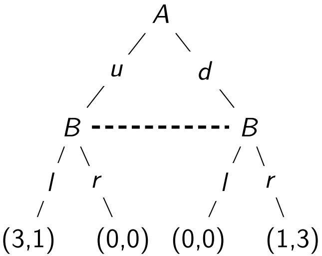
Games

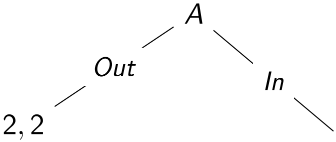
		B: l	B: r
A	u	3, 1	0, 0
	d	0, 0	1, 3



Games

		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

Knowledge and beliefs in game situations

J. Harsanyi. *Games with incomplete information played by "Bayesian" players I-III. Management Science Theory* **14***: 159-182, 1967-68.*

R. Aumann. *Interactive Epistemology I & II*. International Journal of Game Theory (1999).

P. Battigalli and G. Bonanno. *Recent results on belief, knowledge and the epistemic foundations of game theory*. Research in Economics (1999).

R. Myerson. *Harsanyi's Games with Incomplete Information*. Special 50th anniversary issue of *Management Science*, 2004.

John C. Harsanyi, nobel prize winner in economics, developed a theory of games with **incomplete information**.

J. Harsanyi. *Games with incomplete information played by "Bayesian" players I-III. Management Science Theory* **14***: 159-182, 1967-68.*

John C. Harsanyi, nobel prize winner in economics, developed a theory of games with **incomplete information**.

1. incomplete information: uncertainty about the *structure* of the game (outcomes, payoffs, strategy space)
2. imperfect information: uncertainty *within the game* about the previous moves of the players

J. Harsanyi. *Games with incomplete information played by "Bayesian" players I-III. Management Science Theory* **14***: 159-182, 1967-68.*

Information in games situations

- Various states of information disclosure.

Information in games situations

- Various states of information disclosure.
  - *ex ante*, *ex interim*, *ex post*

Information in games situations

- Various states of information disclosure.
  - *ex ante*, *ex interim*, *ex post*

- Various "types" of information:

Information in games situations

- Various states of information disclosure.
  - *ex ante*, *ex interim*, *ex post*

- Various "types" of information:
  - imperfect information about the play of the game
  - incomplete information about the structure of the game
  - strategic information (what will the other players do?)
  - higher-order information (what are the other players thinking?)

Information in games situations

- Various states of information disclosure.
  - *ex ante*, *ex interim*, *ex post*
- Various "types" of information:
  - imperfect information about the play of the game
  - incomplete information about the structure of the game
  - strategic information (what will the other players do?)
  - higher-order information (what are the other players thinking?)
- Varieties of informational attitudes

Information in games situations

- Various states of information disclosure.
  - *ex ante*, *ex interim*, *ex post*

- Various "types" of information:
  - imperfect information about the play of the game
  - incomplete information about the structure of the game
  - strategic information (what will the other players do?)
  - higher-order information (what are the other players thinking?)

- Varieties of informational attitudes
  - hard ("knowledge")
  - soft ("beliefs")

Ex ante vs. *ex interim* knowledge

- $\mathcal{M}, m/h \models \mathsf{K}_\alpha A$ if and only if, for all m'/h', if $m/h \sim_\alpha m'/h'$, then $\mathcal{M}, m'/h' \models A$

- $\mathcal{M}, m/h \models \mathsf{K}^{\mathsf{act}}_\alpha A$ if and only if, for all m'/h', if $m/h \sim_\alpha m'/h'$ and $h' \in [\mathit{Type}^m_\alpha(h)]^{m'}_\alpha$, $\mathcal{M}, m'/h' \models A$

Discussion

- Language/validities

$\Box A \supset [\alpha\ \mathit{stit}\colon A]$
$\mathsf{K}_\alpha \Box A \supset [\alpha\ \mathit{kstit}\colon A]$
$[\alpha\ \mathit{kstit}\colon A] \equiv \mathsf{K}^{\mathsf{act}}_\alpha [\alpha\ \mathit{stit}\colon A]$
...

- What do the agents know vs. What do the agents know *given what they are doing*.

Broersen and Duijf

$m/h \sim'_\alpha m'/h'$ if and only if, $m/h \sim_\alpha m'/h'$ and $h' \in [Type^m_\alpha(h)]^{m'}_\alpha$

Broersen and Duijf

Theorem 1. *Let $\mathcal{M}$ be a static labelled stit model, and let $\mathcal{M}'$ be the associated transform epistemic stit model. Let $i \in Ags$ be an agent, $\varphi \in \mathfrak{L}_{stit}$ be a standard stit formula, and m/h be an index. Then the following holds*

(1) $\mathcal{M}, m/h \models [i \text{ kstit}]\varphi$ *if and only if* $\mathcal{M}', m/h \models \mathsf{K}_i[i \text{ stit}]\varphi$;

(2) $\mathcal{M}, m/h \models \mathsf{K}_i\varphi$ *if and only if* $\mathcal{M}', m/h \models \mathsf{K}_i\Box\varphi$;

and

(3) $\mathcal{M}', m/h \models \mathsf{K}_i\varphi \to \mathsf{K}_i[i \text{ stit}]\varphi$;

(4) $\mathcal{M}', m/h \models \Diamond\mathsf{K}_i\varphi \to \mathsf{K}_i\Diamond\varphi$

Concluding Remarks

- Group epistemic agency:

Concluding Remarks

- Group epistemic agency:
  - Collective attitudes: *aggregate attitudes* (e.g., distributed knowledge, collective wisdom, ...) vs. *common attitudes* (e.g., common knowledge)

    (cf. C. List, "Three kinds of collective attitudes", 2015)

Concluding Remarks

- Group epistemic agency:
  - Collective attitudes: *aggregate attitudes* (e.g., distributed knowledge, collective wisdom, ...) vs. *common attitudes* (e.g., common knowledge)

    (cf. C. List, "Three kinds of collective attitudes", 2015)
  - Group action types: products of individual action types vs. labelings of products of individual action tokens

Concluding Remarks

- Group epistemic agency:
  - Collective attitudes: *aggregate attitudes* (e.g., distributed knowledge, collective wisdom, ...) vs. *common attitudes* (e.g., common knowledge)

    (cf. C. List, "Three kinds of collective attitudes", 2015)
  - Group action types: products of individual action types vs. labelings of products of individual action tokens
  - Group oughts (Kooi and Tamminga: relativized to the interests of other groups of agents)

Conclude Remarks

- Making assumptions about what the other agents are going to do (what should you do when you know that the other agents will do what they ought to do?)

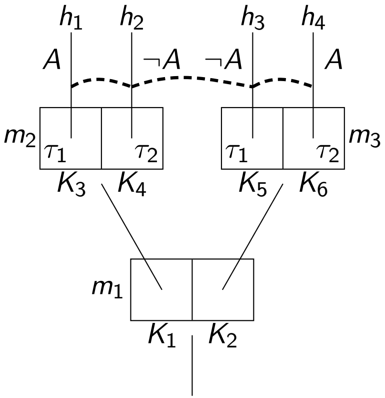
Conclude Remarks

- Making assumptions about what the other agents are going to do (what should you do when you know that the other agents will do what they ought to do?)

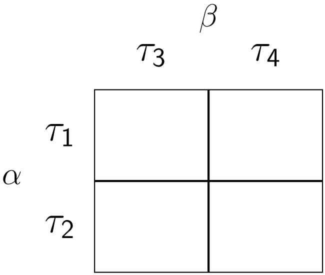
- Moving beyond common payoffs

Conclude Remarks

- Making assumptions about what the other agents are going to do (what should you do when you know that the other agents will do what they ought to do?)
- Moving beyond common payoffs
- When do two labeled stit models represent the *same* situation? (cf. when are two *games* the same?)



h_1
h_2
h_3
h_4
A
$\neg A$
$\neg A$
A
m_2
τ_1
τ_2
τ_1
τ_2
m_3
K_3
K_4
K_5
K_6
m_1
K_1
K_2



β
τ_3
τ_4
α
τ_1
τ_2

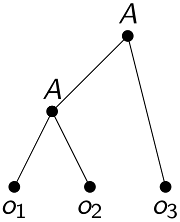
Thompson Transformations

Game-theoretic analysis should not depend on "irrelevant" features of the (mathematical) description of the game.

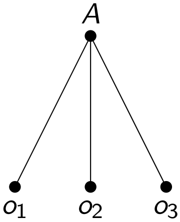
F. B. Thompson. *Equivalence of Games in Extensive Form*. Classics in Game Theory, pgs 36 - 45, 1952.

(Osborne and Rubinstein, pgs. 203 - 212)

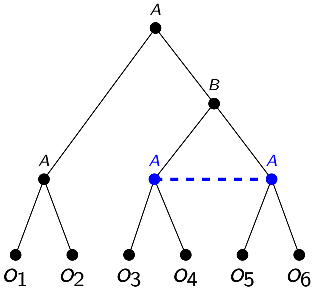
The same decision problem

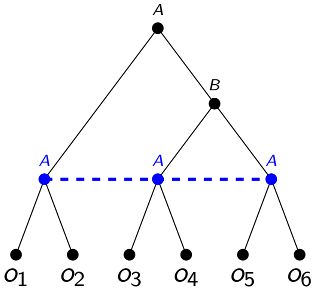


D_1

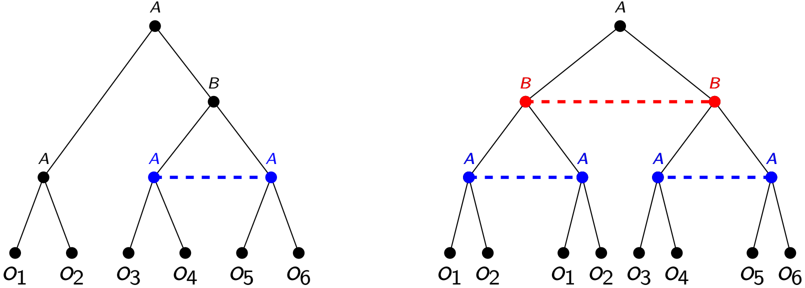


D_2

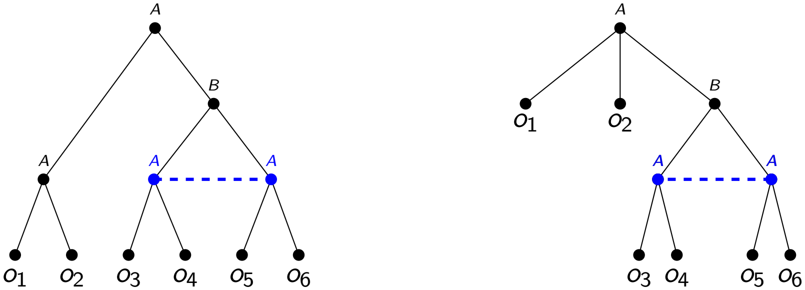




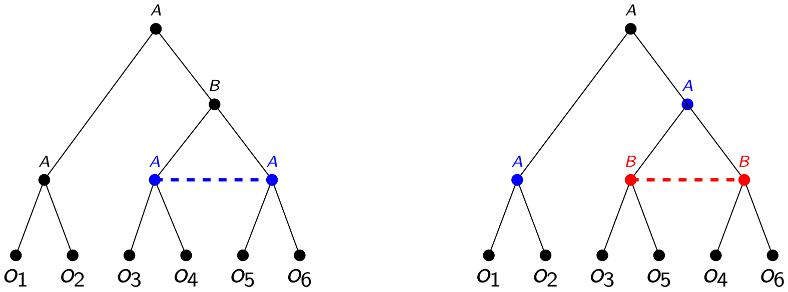
Inflation/deflation



Addition of Superfluous Move



Coalescing of moves



Interchange of moves

Theorem (Thompson) Each of the previous transformations preserves the reduced strategic form of the game. In finite extensive games (without uncertainty between subhistories), if any two games have the same reduced normal form then one can be obtained from the other by a sequence of the four transformations.

Extensive vs. Normal Forms

G. Bonanno. *Set-Theoretic Equivalence of Extensive-Form Games*. IJGT (1992).

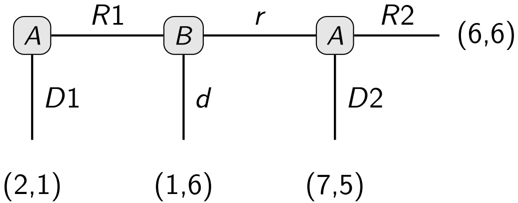
S. Elmes and P. Reny. *On The Strategic Equivalence of Extensive Form Games*. Journal of Economic Theory (1994).

E. Kholberg and F. Mertens. *On Strategic Stability of Equilibria*. Econometrica (1986).

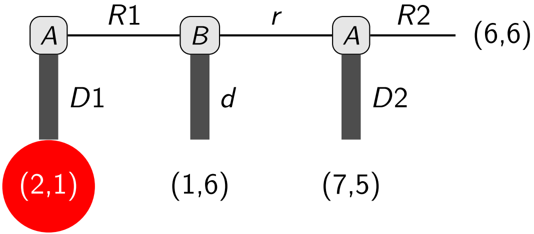
T. Seidenfeld. *When Normal and Extensive Form Decisions Differ*. Logic, Methodology and Philosophy of Science, 1994.

Strategic stit/obligations and incorporating game-theoretic reasoning (and beliefs)

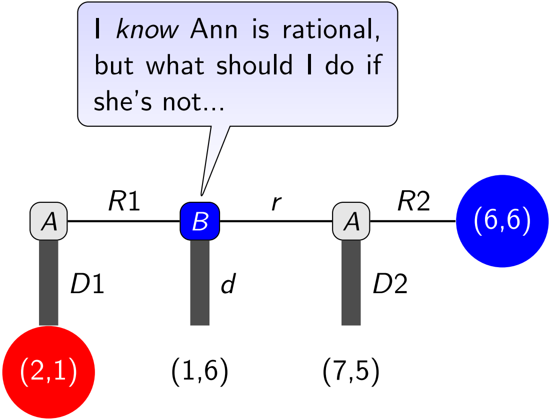
BI Puzzle?

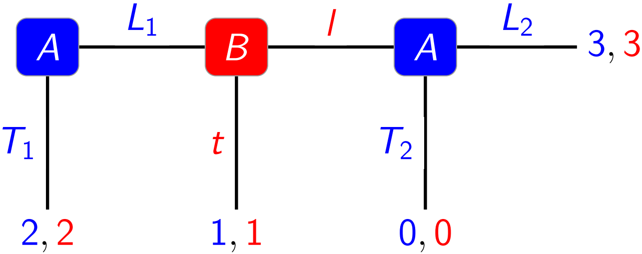


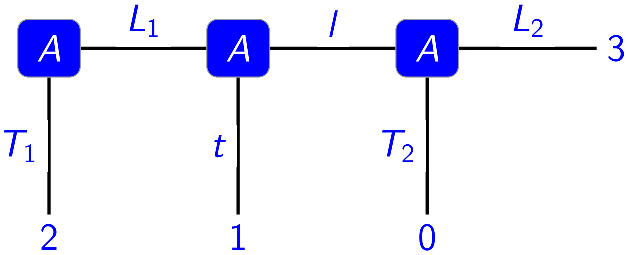
BI Puzzle?

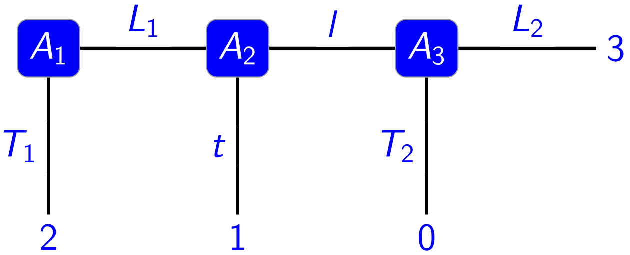


BI Puzzle?

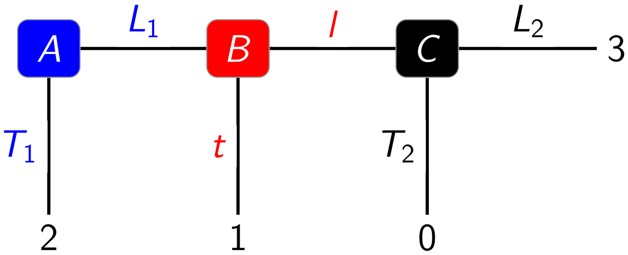

I *know* Ann is rational, but what should I do if she's not...
A
R1
B
r
A
R2
(6,6)
D1
d
D2
(2,1)
(1,6)
(7,5)

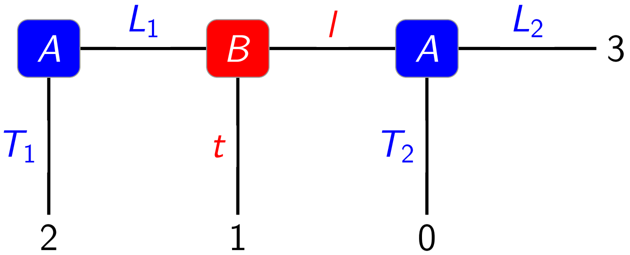

A
L_1
B
l
A
L_2
3, 3
T_1
t
T_2
2, 2
1, 1
0, 0


A
A
A
L1
l
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
L_1
B
l
C
L_2
3
T_1
t
T_2
2
1
0



A
L1
B
l
A
L2
3
T1
t
T2
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.

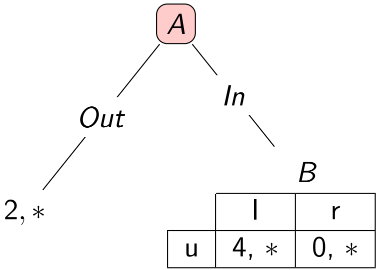
Informal characterizations of BI

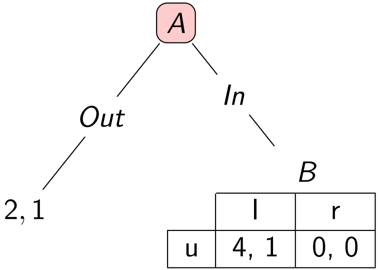
- Future choices are *epistemically independent* of any observed behavior
- Any "off-equilibrium" choice is interpreted simply as a mistake (which will not be repeated)
- At each choice point in a game, the players only reason about future paths

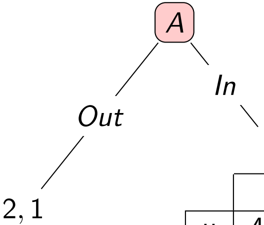
Rationalizing Observed Actions

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.

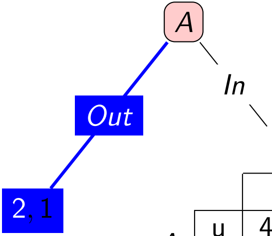

A
Out
In
2, ∗
B
l
r
u
4, ∗
0, ∗


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



B

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



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

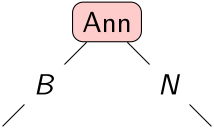
		B	
		l	r
	Out	2, 1	2, 1
A	u	4, 1	0, 0
	d	0, 0	1, 4

		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
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



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

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

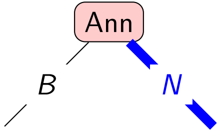
		Bob			
		ll	lr	rl	rr
Ann	Bu	2, 1	2, 1	-2, 0	-2, 0
	Bd	-2, 0	-2, 0	-1, 4	-1, 4
	Nu	4, 1	0, 0	4, 1	0, 0
	Nd	0, 0	1, 4	0, 0	1, 4

		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			
		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

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

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.

Thank you!