A Dynamic Logic of Belief and Intention

Eric Pacuit (with Thomas Icard and Yoav Shoham)

Tilburg University

September 8, 2009

Plan

- Introduction, Motivation and Background: Logics of Rational Agency
- (Very!) Brief Discussion of Existing literature
- Belief-Intention Models
- Dynamics

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

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

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

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

What is a rational agent?

- maximize expected utility (instrumentally rational)
- react to observations
- revise beliefs when learning a *surprising* piece of information
- understand higher-order information
- plans for the future
- ????

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

There is a jungle of formal systems!

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

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

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

There is a jungle of formal systems!

- How do we compare different logical systems studying the same phenomena?
- How *complex* is it to reason about rational agents?
- (How) should we *merge* the various logical systems?
- What do the logical frameworks contribute to the discussion on rational agency?

and logical languages for reasoning about them)

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

- playing a card game
- having a conversation
- executing a *social procedure*
- ....

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

Formally, a game is described by its strategy sets and payoff functions.

Formally, a game is described by its strategy sets and payoff functions. But in real life, may other parameters are relevant; there is a lot more going on. Situations that substantively are vastly different may nevertheless correspond to precisely the same strategic game.

Formally, a game is described by its strategy sets and payoff functions. But in real life, may other parameters are relevant; there is a lot more going on. Situations that substantively are vastly different may nevertheless correspond to precisely the same strategic game. For example, in a parliamentary democracy with three parties, the winning coalitions are the same whether the parties hold a third of the seats, or, say, 49%, 39%, and 12 % respectively.

Formally, a game is described by its strategy sets and payoff functions. But in real life, may other parameters are relevant; there is a lot more going on. Situations that substantively are vastly different may nevertheless correspond to precisely the same strategic game. For example, in a parliamentary democracy with three parties, the winning coalitions are the same whether the parties hold a third of the seats, or, say, 49%, 39%, and 12 % respectively. But the political situations are quite different.

Formally, a game is described by its strategy sets and payoff functions. But in real life, may other parameters are relevant; there is a lot more going on. Situations that substantively are vastly different may nevertheless correspond to precisely the same strategic game. For example, in a parliamentary democracy with three parties, the winning coalitions are the same whether the parties hold a third of the seats, or, say, 49%, 39%, and 12 % respectively. But the political situations are quite different. The difference lies in the attitudes of the players, in their expectations about each other, in custom, and in history, though the rules of the game do not distinguish between the two situations.

R. Aumann and J. H. Dreze. *Rational Expectation in Games*. American Economic Review (2008).

Logics of Rational Agency

Basic Ingredients

- What are the basic building blocks? (the nature of time (continuous or discrete/branching or linear), how (primitive) *events* or *actions* are represented, how *causal* relationships are represented and what constitutes a *state of affairs*.)

Basic Ingredients

- What are the basic building blocks? (the nature of time (continuous or discrete/branching or linear), how (primitive) *events* or *actions* are represented, how *causal* relationships are represented and what constitutes a *state of affairs*.)

- Single agent vs. many agents.

Basic Ingredients

- What are the basic building blocks? (the nature of time (continuous or discrete/branching or linear), how (primitive) *events* or *actions* are represented, how *causal* relationships are represented and what constitutes a *state of affairs*.)
- Single agent vs. many agents.
- What the the primitive operators?
  - Informational attitudes
  - Motivational attitudes
  - Normative attitudes

Basic Ingredients

- What are the basic building blocks? (the nature of time (continuous or discrete/branching or linear), how (primitive) *events* or *actions* are represented, how *causal* relationships are represented and what constitutes a *state of affairs*.)
- Single agent vs. many agents.
- What the the primitive operators?
  - Informational attitudes
  - Motivational attitudes
  - Normative attitudes
- Static vs. dynamic

- ✓ informational attitudes (eg., knowledge, belief, certainty)
- ✓ group notions (eg., common knowledge and coalitional ability)
- ✓ time, actions and ability
- ✓ motivational attitudes (eg., preferences)
- ✓ normative attitudes (eg., obligations)

General Issues

Once a semantics and language are fixed, then standard questions can be asked: eg. develop a proof theory, completeness, decidability, model checking.

General Issues

How should we *compare* the different logical systems?

- Embedding one logic in another:

General Issues

How should we *compare* the different logical systems?

- Embedding one logic in another: *coalition logic* is a fragment of ATL ($tr([C]\varphi) = \langle\langle C\rangle\rangle \bigcirc \varphi$)

General Issues

How should we *compare* the different logical systems?

- Embedding one logic in another: *coalition logic* is a fragment of ATL ($tr([C]\varphi) = \langle\langle C \rangle\rangle \bigcirc \varphi$)

- Compare different models for a fixed language:

General Issues

How should we *compare* the different logical systems?

- Embedding one logic in another: *coalition logic* is a fragment of ATL ($tr([C]\varphi) = \langle\langle C \rangle\rangle \bigcirc \varphi$)

- Compare different models for a fixed language:
  - Alternating-Time Temporal Logics: Three different semantics for the ATL language.

V. Goranko and W. Jamroga. *Comparing Semantics of Logics for Multiagent Systems*. KRA, 2004.

General Issues

How should we *compare* the different logical systems?

- Embedding one logic in another: *coalition logic* is a fragment of ATL ($tr([C]\varphi) = \langle\langle C \rangle\rangle \bigcirc \varphi$)

- Compare different models for a fixed language:
    - Alternating-Time Temporal Logics: Three different semantics for the ATL language.

V. Goranko and W. Jamroga. *Comparing Semantics of Logics for Multiagent Systems*. KRA, 2004.

- Comparing different frameworks:

General Issues

How should we *compare* the different logical systems?

- Embedding one logic in another: *coalition logic* is a fragment of ATL ($tr([C]\varphi) = \langle\langle C\rangle\rangle \bigcirc \varphi$)
- Compare different models for a fixed language:
  - Alternating-Time Temporal Logics: Three different semantics for the ATL language.

V. Goranko and W. Jamroga. *Comparing Semantics of Logics for Multiagent Systems*. KRA, 2004.

- Comparing different frameworks: eg. PDL vs. Temporal Logic, PDL vs. STIT, STIT vs. ATL, etc.

General Issues

How should we *merge* the different logical systems?

General Issues

How should we *merge* the different logical systems?

- Combining logics is hard!

D. Gabbay, A. Kurucz, F. Wolter and M. Zakharyaschev. *Many Dimensional Modal Logics: Theory and Applications*. 2003.

General Issues

How should we *merge* the different logical systems?

- Combining logics is hard!

D. Gabbay, A. Kurucz, F. Wolter and M. Zakharyaschev. *Many Dimensional Modal Logics: Theory and Applications*. 2003.

Theorem $\Box\varphi \leftrightarrow \varphi$ is provable in combinations of Epistemic Logics and PDL with certain "cross axioms" ($\Box[a]\varphi \leftrightarrow [a]\Box\varphi$) (and full substitution).

R. Schmidt and D. Tishkovsky. *On combinations of propositional dynamic logic and doxastic modal logics*. JOLLI, 2008.

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)

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)

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 many 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 many 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.

Bratman's Planning Theory of Intention

M. Bratman. *Intentions, Plans and Practical Reason*. Harvard University Press (1987).

Bratman's Planning Theory of Intention

M. Bratman. *Intentions, Plans and Practical Reason*. Harvard University Press (1987).

A plan is a *conduct-controlling* mental attitude

Bratman's Planning Theory of Intention

M. Bratman. *Intentions, Plans and Practical Reason*. Harvard University Press (1987).

A plan is a *conduct-controlling* mental attitude

An intention is a component of a future-directed plan.

Bratman's Planning Theory of Intention

An agent commits to a (partial) plan that is

Bratman's Planning Theory of Intention

An agent commits to a (partial) plan that is

1. means-end coherent,

Bratman's Planning Theory of Intention

An agent commits to a (partial) plan that is

1. means-end coherent,
2. consistent with the agent's current beliefs and

Bratman's Planning Theory of Intention

An agent commits to a (partial) plan that is

1. means-end coherent,
2. consistent with the agent's current beliefs and
3. *stable* (i.e., plans *normally* resist reconsideration)

Bratman's Planning Theory of Intention

An agent commits to a (partial) plan that is

1. means-end coherent,
2. consistent with the agent's current beliefs and
3. *stable* (i.e., plans *normally* resist reconsideration) *"an agent's habits and dispositions concerning the reconsideration or nonreconsideration of a prior intention or plan determine the stability of that intention or plan". Furthermore, "The stability of [the agent's] plans will generally not be an isolated feature of those plans but will be linked to other features of [the agent's] psychology"*

Bratman's Planning Theory of Intention

Central to Bratman's theory is the idea that these partial plans direct the agent's deliberation by "constrain[ing] what options are considered relevant":

> *"plans narrow the scope of the deliberation to a limited set of options. And they help to answer a question that tends to remain unanswered in traditional decision theory, namely: where do decision problems come from?"*

A Methodological Issue

What are we formalizing? How will the logical framework be *used*?

A Methodological Issue

What are we formalizing? How will the logical framework be *used*?

Two Extremes:

1. Formalizing a (philosophical) theory of rational agency:

A Methodological Issue

What are we formalizing? How will the logical framework be *used*?

Two Extremes:

1. Formalizing a (philosophical) theory of rational agency: philosophers as intuition pumps generating "problems" for the logical frameworks.

A Methodological Issue

What are we formalizing? How will the logical framework be *used*?

Two Extremes:

1. Formalizing a (philosophical) theory of rational agency: philosophers as intuition pumps generating "problems" for the logical frameworks.

2. Reasoning *about* multiagent systems.

A Methodological Issue

What are we formalizing? How will the logical framework be *used*?

Two Extremes:

1. Formalizing a (philosophical) theory of rational agency: philosophers as intuition pumps generating "problems" for the logical frameworks.

2. Reasoning *about* multiagent systems.Three main applications of BDI logics: 1. a specification language for a MAS, 2. a programming language, and 3. verification language.

W. van der Hoek and M. Wooldridge. *Towards a logic of rational agency*. Logic Journal of the IGPL 11 (2), 2003.

C & L Logic of Intention

1. Intentions normally pose problems for the agent; the agent needs to determine a way to achieve them.
2. Intentions provide a “screen of admissibility” for adopting other intentions.
3. Agents “track” the success of their attempts to achieve their intentions.
4. If an agent intends to achieve p, then
   - 4.1 The agent believes p is possible
   - 4.2 The agent does not believe he will not bring abut p
   - 4.3 Under certain conditions, the agent believes he will bring about p
   - 4.4 Agents need not intend all the expected side-effects of their intentions.

C & L Logic of Intention

$$(\mathsf{PGOAL}_i p) := (\mathsf{GOAL}_i(\mathsf{LATER} p)) \wedge$$
$$(\mathsf{BEL}_i \neg p) \wedge [\mathsf{BEFORE}((\mathsf{BEL}_i p) \vee (\mathsf{BEL}_i \Box \neg p)) \neg (\mathsf{GOAL}_i(\mathsf{LATER} p))]$$

$$(\mathsf{INTEND}_i a) := (\mathsf{PGOAL}_i[\mathsf{DONE}_i(\mathsf{BEL}_i(\mathsf{HAPPENS} a))?; a])$$

Methodological Issues

A third alternative:

3. Start from an explicit description of *what is being modeled*.

Methodological Issues

A third alternative:

3. Start from an explicit description of *what is being modeled*.

Database/Planner Picture: Planner using a database to maintain its current set of *beliefs*.

Planning vs. Database Management

1. How does an agent *generate* new intentions?

2. Given that the agent's intentions specify a *partial plan*, how and when is the plan "filled out"?

3. How does an agent choose a particular *action* (that is under its control) given its current intentions?

4. How should an agent *maintain* its current state of beliefs and intentions in the presence of new information or new intentions?

5. When should an agent *reconsider* its intentions?

Our Framework

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

Elements of a Logic of Intention Revision

Elements of a Logic of Intention Revision

- Beliefs in a dynamic environment: certainty (irrevocable knowledge, hard information), belief (revisable, soft information), *safe* belief

Elements of a Logic of Intention Revision

- Beliefs in a dynamic environment: certainty (irrevocable knowledge, hard information), belief (revisable, soft information), *safe* belief

- Three views of actions: PDL (state changing), Temporal (lay out time and actions are sequences of time points), STIT (choices, or actions, constrain the future).

Elements of a Logic of Intention Revision

- Beliefs in a dynamic environment: certainty (irrevocable knowledge, hard information), belief (revisable, soft information), *safe* belief

- Three views of actions: PDL (state changing), Temporal (lay out time and actions are sequences of time points), STIT (choices, or actions, constrain the future).

- Two types of beliefs: those about the state of the world and those about the future *which are governed by the agent's plans*

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 ($[\Omega]\varphi$)

time for some details.

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).

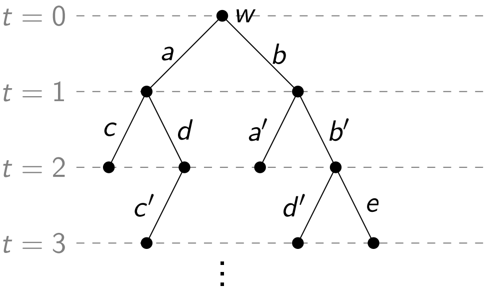
2. *At a fixed moment*, a **model** describes the agent's (current) beliefs (about the current state-of-the-world and what will become true in the future including options that will become available) and the agent's (current) *instructions from the Planner* (about future choices).

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

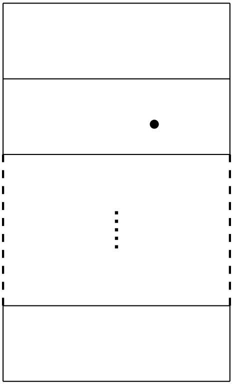
Standard picture where worlds are choice situations

$\mathcal{M}_w \preceq \mathcal{N}_v$: 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

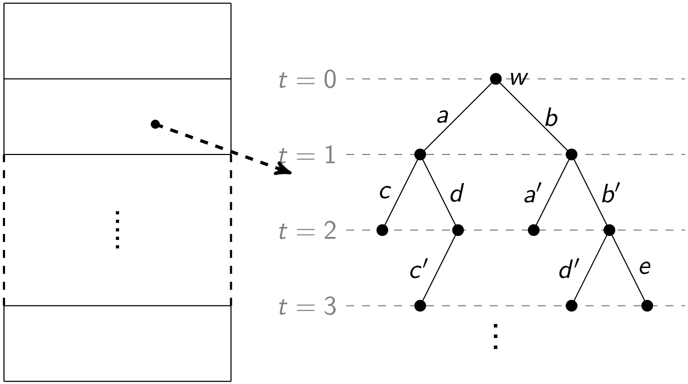
Belief Structures

$$\mathcal{B} = (S, \preceq, \mathcal{M}_w)$$



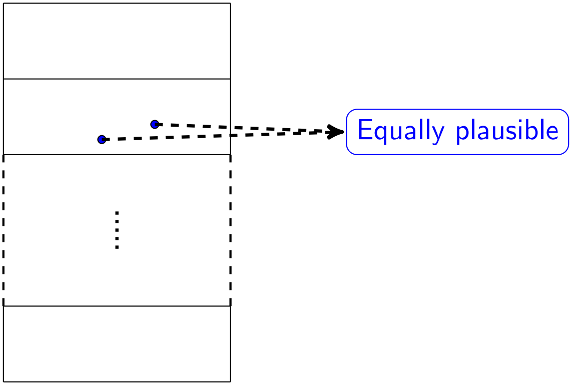
Belief Structures

$$\mathcal{B} = (S, \preceq, \mathcal{M}_w)$$



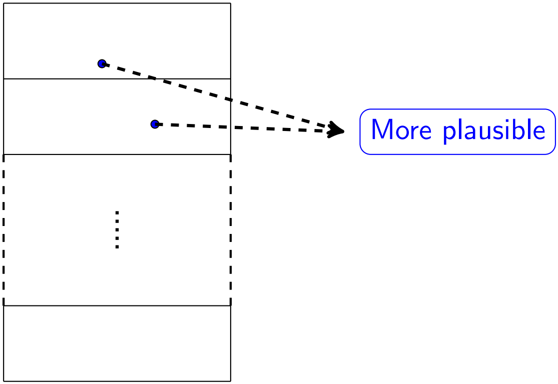
Belief Structures

$$\mathcal{B} = (S, \preceq, \mathcal{M}_w)$$



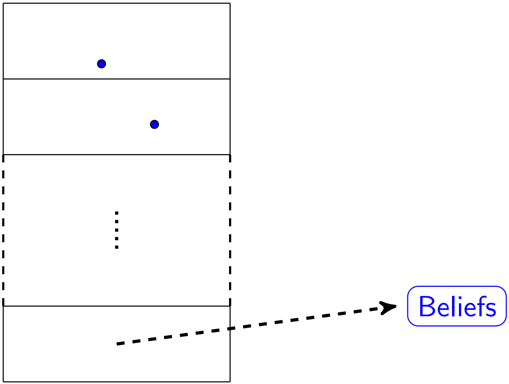
Belief Structures

$$\mathcal{B} = (S, \preceq, \mathcal{M}_w)$$



Belief Structures

$$\mathcal{B} = (S, \preceq, \mathcal{M}_w)$$



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

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

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

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

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

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

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 $a \in \mathrm{Act}$ 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 $a \in \mathsf{Act}$ the agent will perform.
2. The instructions may be *partial*: finite list of pairs (a, t) where $a \in \mathsf{Act}$ 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 a 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.

Consistency of Intentions and Beliefs

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

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).

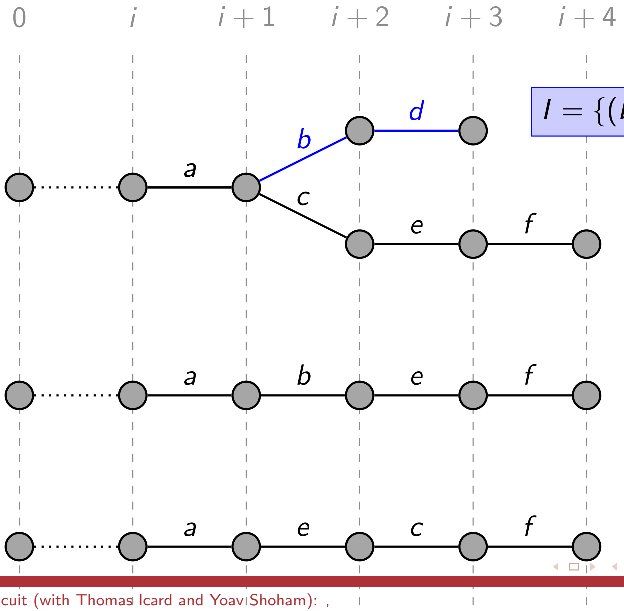
3. The Planner can amend the agent's current set of instructions.

Dynamics

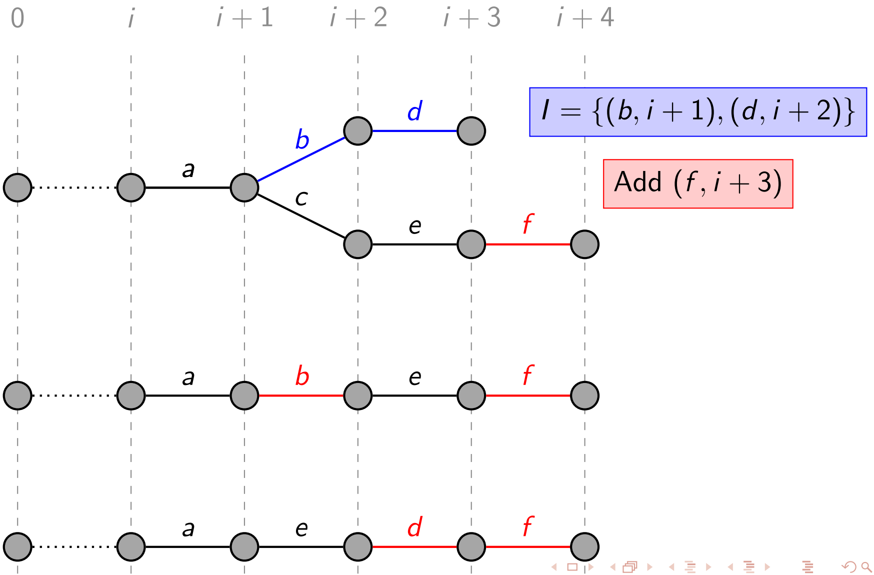
There are three sources of dynamics:

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 Planner 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
$I = \{(b, i+1), (d, i+2)\}$
Add $(f, i+3)$
a
b
e
f
a
e
d
f

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:
$\gamma : \mathcal{P}(\mathit{ChoiceSit}) \times \mathcal{P}_{<\omega}(\mathit{Int}) \to \mathcal{P}_{<\omega}(\mathit{Int})$

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.

Where we are going

Completeness with dynamic operators get considerably more technical (eg., *reduction axioms* are not available), but standard methods work.

Where we are going

Completeness with dynamic operators get considerably more technical (eg., *reduction axioms* are not available), but standard methods work.

Moving to complex plans (with choice, concatenation and test):

1. The notion of Belief-Plan consistency must be updated
2. $\mathcal{I}_{a,t}$ is now defined *semantically*: the agent "intends a, t just in case it is a *necessary component* of the current plan".
3. Axiomatization issues

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).

Logics of Rational Agency

- What's going on in the area:
  www.loriweb.org
- LORI-II, October 8 - 11, 2009, Chongqing, China
  loriweb.org/lori2009
- 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 logic and rational agency
  (eds. J. van Benthem, EP, 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.

Thank You!