Logics of Rational Agency

Lecture 5

Eric Pacuit

Tilburg Institute for Logic and Philosophy of Science
Tilburg Univeristy
ai.stanford.edu/~epacuit

July 31, 2009

✓ Introduction, Motivation and Background

✓ Basic Ingredients for a Logic of Rational Agency

✓ Logics of Rational Agency and Social Interaction, Part I

✓ Logics of Rational Agency and Social Interaction, Part II

Lecture 5: Conclusions and General Issues

Merging logics of rational agency

- Reasoning about information change (knowledge and time/actions)
- Knowledge, beliefs and certainty
- "Epistemizing" logics of action and ability: *knowing how to achieve* φ vs. *knowing that you can achieve* φ
- Entangling knowledge and preferences
- Planning/intentions (BDI)

Thus logical properties of beliefs can be derived from properties of preferences.

S. Morris. *The Logic of Belief and Belief Change: A Decision Theoretic Approach*. Journal of Economic Theory (1996).

The Framework

Let Ω be a set of states.

An **act** is a function $x : \Omega \to \mathbb{R}$. Let $\Re^\Omega$ be the set of all acts.

x_w for $w \in \Omega$ means that if the true state is w, then the agent receives prize x.

We write $x \succeq_w y$ the agent prefers x over y *provided the true state is w*

Belief Operators

A **belief operator** is a function $B : 2^\Omega \rightarrow 2^\Omega$

For $E \subseteq \Omega$, $w \in B(E)$ means the agent believes E at state w

B is normal if

- $B(\Omega) = \Omega$
- $B(E \cap F) = B(E) \cap B(F)$

Possibility function: $P : \Omega \rightarrow 2^\Omega$: set of states the agent considers possible at w

Defining Beliefs from Preferences

For $E \subseteq \Omega$ and two acts x and y, let (x_E, y_{-E}) denote the new act that is x on E and y on $-E$.

Defining Beliefs from Preferences

For $E \subseteq \Omega$ and two acts x and y, let (x_E, y_{-E}) denote the new act that is x on E and y on $-E$.

B **reflects** $\{\succeq_w\}_{w\in\Omega}$ provided for each $E \subseteq \Omega$

$$B(E) = \{w \mid (x_E, y_{-E}) \sim_w (x_E, z_{-E}) \quad \text{for all } x, y, z \in \Re^\Omega\}$$

Theorem If the preference relations are complete and transitive, then the derived belief operator is normal.

Alternative Definitions

For $x, y \in \Re^{\Omega}$, write $x \geq y$ if for each $w \in \Omega$, $x_w \geq y_w$

$x \gg y$ iff $x_w > y_w$ for each $w \in \Omega$

Alternative Definitions

For $x, y \in \Re^{\Omega}$, write $x \geq y$ if for each $w \in \Omega$, $x_w \geq y_w$

$x \gg y$ iff $x_w > y_w$ for each $w \in \Omega$

$$B^*(E) = \{w \mid (x_E, z_{-E}) \succeq_w (y_E, v_{-E}) \text{ for all } \quad x \gg y,\ x, y, z, v \in \Re^{\Omega}\}$$

Preferences are **monotone** if $x \gg y$ implies $x \succ_w y$ and $x \geq y$ implies $x \succeq_w y$ for all $w \in \Omega$.

Theorem B^* is normal if the preference relations are monotone, non-trivial and transitive.

S. Morris. *Alternative Definitions of Knowledge*. in *Epistemic Logic and the Foundations of Decision and Game Theory*, eds. Bacharach et al..

Coherency

A minimal rationality property relating together preferences at different states of Ω.

Preferences are **coherent** if choices made at different states can be seen as reflecting a true, metapreference ordering over acts.

A preference relation $\succ$ is a **meta-ordering** if it is complete, transitive, continuous and for all $x, y, z \in \Re^\Omega$

$$(x_w, z_{-w}) \succeq (y_w, z_{-w}) \Leftrightarrow x_w \geq y_w$$

Decision Problem

A decision problem is a finite set of acts D

$C_w[D] = \{x \in D \mid x \succeq_w y \text{for all } y \in D\}$

A decision rule is a function $f : \Omega \rightarrow D$

f is optimal provided for each $w \in \Omega$, $f(w) \in C_w[D]$

$C^*[D] = \{x \in \Re^\Omega \mid x_w = f_w(w) \text{ for all } w \in \Omega \text{ for some optimal } f\}$

Coherency

Preferences are coherent if there exists a meta-ordering $\succeq_*$ such that for each finite D, there exists $x \in C^*[D]$ such that $x \succeq_* y$, for all $y \in D$.

Theorem If preferences are coherent, then the beliefs reflecting them satisfy the knowledge axiom ($B(E) \subseteq E$) and positive introspection ($B(E) \subseteq B(B(E))$).

Knowledge and Voting

Theorem (Gibbard-Satterthwaite) There must be situations where it 'profits' a voter to vote *strategically*, i.e., not according to his or her *actual preference*.

Knowledge and Voting

Theorem (Gibbard-Satterthwaite) There must be situations where it 'profits' a voter to vote *strategically*, i.e., not according to his or her *actual preference*.

Under suitable conditions,

1. If P denotes the actual preference ordering of voter i,
2. and $\vec{Y}$ denotes the profile consisting of the preference orderings of all the other voters,
3. and S the aggregation rule,

Then the theorem says that there must exist P, Y, P' such that $S(P', Y) >_P S(P, Y)$.

Two Issues

Two Issues

1. What does it *mean* to vote strategically?

    - Voting as a game vs. voting as an act of communication

Two Issues

1. What does it *mean* to vote strategically?
   - Voting as a game vs. voting as an act of communication

2. When is the Gibbard-Satterthwaite Theorem '*effective*'?
   - The decision to strategize depends on the agents' *information* (eg. poll information).

S. Chopra, E. Pacuit and R. Parikh. *Knowledge-theoretic Properties of Strategic Voting*. JELIA 2004.

Voting Problem

Given a (finite) set X of **candidates**

and a (finite) set A of **voters**

each of whom have a **preference** over X

Devise a method F which aggregates the individual preferences to produce a collective decision (typically a subset of X)

Voting Procedures

- Type of vote, or **ballot**, that is recognized as admissible by the procedure: let $\mathcal{B}(X)$ be the set of admissible ballots for a set X of candidates
- A method to **count** a vector of ballots (one ballot for each voter) and select a winner (or winners)

Voting Procedures

- Type of vote, or **ballot**, that is recognized as admissible by the procedure: let $\mathcal{B}(X)$ be the set of admissible ballots for a set X of candidates
- A method to **count** a vector of ballots (one ballot for each voter) and select a winner (or winners)

Formally, A voting procedure for a set A of agents (with $|A| = n$) and a set X of candidates is a pair

$$(\mathcal{B}(X), \mathrm{Ag})$$

- $\mathcal{B}(X)$ is a set of ballots; and
- $\mathrm{Ag} : \mathcal{B}(X)^n \rightarrow 2^X$ (typically we are interested in the case where $|\mathrm{Ag}(\vec{b})| = 1$).

Examples

Plurality (Simple Majority)

- $\mathcal{B}(X) = X$
- Given $\vec{b} \in X^n$ and $x \in X$, let $\#_x(\vec{b}) = \sum_{\{i \mid b_i = x\}} 1$

$$\mathrm{Ag}(\vec{b}) = \{x \mid \#_x(\vec{b}) \text{ is maximal}\}$$

Examples

Plurality (Simple Majority)

- $\mathcal{B}(X) = X$
- Given $\vec{b} \in X^n$ and $x \in X$, let $\#_x(\vec{b}) = \sum_{\{i \mid b_i = x\}} 1$

$$\mathrm{Ag}(\vec{b}) = \{x \mid \#_x(\vec{b}) \text{ is maximal}\}$$

Approval Voting

- $\mathcal{B}(X) = 2^X$
- $\mathrm{Ag}(\vec{b}) = \{x \mid \#_x(\vec{b}) \text{ is maximal}\}$

Strategizing Functions

Fix the voters' true preferences: $\mathcal{P}^* = (P_1^*, \ldots, P_n^*)$

Given a vote profile $\vec{v}$ of *actual* votes, we ask whether voter i will change its vote if given another chance to vote.

Example I

The following example is due to [Brams & Fishburn]

$$P_A^* = o_1 > o_3 > o_2$$
$$P_B^* = o_2 > o_3 > o_1$$
$$P_C^* = o_3 > o_1 > o_2$$

Size	Group	I	II
4	A	$\mathbf{o_1}$	$\mathbf{o_1}$
3	B	o_2	o_2
2	C	o_3	$\mathbf{o_1}$

If the current winner is o, then agent i will switch its vote to some candidate o' provided

1. o' is one of the top two candidates as indicated by a poll
2. o' is preferred to the other top candidate

Example I

The following example is due to [Brams & Fishburn]

$$P^*_A = o_1 > o_3 > o_2$$
$$P^*_B = o_2 > o_3 > o_1$$
$$P^*_C = o_3 > o_1 > o_2$$

Size	Group	I	II
4	A	$\mathbf{o_1}$	$\mathbf{o_1}$
3	B	o_2	o_2
2	C	o_3	$\mathbf{o_1}$

If the current winner is o, then agent i will switch its vote to some candidate o' provided

1. o' is one of the top two candidates as indicated by a poll
2. o' is preferred to the other top candidate

Example II

$$P_A^* = (o_1, o_4, o_2, o_3)$$
$$P_B^* = (o_2, o_1, o_3, o_4)$$
$$P_C^* = (o_3, o_2, o_4, o_1)$$
$$P_D^* = (o_4, o_1, o_2, o_3)$$
$$P_E^* = (o_3, o_1, o_2, o_4)$$

Size	Group	I	II	III	IV
40	A	$\mathbf{o_1}$	o_1	o_4	$\mathbf{o_1}$
30	B	o_2	$\mathbf{o_2}$	$\mathbf{o_2}$	o_2
15	C	o_3	$\mathbf{o_2}$	$\mathbf{o_2}$	o_2
8	D	o_4	o_4	o_1	o_4
7	E	o_3	o_3	o_1	$\mathbf{o_1}$

If the current winner is o, then agent i will switch its vote to some candidate o' provided

1. i prefers o' to o, and
2. the current total for o' plus agent i's votes for o' is greater than the current total for o.

Example II

$$P_A^* = (o_1, o_4, o_2, o_3)$$
$$P_B^* = (o_2, o_1, o_3, o_4)$$
$$P_C^* = (o_3, o_2, o_4, o_1)$$
$$P_D^* = (o_4, o_1, o_2, o_3)$$
$$P_E^* = (o_3, o_1, o_2, o_4)$$

Size	Group	I	II	III	IV
40	A	$\mathbf{o_1}$	o_1	o_4	$\mathbf{o_1}$
30	B	o_2	$\mathbf{o_2}$	$\mathbf{o_2}$	o_2
15	C	o_3	$\mathbf{o_2}$	$\mathbf{o_2}$	o_2
8	D	o_4	o_4	o_1	o_4
7	E	o_3	o_3	o_1	$\mathbf{o_1}$

If the current winner is o, then agent i will switch its vote to some candidate o' provided

1. i prefers o' to o, and
2. the current total for o' plus agent i's votes for o' is greater than the current total for o.

Example II

$$P^*_A = (o_1, o_4, o_2, o_3)$$
$$P^*_B = (o_2, o_1, o_3, o_4)$$
$$P^*_C = (o_3, o_2, o_4, o_1)$$
$$P^*_D = (o_4, o_1, o_2, o_3)$$
$$P^*_E = (o_3, o_1, o_2, o_4)$$

Size	Group	I	II	III	IV
40	A	$\mathbf{o_1}$	o_1	o_4	$\mathbf{o_1}$
30	B	o_2	$\mathbf{o_2}$	$\mathbf{o_2}$	o_2
15	C	o_3	$\mathbf{o_2}$	$\mathbf{o_2}$	o_2
8	D	o_4	o_4	o_1	o_4
7	E	o_3	o_3	o_1	$\mathbf{o_1}$

If the current winner is o, then agent i will switch its vote to some candidate o' provided

1. i prefers o' to o, and
2. the current total for o' plus agent i's votes for o' is greater than the current total for o.

Example II

$$P_A^* = (o_1, o_4, o_2, o_3)$$
$$P_B^* = (o_2, o_1, o_3, o_4)$$
$$P_C^* = (o_3, o_2, o_4, o_1)$$
$$P_D^* = (o_4, o_1, o_2, o_3)$$
$$P_E^* = (o_3, o_1, o_2, o_4)$$

Size	Group	I	II	III	IV
40	A	$\mathbf{o_1}$	o_1	o_4	$\mathbf{o_1}$
30	B	o_2	$\mathbf{o_2}$	$\mathbf{o_2}$	o_2
15	C	o_3	$\mathbf{o_2}$	$\mathbf{o_2}$	o_2
8	D	o_4	o_4	o_1	o_4
7	E	o_3	o_3	o_1	$\mathbf{o_1}$

If the current winner is o, then agent i will switch its vote to some candidate o' provided

1. i prefers o' to o, and
2. the current total for o' plus agent i's votes for o' is greater than the current total for o.

Example III

$$P_A^* = (o_1, o_2, o_3)$$
$$P_B^* = (o_2, o_3, o_1)$$
$$P_C^* = (o_3, o_1, o_2)$$

Size	Group	I	II	III	IV	V	VI	VII	$\cdots$
40	A	o_1	o_1	o_2	o_2	o_2	o_1	o_1	o_1
30	B	o_2	o_3	o_3	o_2	o_2	o_2	o_3	o_3
30	C	o_3	o_3	o_3	o_3	o_1	o_1	o_1	o_3

Example III

$$P_A^* = (o_1, o_2, o_3)$$
$$P_B^* = (o_2, o_3, o_1)$$
$$P_C^* = (o_3, o_1, o_2)$$

Size	Group	I	II	III	IV	V	VI	VII	$\cdots$
40	A	o_1	o_1	o_2	o_2	o_2	o_1	o_1	o_1
30	B	o_2	o_3	o_3	o_2	o_2	o_2	o_3	o_3
30	C	o_3	o_3	o_3	o_3	o_1	o_1	o_1	o_3

Example III

$$P_A^* = (o_1, o_2, o_3)$$
$$P_B^* = (o_2, o_3, o_1)$$
$$P_C^* = (o_3, o_1, o_2)$$

Size	Group	I	II	III	IV	V	VI	VII	$\cdots$
40	A	o_1	o_1	o_2	o_2	o_2	o_1	o_1	o_1
30	B	o_2	o_3	o_3	o_2	o_2	o_2	o_3	o_3
30	C	o_3	o_3	o_3	o_3	o_1	o_1	o_1	o_3

Example III

$$P_A^* = (o_1, o_2, o_3)$$
$$P_B^* = (o_2, o_3, o_1)$$
$$P_C^* = (o_3, o_1, o_2)$$

Size	Group	I	II	III	IV	V	VI	VII	$\cdots$
40	A	o_1	o_1	o_2	o_2	o_2	o_1	o_1	o_1
30	B	o_2	o_3	o_3	o_2	o_2	o_2	o_3	o_3
30	C	o_3	o_3	o_3	o_3	o_1	o_1	o_1	o_3

Example III

$$P_A^* = (o_1, o_2, o_3)$$
$$P_B^* = (o_2, o_3, o_1)$$
$$P_C^* = (o_3, o_1, o_2)$$

Size	Group	I	II	III	IV	V	VI	VII	...
40	A	o_1	o_1	o_2	o_2	o_2	o_1	o_1	o_1
30	B	o_2	o_3	o_3	o_2	o_2	o_2	o_3	o_3
30	C	o_3	o_3	o_3	o_3	o_1	o_1	o_1	o_3

Example III

$$P_A^* = (o_1, o_2, o_3)$$
$$P_B^* = (o_2, o_3, o_1)$$
$$P_C^* = (o_3, o_1, o_2)$$

Size	Group	I	II	III	IV	V	VI	VII	...
40	A	o_1	o_1	o_2	o_2	o_2	o_1	o_1	o_1
30	B	o_2	o_3	o_3	o_2	o_2	o_2	o_3	o_3
30	C	o_3	o_3	o_3	o_3	o_1	o_1	o_1	o_3

Example III

$$P_A^* = (o_1, o_2, o_3)$$
$$P_B^* = (o_2, o_3, o_1)$$
$$P_C^* = (o_3, o_1, o_2)$$

Size	Group	I	II	III	IV	V	VI	VII	...
40	A	o_1	o_1	o_2	o_2	o_2	o_1	o_1	o_1
30	B	o_2	o_3	o_3	o_2	o_2	o_2	o_3	o_3
30	C	o_3	o_3	o_3	o_3	o_1	o_1	o_1	o_3

Example III

$$P_A^* = (o_1, o_2, o_3)$$
$$P_B^* = (o_2, o_3, o_1)$$
$$P_C^* = (o_3, o_1, o_2)$$

Size	Group	I	II	III	IV	V	VI	VII	⋯
40	A	o_1	o_1	o_2	o_2	o_2	o_1	o_1	o_1
30	B	o_2	o_3	o_3	o_2	o_2	o_2	o_3	o_3
30	C	o_3	o_3	o_3	o_3	o_1	o_1	o_1	o_3

Example III

$$P_A^* = (o_1, o_2, o_3)$$
$$P_B^* = (o_2, o_3, o_1)$$
$$P_C^* = (o_3, o_1, o_2)$$

Size	Group	I	II	III	IV	V	VI	VII	$\cdots$
40	A	o_1	o_1	o_2	o_2	o_2	o_1	o_1	o_1
30	B	o_2	o_3	o_3	o_2	o_2	o_2	o_3	o_3
30	C	o_3	o_3	o_3	o_3	o_1	o_1	o_1	o_3

Summary

Agents, knowing an aggregation function, will strategize if they know

a. enough about other agents' preferences and
b. that the output of the aggregation function of a changed preference will provide them with a more favorable result.

Beliefs & Plans/Intention

T. Icard, EP and Y. Shoham. *A Dynamic Logic of Belief and Intention*. Manuscript (2009).

Plan

- (Very!) Brief Discussion of Existing literature
- Belief-Intention Models
- Dynamics

Some Literature

Stemming from Bratman's planning theory of intention a number of logics of rational agency have been developed:

- Cohen and Levesque; Rao and Georgeff (BDI); Meyer, van der Hoek (KARO); Bratman, Israel and Pollack (IRMA); and others.

Some Literature

Stemming from Bratman's planning theory of intention a number of logics of rational agency have been developed:

- Cohen and Levesque; Rao and Georgeff (BDI); Meyer, van der Hoek (KARO); Bratman, Israel and Pollack (IRMA); and others.

Some common features

- Underlying temporal model
- Belief, Desire, Intention, Plans, Actions are defined with corresponding operators in a language

J.-J. Meyer and F. Veltman. *Intelligent Agents and Common Sense Reasoning*. Handbook of Modal Logic, 2007.

Intention Revision

Many of the frameworks do discuss some form of intention revision.

Intention Revision

Many of the frameworks do discuss some form of intention revision.

W. van der Hoek, W. Jamroga and M. Wooldridge. *Towards a Theory of Intention Revision*. Synthese, 2007.

Intention Revision

Many of the frameworks do discuss some form of intention revision.

W. van der Hoek, W. Jamroga and M. Wooldridge. *Towards a Theory of Intention Revision*. Synthese, 2007.

- Beliefs are sets of Linear Temporal Logic formulas (eg., $\bigcirc\varphi$)

Intention Revision

Many of the frameworks do discuss some form of intention revision.

W. van der Hoek, W. Jamroga and M. Wooldridge. *Towards a Theory of Intention Revision*. Synthese, 2007.

- Beliefs are sets of Linear Temporal Logic formulas (eg., $\bigcirc\varphi$)
- Desires are (possibly inconsistent) sets of Linear Temporal Logic formulas

Intention Revision

Many of the frameworks do discuss some form of intention revision.

W. van der Hoek, W. Jamroga and M. Wooldridge. *Towards a Theory of Intention Revision*. Synthese, 2007.

- Beliefs are sets of Linear Temporal Logic formulas (eg., $\bigcirc\varphi$)
- Desires are (possibly inconsistent) sets of Linear Temporal Logic formulas
- Practical reasoning rules: $\alpha \leftarrow \alpha_1, \alpha_2, \ldots, \alpha_n$

Intention Revision

Many of the frameworks do discuss some form of intention revision.

W. van der Hoek, W. Jamroga and M. Wooldridge. *Towards a Theory of Intention Revision*. Synthese, 2007.

- Beliefs are sets of Linear Temporal Logic formulas (eg., $\bigcirc\varphi$)
- Desires are (possibly inconsistent) sets of Linear Temporal Logic formulas
- Practical reasoning rules: $\alpha \leftarrow \alpha_1, \alpha_2, \ldots, \alpha_n$
- Intentions are derived from the agents current active plans (trees of practical reasoning rules)

Intention Revision

Many of the frameworks do discuss some form of intention revision.

W. van der Hoek, W. Jamroga and M. Wooldridge. *Towards a Theory of Intention Revision*. Synthese, 2007.

- Two types of beliefs: strong beliefs vs. weak beliefs (beliefs that take into account the agent's intentions)
- A dynamic update operator is defined ($[\Omega]\varphi$)

Our Framework

- Database/Planner picture
- Sources of beliefs
- What type of information does a planner provide? How do we represent a *plan*?
- Sources of dynamics: What can cause an agent's database to change?
- Changing/amending plans vs. revising/updating beliefs

Our Framework

1. *At a fixed moment*, a **choice situation** describes the current state-of-affairs (i.e., facts about the state-of-the-world), the tree of options that are available to the agent (i.e., the decision tree) and how actions change state of the world (i.e., the effect that performing an action will have on the state-of-the-world).

Our Framework

1. *At a fixed moment*, a **choice situation** describes the current state-of-affairs (i.e., facts about the state-of-the-world), the tree of options that are available to the agent (i.e., the decision tree) and how actions change state of the world (i.e., the effect that performing an action will have on the state-of-the-world).

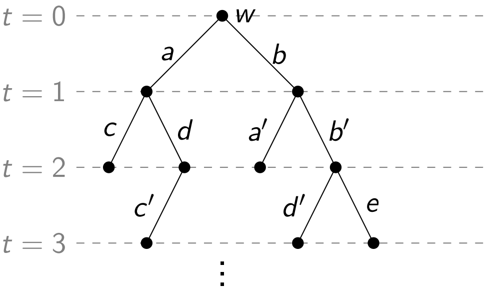
2. *At a fixed moment*, a **model** describes the agent's (current) beliefs (about the current state-of-the-world and what will become true in the future including options that will become available) and the agent's (current) *instructions from the Planner* (about future choices).

Our Framework

3. **Dynamic operators** representing each of the situations that may cause a change in beliefs and/or plans: learning a true fact, doing an action and receiving instructions from the Planner. These operators will describe how to relate models *at different moments*.

Choice Situations

$$\mathcal{M}_w = (W, \{R_a\}_{a \in \mathsf{Act}}, V, w)$$



Choice Situations: $\mathcal{L}_1$

$$\varphi := p \mid \varphi \wedge \varphi \mid \neg\varphi \mid \langle a \rangle \varphi$$

Choice Situations: $\mathcal{L}_1$

$$\varphi := \; p \mid \varphi \wedge \varphi \mid \neg\varphi \mid \langle a \rangle \varphi$$

- $\mathcal{M}_w \models p$ iff $w \in V(p)$
- $\mathcal{M}_w \models \varphi \wedge \psi$ iff $\mathcal{M}_w \models \varphi$ and $\mathcal{M}_w \models \psi$
- $\mathcal{M}_w \models \neg\varphi$ iff $\mathcal{M}_w \not\models \varphi$
- $\mathcal{M}_w \models \langle a \rangle \varphi$ iff $\exists x \; wR_a x$ and $\mathcal{M}_x \models \varphi$.

Choice Situations: $\mathcal{L}_1$

$$\varphi := p \mid \varphi \wedge \varphi \mid \neg\varphi \mid \langle a\rangle\varphi$$

- $\mathcal{M}_w \models p$ iff $w \in V(p)$
- $\mathcal{M}_w \models \varphi \wedge \psi$ iff $\mathcal{M}_w \models \varphi$ and $\mathcal{M}_w \models \psi$
- $\mathcal{M}_w \models \neg\varphi$ iff $\mathcal{M}_w \not\models \varphi$
- $\mathcal{M}_w \models \langle a\rangle\varphi$ iff $\exists x$ $wR_a x$ and $\mathcal{M}_x \models \varphi$.

Notation: If $\alpha = a_1 a_2 a_3 \cdots a_n$, $\langle\alpha\rangle\varphi := \langle a_1\rangle \cdots \langle a_n\rangle\varphi$

$$N\varphi := \bigwedge_{a\in \mathrm{Act}}[a]\varphi \quad [t]\varphi := \overbrace{N...N}^{t \text{ times}}\varphi$$

$$P\varphi := \bigvee_{a\in \mathrm{Act}}\langle a\rangle\varphi \quad \langle t\rangle\varphi := \overbrace{P...P}^{t \text{ times}}\varphi$$

Adding Beliefs

Standard picture where worlds are choice situations

Adding Beliefs

Standard picture where worlds are choice situations

$\mathcal{M}_w \preceq \mathcal{N}_v$: Choice situation $\mathcal{N}_v$ is at least as plausible as $\mathcal{M}_w$.

Adding Beliefs

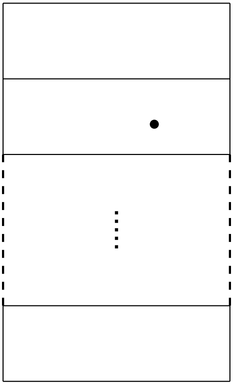
Standard picture where worlds are choice situations

$\mathcal{M}_w \preceq \mathcal{N}_v$: Choice situation $\mathcal{N}_v$ is at least as plausible as $\mathcal{M}_w$.

1. Beliefs are about available options, current and future state of affairs: $Bp \wedge B\langle a \rangle \langle b \rangle q$
2. Immediate options are *known*.
3. *In the static model*, restrict the language to only talk about *current* beliefs: $\langle a \rangle B\varphi$ is not well-formed

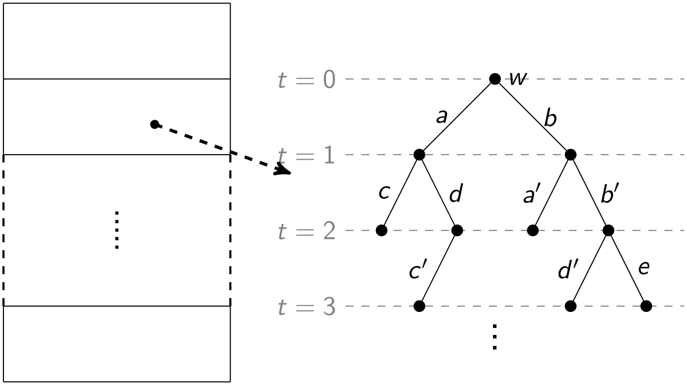
Belief Structures

$$\mathcal{M}_w = (W, \{R_a\}_{a \in \mathsf{Act}}, V, w)$$



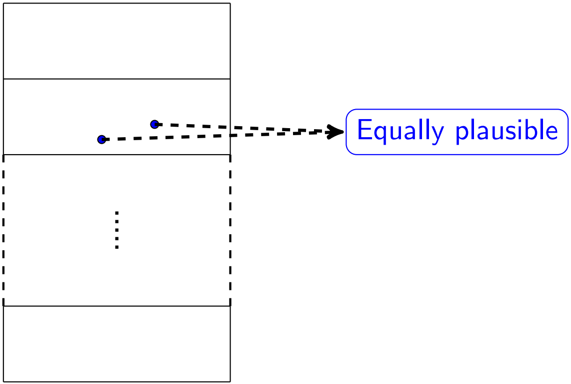
Belief Structures

$$\mathcal{M}_w = (W, \{R_a\}_{a \in \mathsf{Act}}, V, w)$$



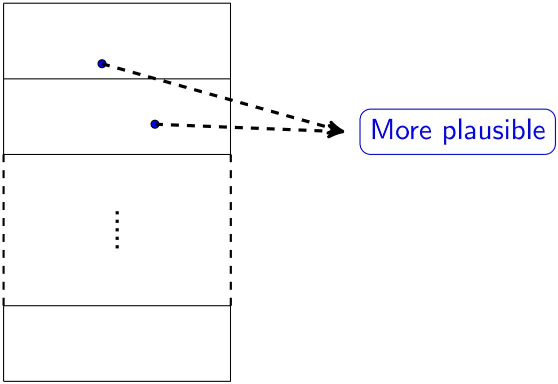
Belief Structures

$$\mathcal{M}_w = (W, \{R_a\}_{a \in \mathsf{Act}}, V, w)$$



Belief Structures

$$\mathcal{M}_w = (W, \{R_a\}_{a \in \mathsf{Act}}, V, w)$$



Belief Structures

Language ($\mathcal{L}_2$): $\varphi := \ \chi \mid \varphi \wedge \varphi \mid \neg\varphi \mid B(\varphi), \quad \chi \in \mathcal{L}_1$

Structures $\mathcal{B} = (S, \preceq, \mathcal{M}_w)$ is a *belief structure* if:

(i) S a set of choice situations

(ii) $\preceq$ is a plausibility ordering (reflexive, transitive, well-founded)

(iii) $\mathcal{M}_w \in S$.

Belief Structures

Language ($\mathcal{L}_2$): $\varphi := \ \chi \mid \varphi \wedge \varphi \mid \neg\varphi \mid B(\varphi), \quad \chi \in \mathcal{L}_1$

Structures $\mathcal{B} = (S, \preceq, \mathcal{M}_w)$ is a *belief structure* if:

(i) S a set of choice situations

(ii) $\preceq$ is a plausibility ordering (reflexive, transitive, well-founded)

(iii) $\mathcal{M}_w \in S$.

(iv) If wR_ax for some x in $\mathcal{M}$, then for all $\mathcal{N}_v \in S$ s.t. $\mathcal{M}_w \preceq \mathcal{N}_v$, there is some x' for which vR_ax' in $\mathcal{N}$.

(v) If $\mathcal{M}_w \preceq \mathcal{N}_v$ and vR_ax for some x in $\mathcal{N}$, there is some $x' \in W$ such that wR_ax' in $\mathcal{M}$.

Belief Structures

$$\mathcal{B} \Vdash \chi, \text{ iff } \mathcal{M}_w \models \chi.$$

$$\mathcal{B} \Vdash \varphi \wedge \psi, \text{ iff } \mathcal{B} \Vdash \varphi, \text{ and } \mathcal{B} \Vdash \psi.$$

$$\mathcal{B} \Vdash \neg\varphi, \text{ iff } \mathcal{B} \nVdash \varphi.$$

$$\mathcal{B} \Vdash B(\varphi), \text{ iff } \text{ for all } \mathcal{N}_v \in Min_{\preceq}(S), \ \mathcal{B}, \mathcal{N}_v \Vdash \varphi.$$

Completeness

1. Standard proof works for the class of choice situations
2. The class of belief structures is also easily axiomatized ($\Box\varphi$ means φ is true an all worlds at least as plausible as the current world):
   - **KD45** for B
   - $\langle a\rangle\top \rightarrow \Box(\langle a\rangle\top)$
   - $\Diamond(\langle a\rangle\top) \rightarrow \langle a\rangle\top$

Instructions

At each moment there are *instructions* from the Planner: We assume that at each moment, there are some instructions about future choices that the agent has agreed to follow (if he can).

Instructions

At each moment there are *instructions* from the Planner: We assume that at each moment, there are some instructions about future choices that the agent has agreed to follow (if he can).

1. A *complete plan*, for each moment the specific action $a \in \mathsf{Act}$ the agent will perform.

Instructions

At each moment there are *instructions* from the Planner: We assume that at each moment, there are some instructions about future choices that the agent has agreed to follow (if he can).

1. A *complete plan*, for each moment the specific action $a \in \text{Act}$ the agent will perform.
2. The instructions may be *partial*: finite list of pairs (a, t) where $a \in \text{Act}$ and $t \in \mathbb{N}$.

Instructions

At each moment there are *instructions* from the Planner: We assume that at each moment, there are some instructions about future choices that the agent has agreed to follow (if he can).

1. A *complete plan*, for each moment the specific action $a \in \text{Act}$ the agent will perform.
2. The instructions may be *partial*: finite list of pairs (a, t) where $a \in \text{Act}$ and $t \in \mathbb{N}$.
3. The instructions may be *conditional*: do a at time t provided φ is true.

Instructions

At each moment there are *instructions* from the Planner: We assume that at each moment, there are some instructions about future choices that the agent has agreed to follow (if he can).

1. A *complete plan*, for each moment the specific action $a \in \text{Act}$ the agent will perform.
2. The instructions may be *partial*: finite list of pairs (a, t) where $a \in \text{Act}$ and $t \in \mathbb{N}$.
3. The instructions may be *conditional*: do a at time t provided φ is true.
4. Rather than instructing the agent to follow a specific (partial, conditional) plan, the Planner simply restricts the choices that are available to the agent in the future.

Instructions

At each moment there are *instructions* from the Planner: We assume that at each moment, there are some instructions about future choices that the agent has agreed to follow (if he can).

1. A *complete plan*, for each moment the specific action $a \in \text{Act}$ the agent will perform.
2. The instructions may be *partial*: finite list of pairs (a, t) where $a \in \text{Act}$ and $t \in \mathbb{N}$.
3. The instructions may be *conditional*: do a at time t provided φ is true.
4. Rather than instructing the agent to follow a specific (partial, conditional) plan, the Planner simply restricts the choices that are available to the agent in the future.
5. The Planner may provide more complicated structure (subplan structure, goals, etc.)

Instructions

At each moment there are *instructions* from the Planner: We assume that at each moment, there are some instructions about future choices that the agent has agreed to follow (if he can).

1. A *complete plan*, for each moment the specific action $a \in \text{Act}$ the agent will perform.
2. The instructions may be *partial*: finite list of pairs (a, t) where $a \in \text{Act}$ and $t \in \mathbb{N}$.
3. The instructions may be *conditional*: do a at time t provided φ is true.
4. Rather than instructing the agent to follow a specific (partial, conditional) plan, the Planner simply restricts the choices that are available to the agent in the future.
5. The Planner may provide more complicated structure (subplan structure, goals, etc.)

Belief-Intention Structures

$\mathfrak{B} = (S, \preceq, I, \mathcal{M}_w)$ is a *belief-intention structure* where

- $(S, \preceq, \mathcal{M}_w)$ is a belief structure
- and I is a finite set of pairs (a, t), such that $a \in$ Act and $t \in \mathbb{N}$, and

Belief-Intention Structures

$\mathfrak{B} = (S, \preceq, I, \mathcal{M}_w)$ is a *belief-intention structure* where

- $(S, \preceq, \mathcal{M}_w)$ is a belief structure
- and I is a finite set of pairs (a, t), such that $a \in \text{Act}$ and $t \in \mathbb{N}$, and
- **Belief-Intention Coherency**: There exists some $\mathcal{N}_v \in Min_{\preceq}(S)$ such and $\vec{a}$ in $\mathcal{N}$, such that for each $(b, t) \in I$, $b = a_t$

  We say $\mathcal{N}_v$ *admits* I, and that the sequence $\vec{a}$ is a *satisfying sequence* for I.

Belief-Intention Structures: Language

Language: $\varphi := \chi \mid \varphi \wedge \varphi \mid \neg\varphi \mid B(\varphi) \mid \mathcal{I}_{a,t} \mid B^I(\varphi)$
(with $\chi \in \mathcal{L}_1$)

$B\varphi$: the agent believes φ

$B^I\varphi$: the agent believes φ given that the instructions are followed

$\mathcal{I}_{a,t}$: the agent intends to do a, t units from now

Belief-Intention Structure: Truth

$\mathfrak{B} = (S, \preceq, I, \mathcal{M}_w)$ is a *belief-intention structure*.

$$\mathfrak{B} \Vdash \mathcal{I}_{a,t}, \text{ iff } (a, t) \in I.$$

Belief-Intention Structure: Truth

$\mathfrak{B} = (S, \preceq, I, \mathcal{M}_w)$ is a *belief-intention structure*.

$\mathfrak{B} \Vdash \mathcal{I}_{a,t}$, iff $(a, t) \in I$.

$\mathfrak{B} \Vdash B(\varphi)$, iff for all $\mathcal{N}_v \in Min_{\preceq}(S)$, $(S, \preceq, I, \mathcal{N}_v) \Vdash \varphi$.

Belief-Intention Structure: Truth

$\mathfrak{B} = (S, \preceq, I, \mathcal{M}_w)$ is a *belief-intention structure*.

$\mathfrak{B} \Vdash \mathcal{I}_{a,t}$, iff $(a, t) \in I$.

$\mathfrak{B} \Vdash B(\varphi)$, iff for all $\mathcal{N}_v \in Min_{\preceq}(S)$, $(S, \preceq, I, \mathcal{N}_v) \Vdash \varphi$.

$\mathfrak{B} \Vdash B^I(\varphi)$, iff for all $\mathcal{N}_v \in Min_{\preceq}(S)$ *admitting* I, $(S', \preceq', I, \mathcal{N}_v^I) \Vdash \varphi$

where all choice situations are restricted to satisfying sequences.

Completeness

Theorem The class of all belief-intention structures is axiomatizable.

Completeness

Theorem The class of all belief-intention structures is axiomatizable.

Axioms for Belief

- **KD45** axioms and rules for B and B^I
- $B(\varphi) \leftrightarrow B^I(B(\varphi))$
- $\neg B(\varphi) \rightarrow B^I(\neg B(\varphi))$
- $B^I(\varphi) \leftrightarrow B(B^I(\varphi))$
- $\neg B^I(\varphi) \rightarrow B(\neg B^I(\varphi))$
- $B^I(\varphi) \rightarrow \widehat{B}(\varphi)$

Completeness

Theorem The class of all belief-intention structures is axiomatizable.

<u>Consistency of Intentions and Beliefs</u>

- $\mathcal{I}_{a,t} \leftrightarrow B(\mathcal{I}_{a,t}) \leftrightarrow B^I(\mathcal{I}_{a,t})$
- $\neg\mathcal{I}_{a,t} \leftrightarrow B(\neg\mathcal{I}_{a,t}) \leftrightarrow B^I(\neg\mathcal{I}_{a,t})$
- $\mathcal{I}_{a,t} \rightarrow B^I(\langle\!\![t]\!\!\rangle(\langle a\rangle\top \wedge \bigwedge_{b\neq a\in \mathsf{Act}}[b]\bot))$
- $B^I(\bigvee[\vec{a}]\varphi) \rightarrow (B(\bigvee[\vec{a}]\varphi) \vee \bigvee \underline{\vec{a}})$
- $B(\bigwedge[\vec{a}]\varphi \rightarrow \bigvee[\vec{b}]\psi) \rightarrow (B^I(\bigwedge[\vec{a}]\varphi \rightarrow \bigvee[\vec{b}]\psi) \vee \bigvee \underline{\vec{a}})$

Dynamics

There are three sources of dynamics:

1. Nature can reveal (true) facts about the current choice situation (eg., facts that are true, choices that are available/not available in the future).

Dynamics

There are three sources of dynamics:

1. Nature can reveal (true) facts about the current choice situation (eg., facts that are true, choices that are available/not available in the future).

2. The agent can decide to perform an action (which in turn forces Nature to reveal certain information such as which actions become available).

Dynamics

There are three sources of dynamics:

1. Nature can reveal (true) facts about the current choice situation (eg., facts that are true, choices that are available/not available in the future).

2. The agent can decide to perform an action (which in turn forces Nature to reveal certain information such as which actions become available).

3. The Panner can amend the agent's current set of instructions.

Dynamics

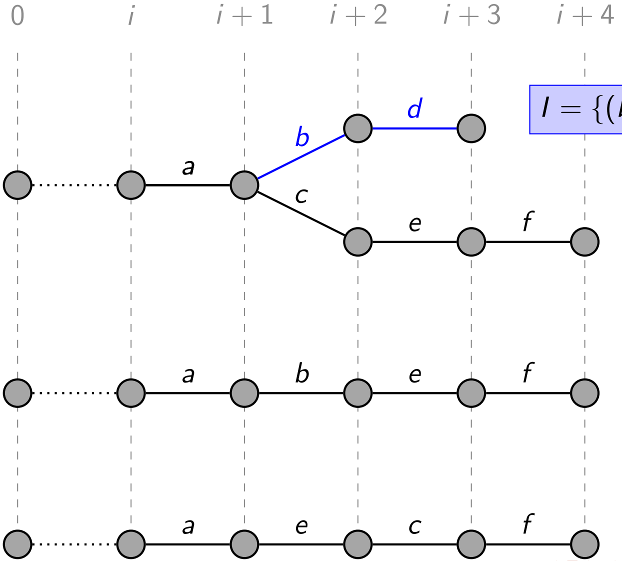
There are three sources of dynamics:

1. Nature can reveal (true) facts about the current choice situation (eg., facts that are true, choices that are available/not available in the future).

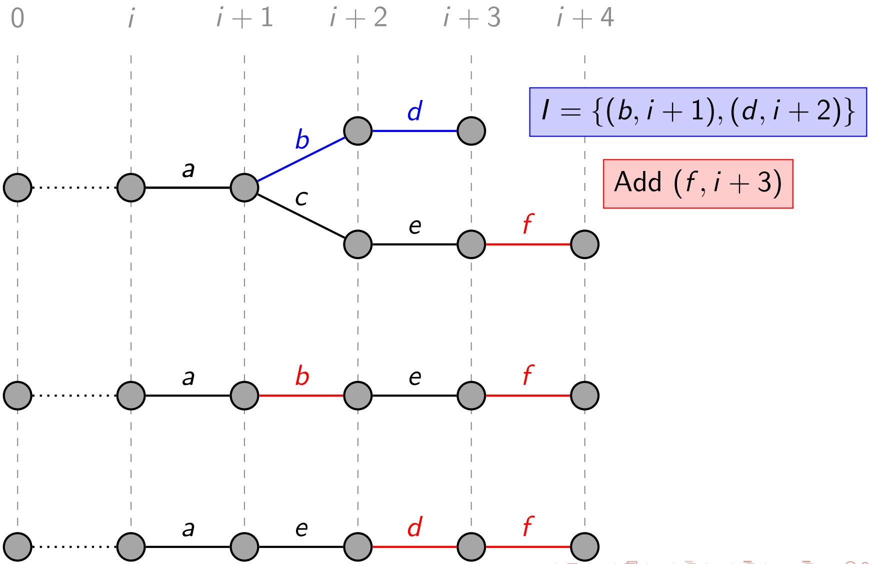
2. The agent can decide to perform an action (which in turn forces Nature to reveal certain information such as which actions become available).

3. The Panner can amend the agent's current set of instructions.

We assume that only doing an action moves time forward. However, all three types of events may change the agent's beliefs and current instructions.



$I = \{(b, i+1), (d, i+2)\}$


0
i
i + 1
i + 2
i + 3
i + 4
a
b
c
d
e
f
a
b
e
f
a
e
d
f
I = {(b, i + 1), (d, i + 2)}
Add (f, i + 3)

Selection Function

A **selection function** γ maps a set of choice situations $\mathcal{B}$ a finite set of action-time pairs C to a finite set of action time pairs:
$\gamma : \mathcal{P}(ChoicSit) \times \mathcal{P}_{<\omega}(Int) \rightarrow \mathcal{P}_{<\omega}(Int)$

1. $\gamma(\mathcal{B}, C) \subseteq C$
2. $\gamma(\mathcal{B}, C)$ is coherent with $\mathcal{B}$.

Selection Functions

let $\mathcal{B}$ be a set of choice situations (representing the agents current beliefs), I a set of action-time pairs (representing the agents current intentions) and (a, t) an intention

Selection Functions

let $\mathcal{B}$ be a set of choice situations (representing the agents current beliefs), I a set of action-time pairs (representing the agents current intentions) and (a, t) an intention

- (consistency) If $I \cup \{(a, t)\}$ is consistent with $\mathcal{B}$ then $\gamma(\mathcal{B}, I \cup \{(a, t)\}) = I \cup \{(a, t)\}$

Selection Functions

let $\mathcal{B}$ be a set of choice situations (representing the agents current beliefs), I a set of action-time pairs (representing the agents current intentions) and (a, t) an intention

- (consistency) If $I \cup \{(a, t)\}$ is consistent with $\mathcal{B}$ then $\gamma(\mathcal{B}, I \cup \{(a, t)\}) = I \cup \{(a, t)\}$
- (success) $(a, t) \in \gamma(\mathcal{B}, I \cup \{(a, t)\})$

Selection Functions

let $\mathcal{B}$ be a set of choice situations (representing the agents current beliefs), I a set of action-time pairs (representing the agents current intentions) and (a, t) an intention

- (consistency) If $I \cup \{(a, t)\}$ is consistent with $\mathcal{B}$ then $\gamma(\mathcal{B}, I \cup \{(a, t)\}) = I \cup \{(a, t)\}$
- (success) $(a, t) \in \gamma(\mathcal{B}, I \cup \{(a, t)\})$
- (minimal change) If $I' \subseteq I$ and $I' \cup \{(a, t)\}$ is consistent with $\mathcal{B}$ then $I' \subseteq \gamma(\mathcal{B}, I \cup \{(a, t)\})$

Selection Functions

let $\mathcal{B}$ be a set of choice situations (representing the agents current beliefs), I a set of action-time pairs (representing the agents current intentions) and (a, t) an intention

- (consistency) If $I \cup \{(a, t)\}$ is consistent with $\mathcal{B}$ then $\gamma(\mathcal{B}, I \cup \{(a, t)\}) = I \cup \{(a, t)\}$
- (success) $(a, t) \in \gamma(\mathcal{B}, I \cup \{(a, t)\})$
- (minimal change) If $I' \subseteq I$ and $I' \cup \{(a, t)\}$ is consistent with $\mathcal{B}$ then $I' \subseteq \gamma(\mathcal{B}, I \cup \{(a, t)\})$
- Other properties may depend on the structure of the plans:
  - if $\{(a_1, t_1), \ldots, (a_n, t_n)\} \subseteq I$ form a (sub)plan, then either $\{(a_1, t_1), \ldots, (a_n, t_n)\} \subseteq \gamma(\mathcal{B}, I \cup \{(a, t)\})$ or $\{(a_1, t_1), \ldots, (a_n, t_n)\} \cap \gamma(\mathcal{B}, I \cup \{(a, t)\}) = \emptyset$

Incorporating a new intention

.

- $[+(a,t)]\varphi$: after adopting the intention to do a at time t, φ is true.

- Given a selection function γ, let $I + a = \gamma(\mathcal{B}, I \cup \{(a,t)\})$ be the new set of intentions where $\mathcal{B}$ is the current minimal set of choice situations and I the current set of intentions.

Observing a true fact

- $[\varphi]\psi$ after observing that φ is true then ψ is true.

- The precondition is that φ is true. We also assume that φ is in the language $\mathcal{L}_1$.

- $\mathfrak{B}^\varphi = (S', \preceq', I', \mathcal{M}'_w)$ where $S' = \{\mathcal{N}_v \in S \mid \mathcal{N}_v \models \varphi\}$, $\preceq' = \preceq \cap S'$, $I' = I$ and $V'(p) = V(p) \cap S'$.

Doing an action

- $[DO(a)]\varphi$: "after the agent does action a, then φ is true"
- The precondition is that action a is possible in the actual choice situation
- We may assume further that the agent can only do something *currently* consistent with his intentions.

Doing an action

- The result of doing an action a is the belief-intention structure $\mathfrak{B}_a$ is constructed by first incorporating the fact that a has been executed, so the new set of states are $S' = \{\mathcal{N}_{v'}^{do(a)} \mid \mathcal{N}_v \in S\}$.

  Next the agent observes which actions are available. I.e., if Opt is the (finite) set of immediately available in $\mathcal{M}_{w'}^{do(a)}$ then

$$\bigwedge_{a \in Opt} \langle a \rangle \top \wedge \bigwedge_{b \notin Opt} [b] \bot$$

  is announced

- This may result in a situation where the agents intention set I is no longer consistent with the new beliefs.

Conclusions

Conclusions

- Reexamined some foundational issues in logics of belief and intentions/plans.

Conclusions

- Reexamined some foundational issues in logics of belief and intentions/plans.

- A number of questions remain about how to axiomatize the dynamic operators

Conclusions

We are interested in reasoning about rational agents interacting in *social* situations.

Conclusions

We are interested in reasoning about rational agents interacting in *social* situations.

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

Conclusions

We are interested in reasoning about rational agents interacting in *social* situations.

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

- Normative vs. Descriptive

Conclusions

We are interested in reasoning about rational agents interacting in *social* situations.

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

- Normative vs. Descriptive
- refine and test our intuitions: provide many answers to the question *what is a rational agent*?

Conclusions

We are interested in reasoning about rational agents interacting in *social* situations.

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

- Normative vs. Descriptive
- refine and test our intuitions: provide many answers to the question *what is a rational agent*?
- (epistemic) foundations of game theory
  Logic *and* Game Theory, not Logic in place of Game Theory.

Conclusions

We are interested in reasoning about rational agents interacting in *social* situations.

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

- Normative vs. Descriptive
- refine and test our intuitions: provide many answers to the question *what is a rational agent*?
- (epistemic) foundations of game theory
  Logic *and* Game Theory, not Logic in place of Game Theory.
- Social Software: Verify properties of social procedures
  - *Refine existing social procedures or suggest new ones*

R. Parikh. *Social Software*. *Synthese* **132** (2002).

Conclusions

- Many types of informational attitudes: "hard" knowledge, belief, belief about the future state of affairs, "intention" based beliefs, revisable beliefs, safe beliefs.

- Where does the "protocol" come from? What do the agents know about the protocol?

Logics of Rational Agency

- What's going on in the area:
  www.loriweb.org
- LORI-II, October 8 - 11, 2009, Chongqing, China
  loriweb.org
- Special Issue of Synthese: Knowledge, Rationality and Interaction. *Logic and Intelligent Interaction*, Volume 169, Number 2 / July, 2009
  (eds. T. Agotnes, J. van Benthem and EP)
- New subarea of Stanford Encyclopedia of Philosophy on logics and rational agency
  (eds. J. van Benthem, E. Pacuit, and O. Roy)

Calls for....

- **Papers**: LOFT 2010. University of Toulouse, July 21 - 23. Deadline: March 15, 2010.
- **Course/Workshop Proposals**: NASSLLI, Indiana Univeristy, Bloomington. Deadline: September 15.
- **Ph.D. position**: TiLPS, Tilburg University, "A formal analysis of social procedures". Deadline: October 15 (to start in February).

Thank you!