Logic and Artificial Intelligence

Lecture 26

Eric Pacuit

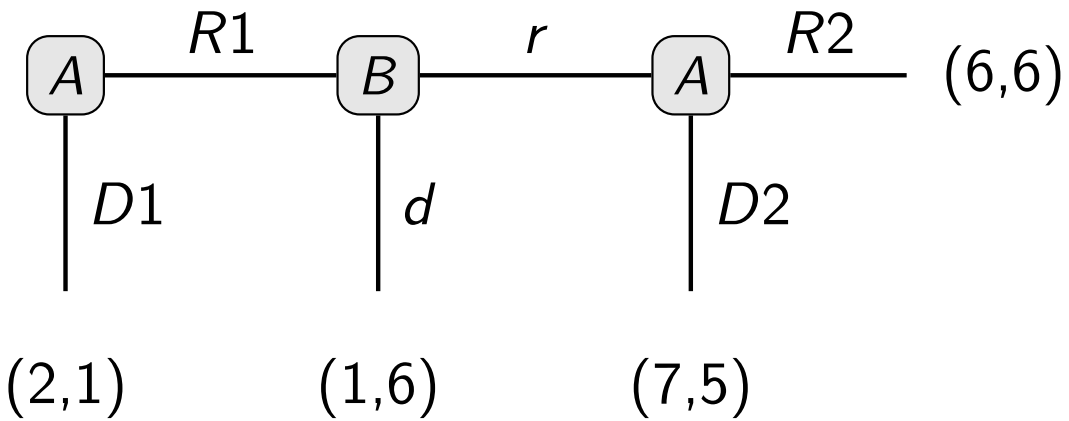
Currently Visiting the Center for Formal Epistemology, CMU

Center for Logic and Philosophy of Science
Tilburg University

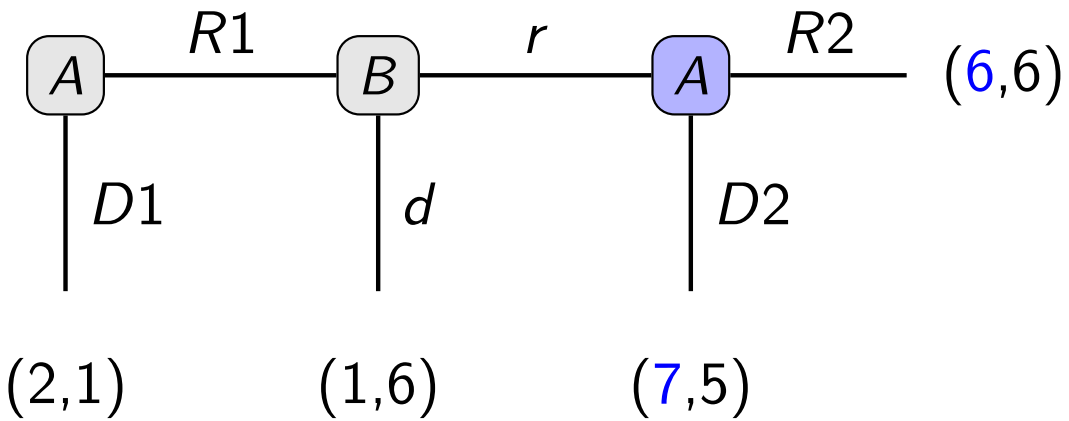
ai.stanford.edu/∼epacuit
e.j.pacuit@uvt.nl

December 7, 2011

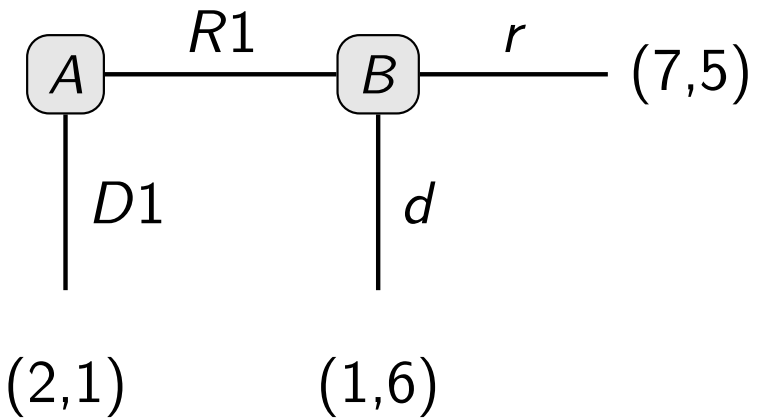
BI Puzzle



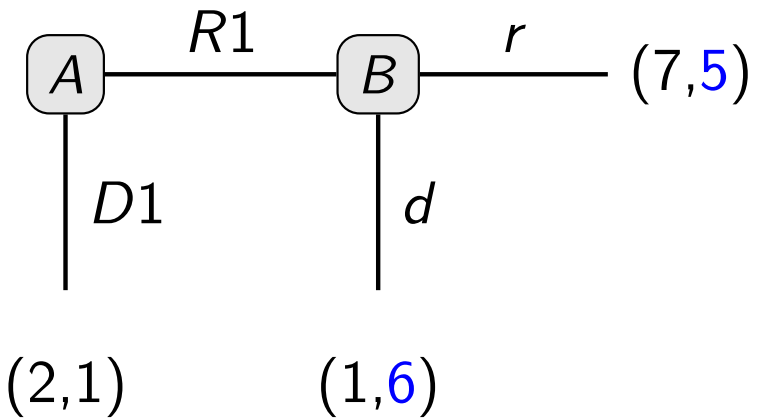
BI Puzzle



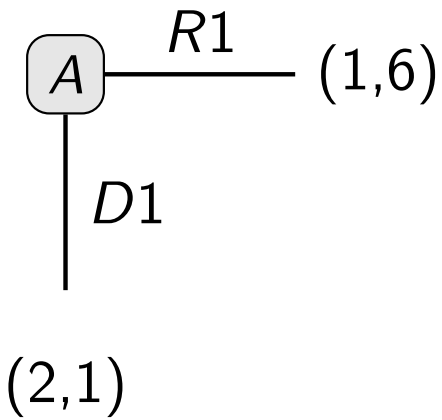
BI Puzzle



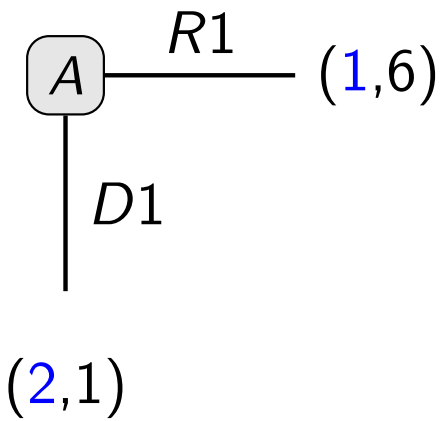
BI Puzzle



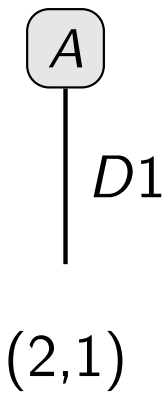
BI Puzzle



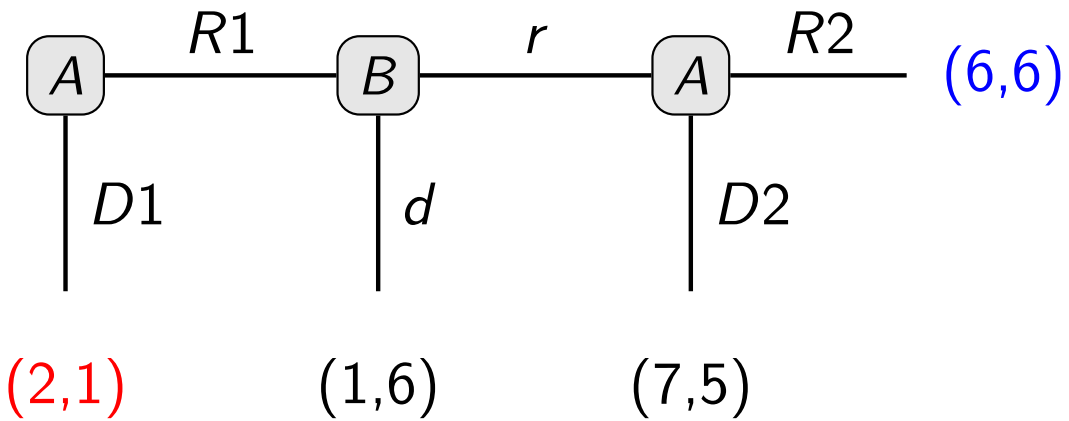
BI Puzzle



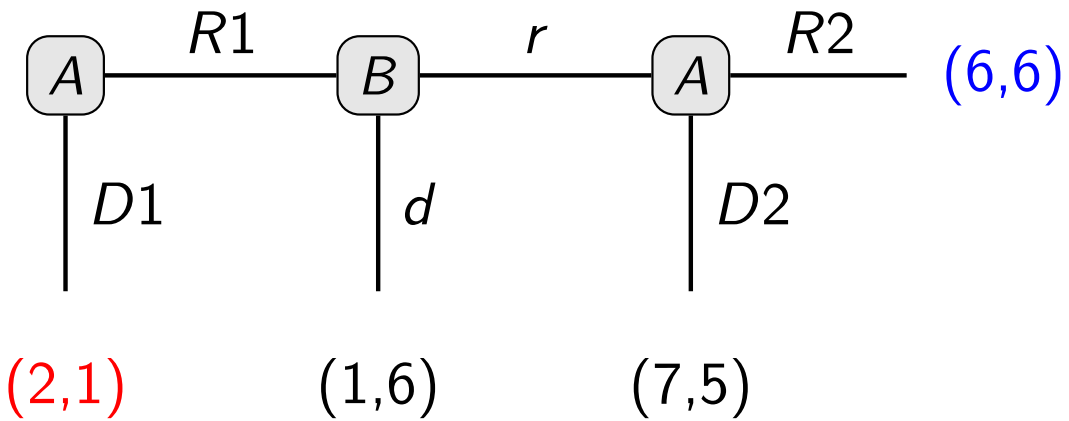
BI Puzzle



But what if...



But what if...



"On the one hand, Under common knowledge of rationality, A *must* go out on the first move. On the other hand, the backward induction argument for this is based on what the players *would* do if A stayed in. But, if she did stay in, then common knowledge of rationality is violated, so the argument that she will go out no longer has a basis."

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.

Models of Extensive Games

$\mathcal{M}(\Gamma) = \langle W, \sim_i, f, \mathbf{s} \rangle$ where

(A1) If $w \sim_i w'$ then $\mathbf{s}_i(w) = \mathbf{s}_i(w')$.

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

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

(F3) $\mathbf{s}(f(w, v))$ and $\mathbf{s}(w)$ agree on the subtree of Γ below v

(F4) For all players i and vertices v, if $w' \in [f(w, v)]_i$ then there exists a state $w'' \in [w]_i$ such that $\mathbf{s}(w')$ and $\mathbf{s}(w'')$ agree on the subtree of Γ below v.

Rationality

i **is rational at** v **in** w: for all strategies $s_i \neq \mathbf{s}_i(w)$, $h_i^v(\mathbf{s}(w')) \geq h_i^v((\mathbf{s}_{-i}(w'), s_i))$ for some $w' \in [w]_i$:

$$\bigwedge_{v \in \Gamma_i} \bigwedge_{t_i \in Strat_i(\Gamma)} \neg K_i[h_i^v(s; t_i) > h_i^v(s)]$$

A-Rat: i is rational at vertex v in w for every vertex $v \in \Gamma_i$

S-Rat: i is rational at vertex v is $f(w, v)$ fo every vertex $v \in \Gamma_i$

(A1) If $w \sim_i w'$ then $\mathbf{s}_i(w) = \mathbf{s}_i(w')$.

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

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

(F3) $\mathbf{s}(f(w, v))$ and $\mathbf{s}(w)$ agree on the subtree of Γ below v

(F4) For all players i and vertices v, if $w' \in [f(w, v)]_i$ then there exists a state $w'' \in [w]_i$ such that $\mathbf{s}(w')$ and $\mathbf{s}(w'')$ agree on the subtree of Γ below v.

Theorem (Halpern). If Γ is a non-degenerate game of perfect information, then for every extended model of Γ in which the selection function satisfies F1-F4, we have $C(\text{S-Rat}) \subseteq BI$.

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

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"

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"

"the rationality of choices in a game depends not only on what players believe, but also on their policies for revising their beliefs" (p. 31)

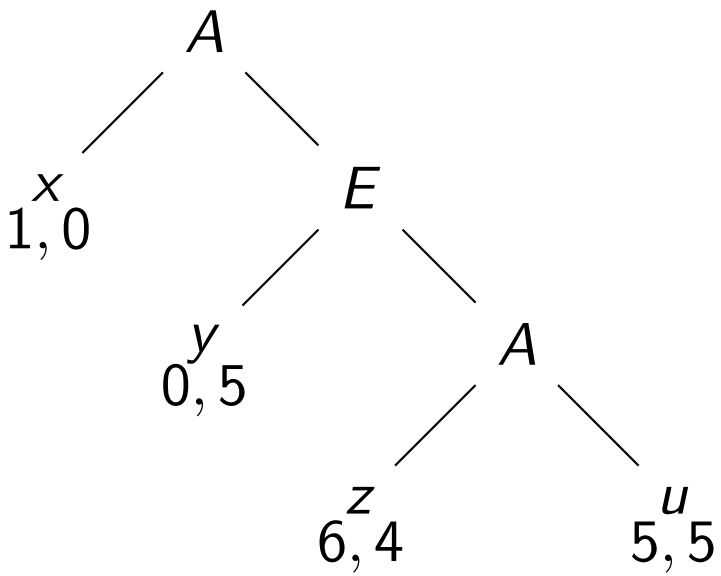
R. Stalnaker. *Belief revision in games: Forward and backward induction*. Mathematical Social Sciences, 36, pgs. 31 - 56, 1998.

"Off-line learning of rationality"

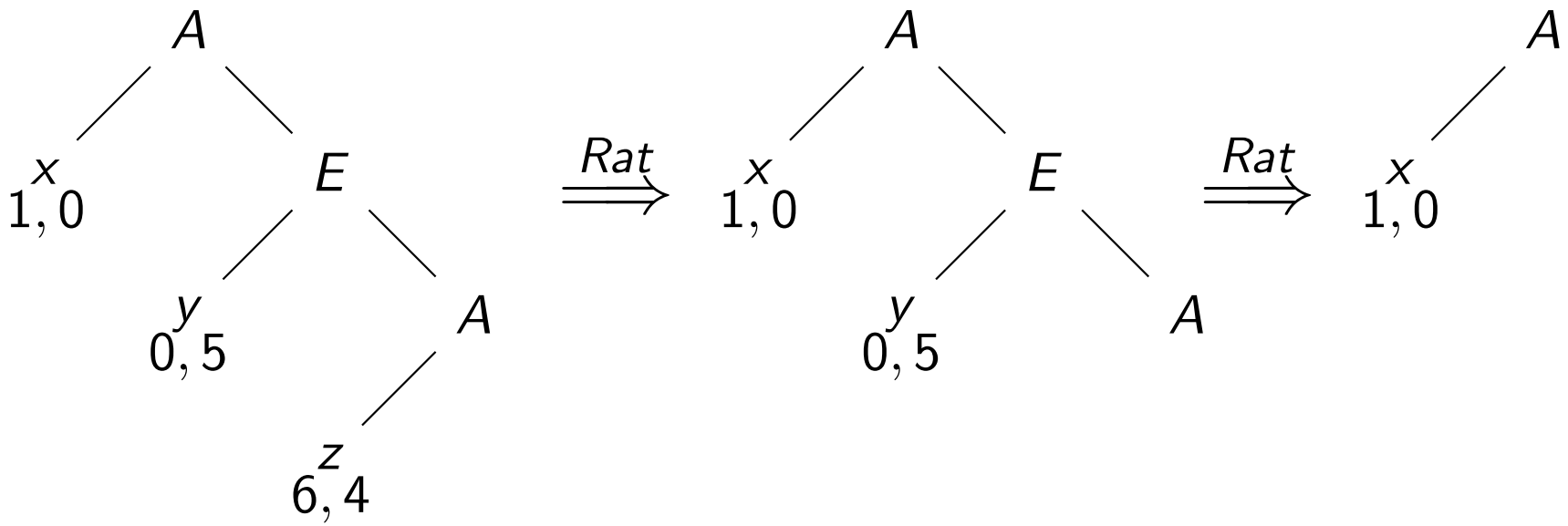
Where do the models satisfying common knowledge/belief of rationality come from?

J. van Benthem. *Rational dynamics and epistemic logic in games*. International Journal of Game Theory Review, 9(1), pgs. 13 - 45, 2007.

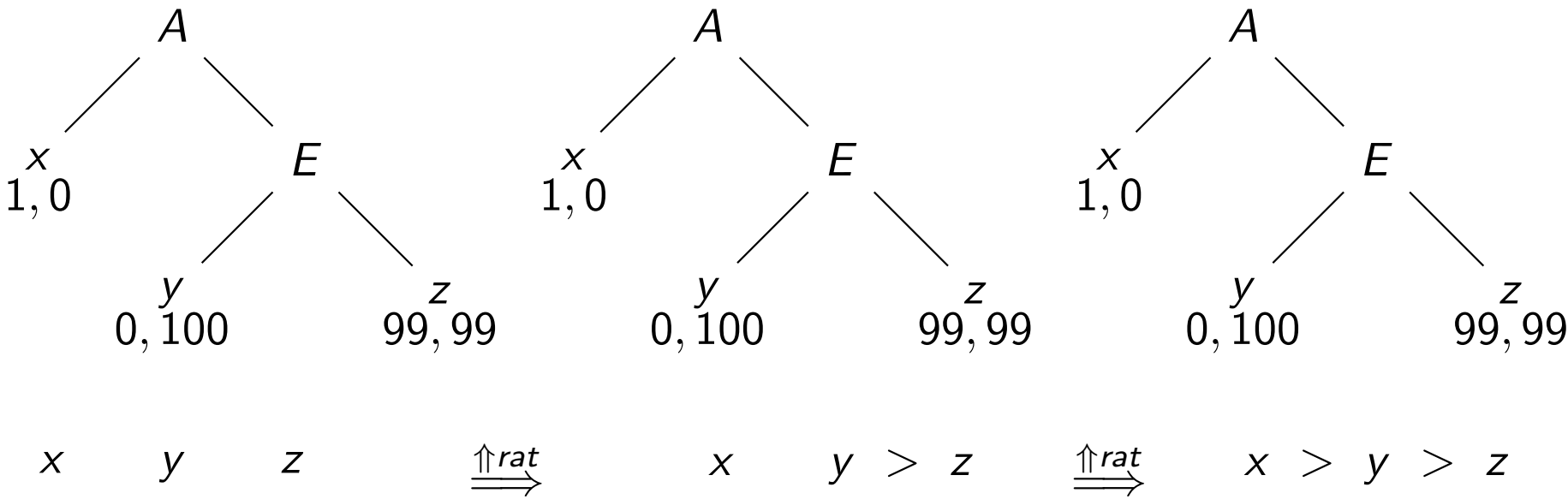
“Off-line learning of rationality”



“Off-line learning of rationality”



"Off-line learning of rationality"



The Dynamics of Rational Play

A. Baltag, S. Smets and J. Zvesper. *Keep 'hoping' for rationality: a solution to the backward induction paradox*. Synthese, 169, pgs. 301 - 333, 2009.

Hard vs. Soft Information in a Game

The structure of the game and past moves are ‘hard information: *irrevocably known*

Hard vs. Soft Information in a Game

The structure of the game and past moves are 'hard information: *irrevocably known*

Players' 'knowledge' of other players' rationality and 'knowledge' of her own future moves at nodes not yet reached are not of the same degree of certainty.

Hard vs. Soft Information in a Game

The structure of the game and past moves are 'hard information: *irrevocably known*

Players' 'knowledge' of other players' rationality and 'knowledge' of her own future moves at nodes not yet reached are not of the same degree of certainty.

What belief revision policy leads to BI?

Dynamic Rationality: The event R that all players are *rational* changes during the play of the game.

Players are assumed to be "incurably optimistic" about the rationality of their opponents.

What belief revision policy leads to BI?

Dynamic Rationality: The event R that all players are *rational* changes during the play of the game.

Players are assumed to be "incurably optimistic" about the rationality of their opponents.

Theorem (Baltag, Smets and Zvesper). Common knowledge of the game structure, of open future and *common stable belief* in dynamic rationality implies common belief in the backward induction outcome.

$$Ck(Struct_G \wedge F_G \wedge [\,!\,]CbRat) \rightarrow Cb(BI_G)$$

Concluding remarks

We are interested in reasoning about rational (and not-so rational) agents engaged in some form of social interaction.

We are interested in reasoning about rational (and not-so rational) agents engaged in some form of social interaction.

- Philosophy (social epistemology, philosophy of action)
- Game Theory
- Social Choice Theory
- AI (multiagent systems)

We are interested in reasoning about rational (and not-so rational) agents engaged in some form of social interaction.

What is a "rational agent"? What *are we modeling?*

We are interested in reasoning about rational (and not-so rational) agents engaged in some form of social interaction.

What is a "rational agent"? What *are we modeling?*

- has consistent preferences (complete, transitive)
- (acts as if she) maximizes expected utility
- reacts to observations
- revises beliefs when learning a *surprising* piece of information
- understands higher-order information
- plans for the future
- asks questions
- ????

We are interested in reasoning about rational (and not-so rational) agents engaged in some form of social interaction.

- playing a (card) game
- having a conversation
- executing a *social procedure* (voting, making a group decision)
- ....

Goal: incorporate/extend existing game-theoretic/social choice analyses

We are interested in reasoning about rational (and not-so rational) agents engaged in some form of social interaction.

There is a jungle of logical frameworks!

- logics of informational attitudes (knowledge, beliefs, certainty)
- logics of action & agency
- temporal logics/dynamic logics
- logics of motivational attitudes (preferences, intentions)
- deontic logics

(*Not to mention various game-theoretic/social choice models and logical languages for reasoning about them*)

We are interested in reasoning about rational (and not-so rational) agents engaged in some form of social interaction.

- How can we compare different logical frameworks addressing similar aspects of rational agency and social interaction?
- How should we combine logical systems which address different aspects of social interaction towards the goal of a comprehensive (formal) theory of rational agency?
- How does a logical analysis contribute to the broader discussion of rational agency and social interaction within philosophy and the social sciences?

Conclusions

We are interested in reasoning about rational (and not-so rational) agents engaged in some form of *social* situations.

Conclusions

We are interested in reasoning about rational (and not-so rational) agents engaged in some form of *social* situations.

What do the logical frameworks contribute to the discussion on rational agency?

Conclusions

We are interested in reasoning about rational (and not-so rational) agents engaged in some form of *social* situations.

What do the logical frameworks contribute to the discussion on rational agency?

Refine and test our intuitions: provide many answers to the question *what is a rational agent*? Explore how different answers *fit together*.

Conclusions

We are interested in reasoning about rational (and not-so rational) agents engaged in some form of *social* situations.

What do the logical frameworks contribute to the discussion on rational agency?

Merge with *Game Theory*/*Social Choice Theory*

- From a *Theory of Games* to a *Theory of Players*

J. van Benthem, EP and O. Roy. *A Theory of Play: A Logical Perspective on Games and Interaction*. Games, 2011.

- (Epistemic) foundations of game theory (rational-choice as a parameter)

Ingredients of a Logical Analysis of Rational Agency

$\Rightarrow$ informational attitudes (eg., knowledge, belief, certainty)

$\Rightarrow$ time, actions and ability

$\Rightarrow$ evaluative/motivational attitudes (eg., preferences)

$\Rightarrow$ pro-attitudes (eg., intentions)

$\Rightarrow$ group notions (eg., common knowledge and coalitional ability)

$\Rightarrow$ normative attitudes (eg., obligations, reasons)

Thank you!

Final Exam: Tuesday, December 13th, 5:30 PM - 8:30 PM in PH 125C