BI Puzzle

![Graph](image)

- Node A with R1 and D1
- Node B with r and d
- Node A with R2 and (6,6)
- Coordinates: (2,1), (1,6), (7,5)
BL Puzzle
BI Puzzle

\[ \begin{align*}
A & \quad R1 \quad B \\
D1 & \quad d \\
(2,1) & \quad (1,6) \\
& \quad r \quad (7,5)
\end{align*} \]
BI Puzzle

\[ (2,1) \quad (1,6) \]

\[ (7,5) \]
BI Puzzle

\[ A \quad R_1 \quad (1,6) \]
\[ D_1 \]
\[ (2,1) \]
BI Puzzle

A

R1

(1,6)

D1

(2,1)
BI Puzzle

\( A \)

\( D1 \)

(2,1)
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.
“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.”


Models of Extensive Games

\[ M(\Gamma) = \langle W, \sim_i, f, s \rangle \text{ where} \]

(A1) If \( w \sim_i w' \) then \( s_i(w) = s_i(w') \).

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

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

(F3) \( s(f(w, v)) \) and \( s(w) \) agree on the subtree of \( \Gamma \) 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 \( s(w') \) and \( s(w'') \) agree on the subtree of \( \Gamma \) below \( v \).
Rationality

\[ i \text{ is rational at } v \text{ in } w: \text{ for all strategies } s_i \neq s_i(w), \]
\[ h_i^v(s(w')) \geq h_i^v((s_{-i}(w'), s_i)) \text{ for some } w' \in [w]_i: \]
\[ \bigwedge_{v \in \Gamma_i} \bigwedge_{t_i \in \text{Strat}_i(\Gamma)} \neg K_i[h_i^v(s; t_i) > h_i^v(s)] \]

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

S-Rat: \ i \text{ is rational at vertex } v \text{ is } f(w, v) \text{ for every vertex } v \in \Gamma_i
(A1) If \( w \sim_i w' \) then \( s_i(w) = s_i(w') \).

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

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

(F3) \( s(f(w, v)) \) and \( s(w) \) agree on the subtree of \( \Gamma \) 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 \( s(w') \) and \( s(w'') \) 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 \( C(S\text{-Rat}) \subseteq BI \).

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 \( v_3 \)”
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)

“Off-line learning of rationality”

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

“Off-line learning of rationality”

```
    A
   / \  
  x   E
 /     \
1,0    y

    E
   /  
  y     A
 /     /  \
0,5   z   u

    A
   /     \
  6,4   5,5
```
“Off-line learning of rationality”
“Off-line learning of rationality”

\[
\begin{align*}
    x &\quad y &\quad z \\
    1,0 & & 0,100 & & 99,99 \\
    A & & E & & x \\
    0,100 & & & & 99,99 \\
    y & & & & \uparrow \text{rat} & & E & & \Rightarrow \\
\end{align*}
\]

Logic and Artificial Intelligence
The Dynamics of Rational Play

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.
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.

\[
Ck(S\text{truct}_G \land F_G \land [ \neg ] CbRat) \rightarrow Cb(Bl_G)
\]
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


- (Epistemic) foundations of game theory (rational-choice as a parameter)
Ingredients of a Logical Analysis of Rational Agency

⇒ informational attitudes (eg., knowledge, belief, certainty)

⇒ time, actions and ability

⇒ evaluative/motivational attitudes (eg., preferences)

⇒ pro-attitudes (eg., intentions)

⇒ group notions (eg., common knowledge and coalitional ability)

⇒ normative attitudes (eg., obligations, reasons)
Thank you!

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