Logics of Rational Agency

Lecture 5

Eric Pacuit

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

June 25, 2010

Questions

How should we represent and reason about the underlying *protocol* (or *plans*) that governs the agents' interactions in a particular social situation?

1. What do agents "know" about the protocol, or plans?

2. Which dynamic operations change the protocol or plans over time?

3. Logical issues: language design, axiomatization issues

Two Methodologies

ETL methodology: when describing a social situation, first write down all possible sequences of events, then at each moment write down the agents' uncertainty, from that infer how the agents' knowledge changes from one moment to the next.

Alternative methodology: describe an initial situations, provide a method for how events change a model that can be described in the formal language, then construct the event tree as needed.

Dynamic Epistemic Logic

Returning to the Example: DEL

Returning to the Example: DEL

$$(\mathcal{M} \otimes E_1) \otimes E_2$$

Returning to the Example: DEL

$$(\mathcal{M} \otimes E_1) \otimes E_2$$

The initial model (Ann and Bob are ignorant about P_{2PM}).

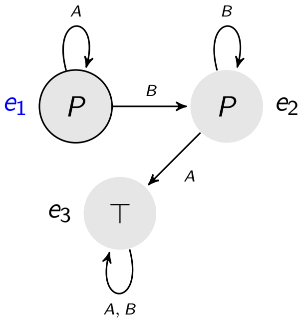
Private announcement to Ann about the talk.

Abstract Description of the Event

Recall the Ann and Bob example: Charles tells Bob that the talk is at 2PM.

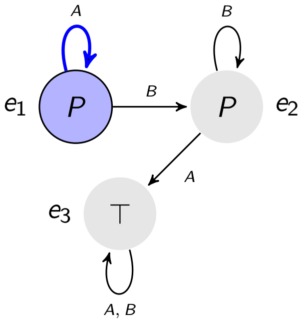
Abstract Description of the Event

Recall the Ann and Bob example: Charles tells Bob that the talk is at 2PM.



Abstract Description of the Event

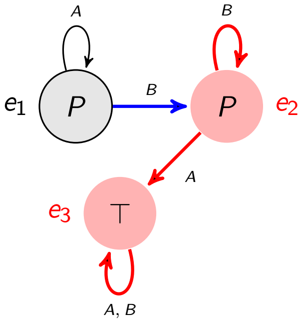
Recall the Ann and Bob example: Charles tells Bob that the talk is at 2PM.



Ann knows which event took place.

Abstract Description of the Event

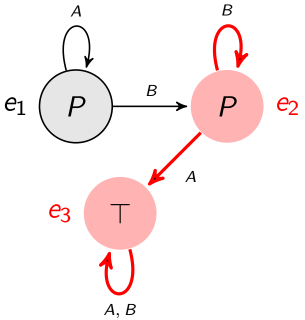
Recall the Ann and Bob example: Charles tells Bob that the talk is at 2PM.



Bob thinks a different event took place.

Abstract Description of the Event

Recall the Ann and Bob example: Charles tells Bob that the talk is at 2PM.

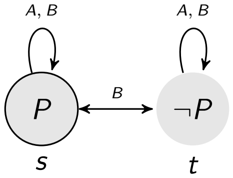


That is, Bob learns the time of the talk, but Ann learns nothing.

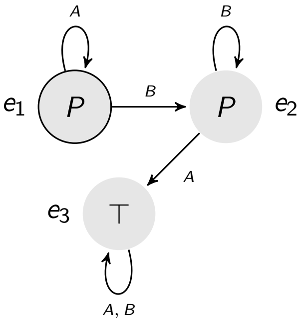
Product Update

Product Update

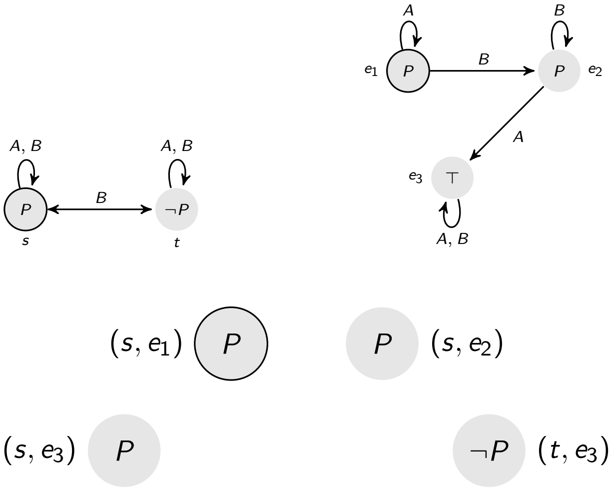
$\mathcal{M} \otimes E_1$



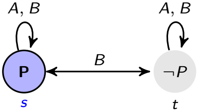
E_2

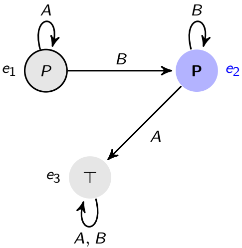


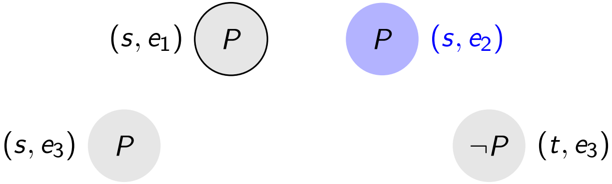
Product Update



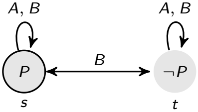
Product Update

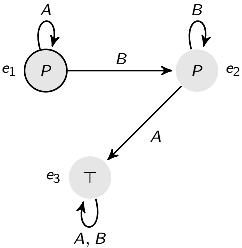




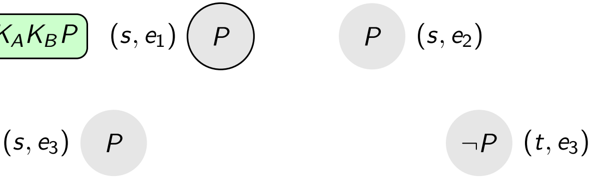


Product Update

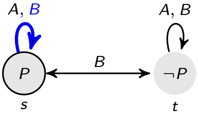


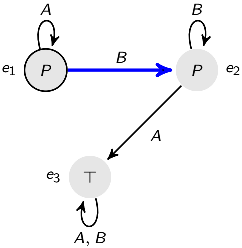


$\boxed{(s, e_1) \models \neg K_B K_A K_B P}$

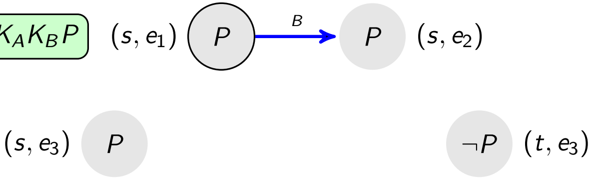


Product Update

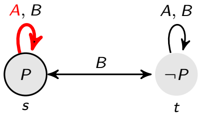


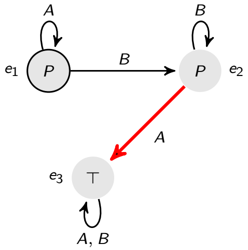


$\boxed{(s, e_1) \models \neg K_B K_A K_B P}$

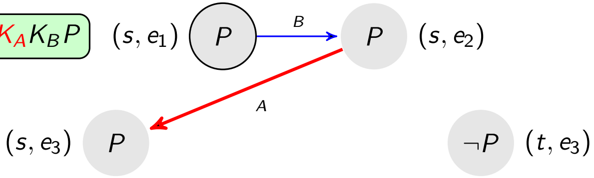


Product Update

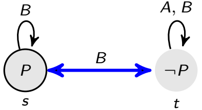


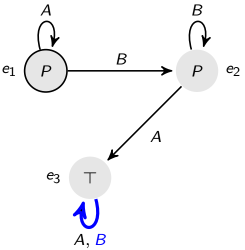


$\boxed{(s, e_1) \models \neg K_B K_A K_B P}$

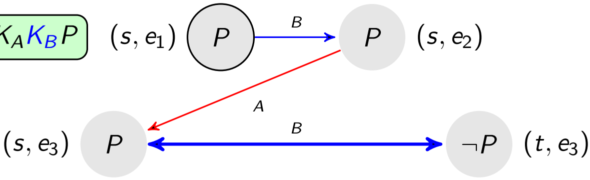


Product Update





$\boxed{(s, e_1) \models \neg K_B K_A K_B P}$



Product Update Details

Let $\mathbb{M} = \langle W, R, V \rangle$ be a Kripke model.

An event model is a tuple $\mathbb{A} = \langle A, S, Pre \rangle$, where $S \subseteq A \times A$ and $Pre : \mathcal{L} \rightarrow \wp(A)$.

Product Update Details

Let $\mathbb{M} = \langle W, R, V \rangle$ be a Kripke model.

An event model is a tuple $\mathbb{A} = \langle A, S, Pre \rangle$, where $S \subseteq A \times A$ and $Pre : \mathcal{L} \to \wp(A)$.

The update model $\mathbb{M} \otimes \mathbb{A} = \langle W', R', V' \rangle$ where

Product Update Details

Let $\mathbb{M} = \langle W, R, V \rangle$ be a Kripke model.

An event model is a tuple $\mathbb{A} = \langle A, S, Pre \rangle$, where $S \subseteq A \times A$ and $Pre : \mathcal{L} \rightarrow \wp(A)$.

The update model $\mathbb{M} \otimes \mathbb{A} = \langle W', R', V' \rangle$ where

- $W' = \{(w, a) \mid w \models Pre(a)\}$

Product Update Details

Let $\mathbb{M} = \langle W, R, V \rangle$ be a Kripke model.

An event model is a tuple $\mathbb{A} = \langle A, S, Pre \rangle$, where $S \subseteq A \times A$ and $Pre : \mathcal{L} \rightarrow \wp(A)$.

The update model $\mathbb{M} \otimes \mathbb{A} = \langle W', R', V' \rangle$ where

- $W' = \{(w, a) \mid w \models Pre(a)\}$
- $(w, a)R'(w', a')$ iff wRw' and aSa'

Product Update Details

Let $\mathbb{M} = \langle W, R, V\rangle$ be a Kripke model.

An event model is a tuple $\mathbb{A} = \langle A, S, Pre\rangle$, where $S \subseteq A \times A$ and $Pre : \mathcal{L} \rightarrow \wp(A)$.

The update model $\mathbb{M} \otimes \mathbb{A} = \langle W', R', V'\rangle$ where

- $W' = \{(w, a) \mid w \models Pre(a)\}$
- $(w, a)R'(w', a')$ iff wRw' and aSa'
- $(w, a) \in V(p)$ iff $w \in V(p)$

Product Update Details

Let $\mathbb{M} = \langle W, R, V \rangle$ be a Kripke model.

An event model is a tuple $\mathbb{A} = \langle A, S, Pre \rangle$, where $S \subseteq A \times A$ and $Pre : \mathcal{L} \rightarrow \wp(A)$.

The update model $\mathbb{M} \otimes \mathbb{A} = \langle W', R', V' \rangle$ where

- $W' = \{(w, a) \mid w \models Pre(a)\}$
- $(w, a)R'(w', a')$ iff wRw' and aSa'
- $(w, a) \in V(p)$ iff $w \in V(p)$

$\mathcal{M}, w \models [\mathrm{A}, a]\varphi$ iff $\mathcal{M}, w \models Pre(a)$ implies $\mathcal{M} \otimes A, (w, a) \models \varphi$.

Literarture

A. Baltag and L. Moss. *Logics for Epistemic Programs*. 2004.

W. van der Hoek, H. van Ditmarsch and B. Kooi. *Dynamic Episetmic Logic*. 2007.

DEL *and* ETL

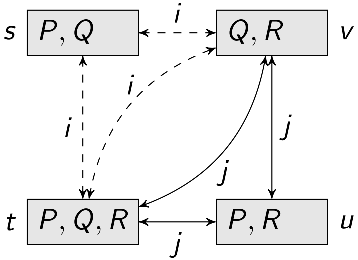
Observation: By repeatedly updating an epistemic model with event models, the machinery of DEL creates ETL models.

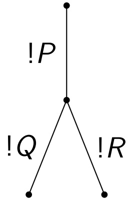
DEL *and* ETL

Observation: By repeatedly updating an epistemic model with event models, the machinery of DEL creates ETL models.

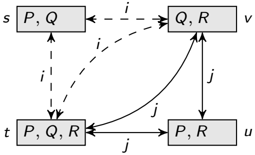
Let M be an epistemic model, and P a DEL protocol (tree of event models). The ETL model generated by M and P, $\mathsf{forest}(M, \mathsf{P})$, represents all possible evolutions of the system obtained by updating M with sequences from P.

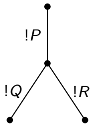
Example: Initial Model and Protocol



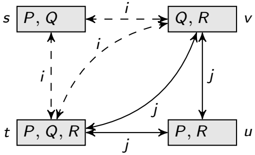


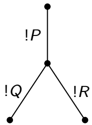
Example

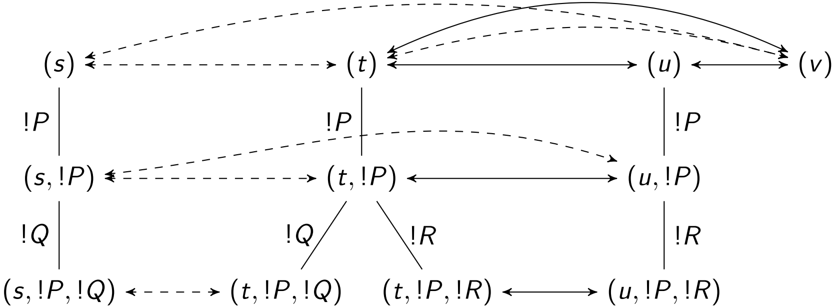




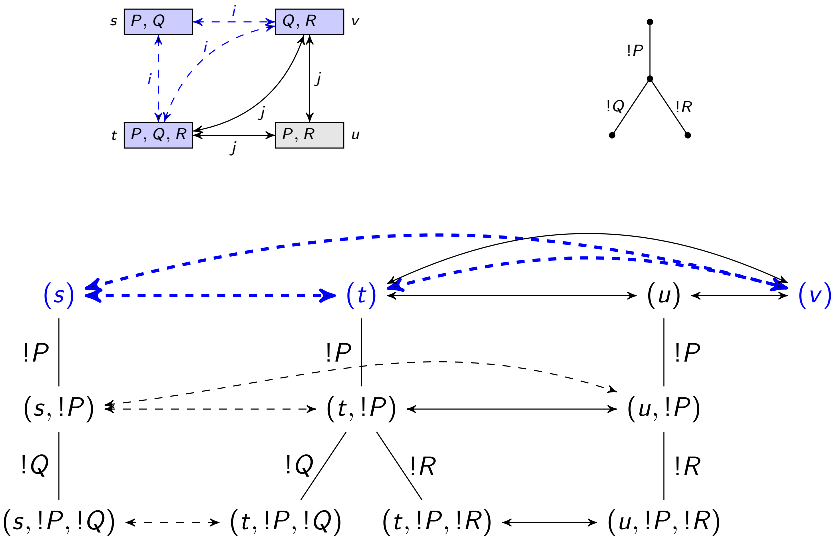
Example


s
P, Q
i
Q, R
v
i
i
j
j
t
P, Q, R
P, R
u
j

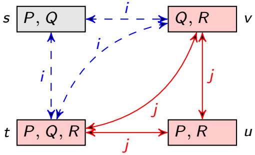

!P
!Q
!R


(s)
(t)
(u)
(v)
!P
!P
!P
(s, !P)
(t, !P)
(u, !P)
!Q
!Q
!R
!R
(s, !P, !Q)
(t, !P, !Q)
(t, !P, !R)
(u, !P, !R)

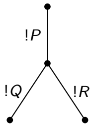
Example


s
P, Q
Q, R
v
i
i
i
j
j
t
P, Q, R
P, R
u
j
!P
!Q
!R
(s)
(t)
(u)
(v)
!P
!P
!P
(s, !P)
(t, !P)
(u, !P)
!Q
!Q
!R
!R
(s, !P, !Q)
(t, !P, !Q)
(t, !P, !R)
(u, !P, !R)

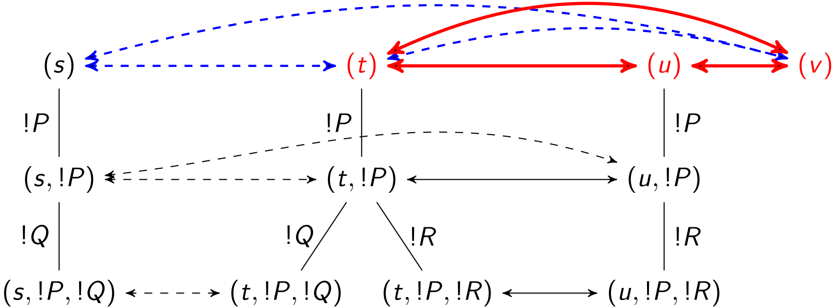
Example



s
P, Q
i
Q, R
v
i
i
j
j
t
P, Q, R
P, R
u
j

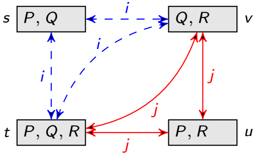


!P
!Q
!R

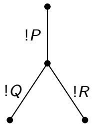


(s)
(t)
(u)
(v)
!P
!P
!P
(s, !P)
(t, !P)
(u, !P)
!Q
!Q
!R
!R
(s, !P, !Q)
(t, !P, !Q)
(t, !P, !R)
(u, !P, !R)

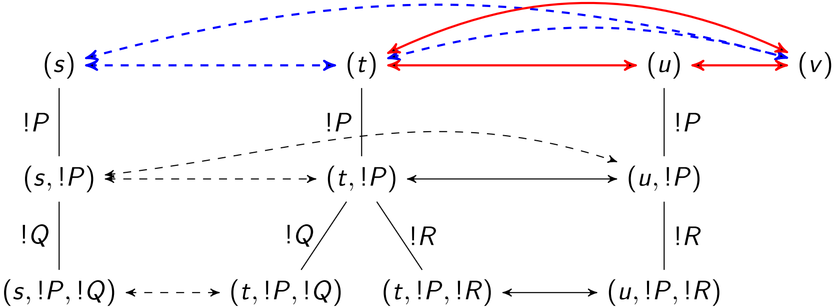
Example



s
P, Q
i
Q, R
v
i
i
j
j
t
P, Q, R
P, R
u
j

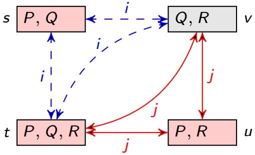


!P
!Q
!R

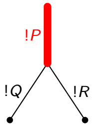


(s)
(t)
(u)
(v)
!P
!P
!P
(s, !P)
(t, !P)
(u, !P)
!Q
!Q
!R
!R
(s, !P, !Q)
(t, !P, !Q)
(t, !P, !R)
(u, !P, !R)

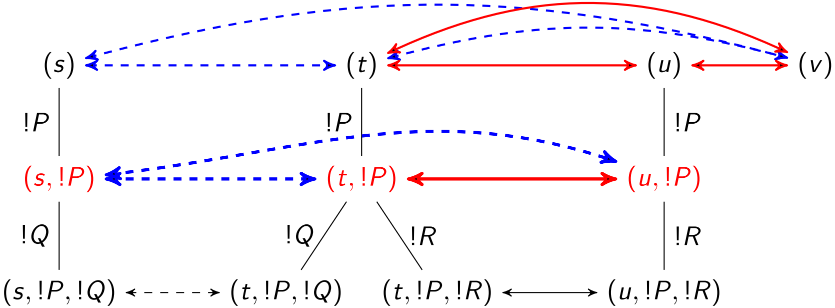
Example



s
P, Q
Q, R
v
i
i
i
j
j
t
P, Q, R
P, R
u
j

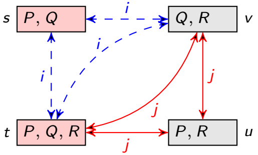


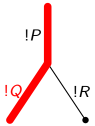
!P
!Q
!R

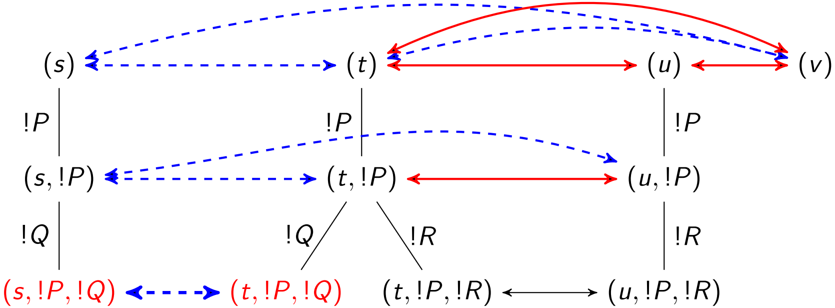


(s)
(t)
(u)
(v)
!P
!P
!P
(s, !P)
(t, !P)
(u, !P)
!Q
!Q
!R
!R
(s, !P, !Q)
(t, !P, !Q)
(t, !P, !R)
(u, !P, !R)

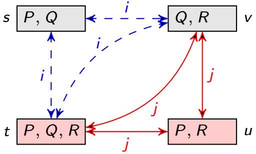
Example

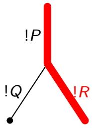


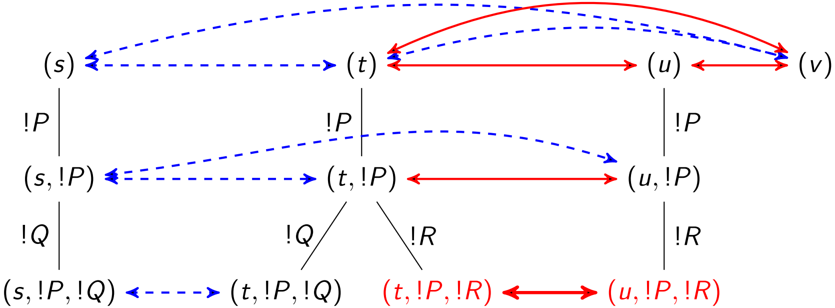




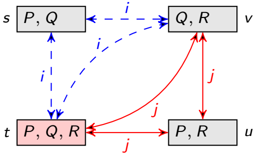
Example

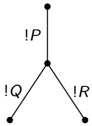

s
P, Q
i
Q, R
v
i
i
j
j
t
P, Q, R
P, R
u
j

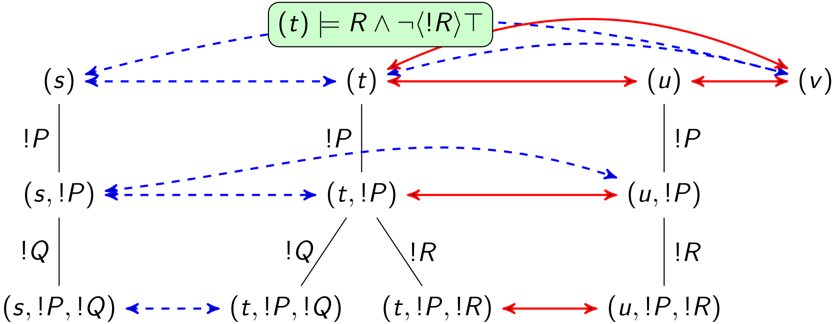

!P
!Q
!R


(s)
(t)
(u)
(v)
!P
!P
!P
(s, !P)
(t, !P)
(u, !P)
!Q
!Q
!R
!R
(s, !P, !Q)
(t, !P, !Q)
(t, !P, !R)
(u, !P, !R)

Example


s
P, Q
i
Q, R
v
i
i
j
j
t
P, Q, R
P, R
u
j


!P
!Q
!R


$(t) \models R \wedge \neg\langle !R \rangle \top$
(s)
(t)
(u)
(v)
$!P$
$!P$
$!P$
$(s, !P)$
$(t, !P)$
$(u, !P)$
$!Q$
$!Q$
$!R$
$!R$
$(s, !P, !Q)$
$(t, !P, !Q)$
$(t, !P, !R)$
$(u, !P, !R)$

$\mathbb{F}(\mathbf{X}) = \{\text{Forest}(\mathcal{M}, \text{P}) \mid \mathcal{M} \text{ an epistemic model and } \text{P} \in \mathbf{X}\}$.

- Can we characterize the class of ETL models $\mathbb{F}(\mathbf{X})$?

- Can we axiomatize interesting classes of DEL-generated ETL models?

J. van Benthem, J. Gerbrandy, T. Hoshi, EP. *Merging Frameworks for Interaction*. JPL, 2009.

A Characterization Theorem

Let Σ be a finite set of events and suppose $\mathbf{X}_{DEL}^{uni}$ is the class of uniform DEL protocols (with a finiteness condition).

Characterization Theorem A model is in $\mathbb{F}(\mathbf{X}_{DEL}^{uni})$ iff it satisfies propositional stability, synchronicity, perfect recall, local no miracles, and local bisimulation invariance.

Constrained Public Announcement

1. $A \rightarrow \langle A \rangle \top$ vs. $\langle A \rangle \top \rightarrow A$

Constrained Public Announcement

1. $A \rightarrow \langle A \rangle \top$ vs. $\langle A \rangle \top \rightarrow A$

2. $\langle A \rangle K_i P \leftrightarrow A \wedge K_i \langle A \rangle P$

Constrained Public Announcement

1. $A \rightarrow \langle A \rangle \top$ vs. $\langle A \rangle \top \rightarrow A$

2. $\langle A \rangle K_i P \leftrightarrow A \wedge K_i \langle A \rangle P$

3. $\langle A \rangle K_i P \leftrightarrow \langle A \rangle \top \wedge K_i(A \rightarrow \langle A \rangle P)$

Constrained Public Announcement

1. $A \rightarrow \langle A \rangle \top$ vs. $\langle A \rangle \top \rightarrow A$
2. $\langle A \rangle K_i P \leftrightarrow A \wedge K_i \langle A \rangle P$
3. $\langle A \rangle K_i P \leftrightarrow \langle A \rangle \top \wedge K_i(A \rightarrow \langle A \rangle P)$
4. $\langle A \rangle K_i P \leftrightarrow \langle A \rangle \top \wedge K_i(\langle A \rangle \top \rightarrow \langle A \rangle P)$

Questions

1. A public announcement is one specific type of event model, can we axiomatize classes of ETL models generated by other types of event models?

2. Which formal languages are best suited to describe these DEL generated ETL models?

$\mathbb{F}(\mathbf{X}) = \{\mathsf{Forest}(\mathcal{M}, \mathsf{P}) \mid \mathcal{M}$ an epistemic model and $\mathsf{P} \in \mathbf{X}\}$.

$\mathbb{F}(\mathbf{X}) = \{\mathsf{Forest}(\mathcal{M}, \mathsf{P}) \mid \mathcal{M}$ an epistemic model and $\mathsf{P} \in \mathbf{X}\}$.

Examples: $\mathbb{F}(\mathsf{PAL})$, $\mathbb{F}(\mathsf{DEL})$, $\mathbb{F}(\mathbf{X}_{PAL})$, $\mathbb{F}(\mathbf{X}_{DEL})$, $\mathbb{F}(\mathbf{X}_{SPriv})$, ...

$\mathbb{F}(\mathbf{X}) = \{\text{Forest}(\mathcal{M}, \text{P}) \mid \mathcal{M} \text{ an epistemic model and } \text{P} \in \mathbf{X}\}$.

Examples: $\mathbb{F}(\text{PAL})$, $\mathbb{F}(\text{DEL})$, $\mathbb{F}(\mathbf{X}_{PAL})$, $\mathbb{F}(\mathbf{X}_{DEL})$, $\mathbb{F}(\mathbf{X}_{SPriv})$, ...

Theorems. Sound and complete axiomatizations (Plaza, Gerbrandy, BMS, van Benthem et al., Hoshi & EP, Balibiani et al., Hoshi, ...)

$\mathbb{F}(\mathbf{X}) = \{\text{Forest}(\mathcal{M}, \text{P}) \mid \mathcal{M} \text{ an epistemic model and } \text{P} \in \mathbf{X}\}$.

Examples: $\mathbb{F}(\text{PAL})$, $\mathbb{F}(\text{DEL})$, $\mathbb{F}(\mathbf{X}_{PAL})$, $\mathbb{F}(\mathbf{X}_{DEL})$, $\mathbb{F}(\mathbf{X}_{SPriv})$, ...

Theorems. Sound and complete axiomatizations (Plaza, Gerbrandy, BMS, van Benthem et al., Hoshi & EP, Balibiani et al., Hoshi, ...)

The logical playground: Decidability of MSO over trees (Rabin); combinations of *PDL* and *S*5; high undecidability just around the corner (Halpern & Vardi, Miller & Moss)

J. van Benthem and EP. *The Tree of Knowledge in Action*. AiML 2006.

Returning to the original questions

How should we represent and reason about the underlying *protocol* (or *plans*) that governs the agents' interactions in a particular social situation?

Returning to the original questions

How should we represent and reason about the underlying *protocol* (or *plans*) that governs the agents' interactions in a particular social situation?

1. What do agents "know" about the protocol, or plans?

Returning to the original questions

How should we represent and reason about the underlying *protocol* (or *plans*) that governs the agents' interactions in a particular social situation?

1. What do agents "know" about the protocol, or plans?

Returning to the original questions

How should we represent and reason about the underlying *protocol* (or *plans*) that governs the agents' interactions in a particular social situation?

1. What do agents "know" about the protocol, or plans?

2. Which dynamic operations change the protocol or plans over time?

Returning to the original questions

How should we represent and reason about the underlying *protocol* (or *plans*) that governs the agents' interactions in a particular social situation?

1. What do agents "know" about the protocol, or plans?

2. Which dynamic operations change the protocol or plans over time?

3. Logical issues: language design, axiomatization issues

What *is* a Protocol?

Given the full tree T of events, a **protocol** is *any* subtree of T.

What *is* a Protocol?

Given the full tree T of events, a **protocol** is *any* subtree of T.

- A protocol is the set of histories compatible with some **process**, i.e., it is the "unwinding" of a (multi-agent) state transition system.

What *is* a Protocol?

Given the full tree T of events, a **protocol** is *any* subtree of T.

- A protocol is the set of histories compatible with some **process**, i.e., it is the "unwinding" of a (multi-agent) state transition system.
- A protocol is the set of histories satisfying some **property**:
  - Physical properties: every message is eventually answered, no message is received before it is sent
  - Agent types: agent i is the **type** of agent who always lies, agent j is the type who always tells the truth

What *is* a Protocol?

Given the full tree T of events, a **protocol** is *any* subtree of T.

- A protocol is the set of histories compatible with some **process**, i.e., it is the "unwinding" of a (multi-agent) state transition system.
- A protocol is the set of histories satisfying some **property**:
  - Physical properties: every message is eventually answered, no message is received before it is sent
  - Agent types: agent i is the **type** of agent who always lies, agent j is the type who always tells the truth
- A protocol is the set of histories of an extensive game consistent with a (partial) **strategy profile.**

Reasoning about protocols, or plans

Reasoning about protocols, or plans

When can an agent *agree* to follow a protocol or plan?

What does the agent need to know about the protocol before agreeing to follow it?

EP and Sunil Simon. *Reasoning with Protocols under Imperfect Information*. manuscript.

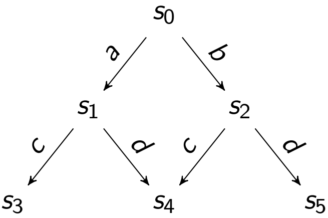
Arena

$$
\begin{tikzcd}
 & & s_0 \arrow[dl, "a"'] \arrow[dr, "b"] & & \\
 & s_1 \arrow[dl, "c"'] \arrow[dr, "d"] & & s_2 \arrow[dl, "c"'] \arrow[dr, "d"] & \\
s_3 & & s_4 & & s_5
\end{tikzcd}
$$

Protocol or Plan

A **protocol** is a finite tree.

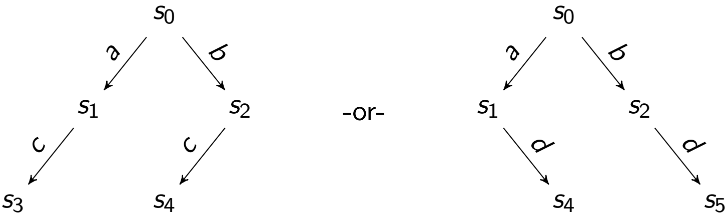
At s_0, the agent agrees to either choose c or choose d:
$(a \cup b); c \cup (a \cup b); d$



Protocol or Plan

A **protocol** is a finite tree.

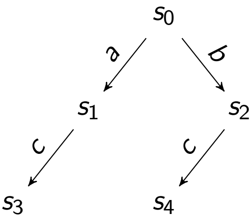
At s_0, the agent agrees to either choose c or choose d:
$(a \cup b); c \cup (a \cup b); d$



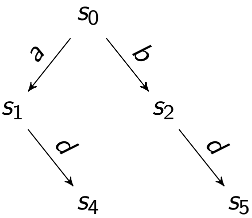
Protocol or Plan

A **protocol** is a finite tree.

At s_0, the agent agrees to either choose c or choose d:
$(a \cup b); c \cup (a \cup b); d$



-or-

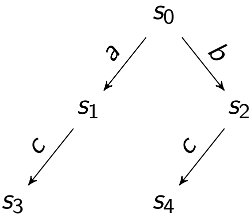


J. van Benthem. *Logical Dynamics of Information and Interaction*. Cambridge University Press, 2010.

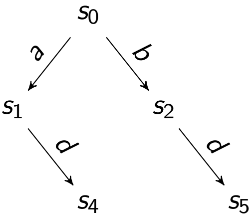
Protocol or Plan

A **protocol** is a finite tree.

At s_0, the agent agrees to either choose c or choose d:
$(a \cup b); c \cup (a \cup b); d$



-or-

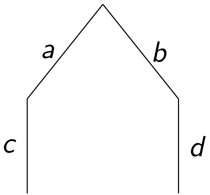


Key idea: of course, PDL action expressions can encode any finite tree, but we want PDL *over trees*

Imperfect Information

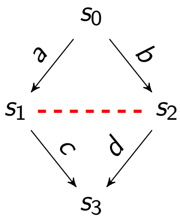
The protocol is **enabled**:

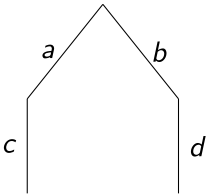
$$\begin{tikzcd} & s_0 \arrow[dl, "a"'] \arrow[dr, "b"] & \\ s_1 \arrow[dr, "c"] & & s_2 \arrow[dl, "d"'] \\ & s_3 & \end{tikzcd}$$



Imperfect Information

The protocol is not **enabled**:





Protocols or Plans

A **protocol** is a finite tree T.

A **complex protocol** is generated by the following grammar:

$$T \mid \pi_1;\pi_2 \mid \pi_1 \cup \pi_2 \mid \pi^*$$

Arena with Imperfect Information

An **arena with imperfect information** is a structure $G^I = (W, \{\Rightarrow_a\}_a \in \Sigma, \rightsquigarrow)$ where $\rightsquigarrow \subseteq W \times W$.

For each position u in an arena, let $\mathcal{I}(u) = \{w \mid u \rightsquigarrow w\}$ be the agent's "*point-of-view*".

For each u, let $\mathcal{A}(u) = \{v \mid \exists a \in \Sigma, w \Rightarrow_a v\}$

Arena with Imperfect Information

- **No Miracles**: for all $a \in \Sigma$ and all $w, v, w', v' \in W$, if $w \rightsquigarrow v$, $w \xrightarrow{a} w'$, and $v \xrightarrow{a} v'$, then $w' \rightsquigarrow v'$.

- **Success**: If $w \rightsquigarrow v$ then $\mathcal{A}(v) \subseteq \mathcal{A}(w)$

- **Awareness**: If $w \rightsquigarrow v$ then $\mathcal{A}(w) \subseteq \mathcal{A}(v)$

- **Certainty of available actions**: If $w \rightsquigarrow v$ and $w \rightsquigarrow v'$ then $\mathcal{A}(v) = \mathcal{A}(v')$

Enabled

A protocol T is **objectively enabled** if T at u in an arena if T can be embedded in the unwinding of u.

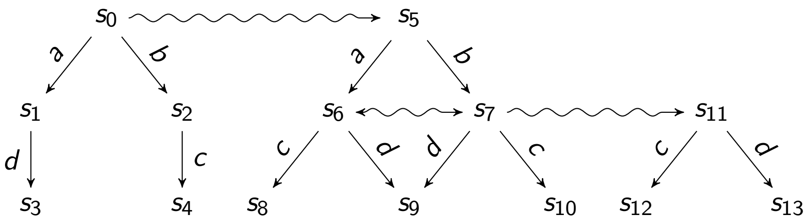
Enabled

A protocol T is **objectively enabled** if T at u in an arena if T can be embedded in the unwinding of u.

A protocol T is **subjectively enabled** at a position u in an arena with imperfect information if

1. the agent is *certain that* T is enabled (for all $v \in \mathcal{I}(u)$, T is enabled at v), and
2. the agent will be certain that she is in fact following the protocol at *every stage* of the protocol.

Subjectively Enabled: The Idea



Adopting a Protocol

- Committing to a basic protocol T *restricts* the choices available to the agent, but there is a trade-off: it also *increases* the ability of the agent to *guarantee* that certain propositions are true.

Adopting a Protocol

- Committing to a basic protocol T *restricts* the choices available to the agent, but there is a trade-off: it also *increases* the ability of the agent to *guarantee* that certain propositions are true.

- Formally, each basic protocol (which is a finite tree) is associated with a set of states X that the agent can "force" the situation to end up in by making choices consistent with the protocol.

Adopting a Protocol

- There are a number of ways to make precise what it means for an agent to "guarantee" that some proposition is true because she adopts the protocols T.

Adopting a Protocol

- There are a number of ways to make precise what it means for an agent to "guarantee" that some proposition is true because she adopts the protocols T.

- Given a complex protocol π, the agent must first decide both *how* to go about adopting π then make her choices"in the moment" consistent with this plan. (Consider committing to $T_1 \cup T_2$).

Reasoning about Abilities under Imperfect Information

Knowledge/belief:

- $\Box\varphi$: "the agent *has the information* that φ is true"

Abilities:

- $\langle\pi\rangle^{\forall}\alpha$: By adopting the protocol π, α is guaranteed to be true.
- $\langle\pi\rangle^{\exists}\alpha$: By adopting the protocol π, the agent can do something consistent with the protocol that will make α true.

"Epistemized" versions of the above operators:

- $\langle\pi\rangle^{\Box}\alpha$: By *agreeing* to adopt the protocol π, the agent is certain that α is guaranteed to be true.
- $\langle\pi\rangle^{\Diamond}\alpha$: By *agreeing* to adopt the protocol π, the agent is can "knowingly" do something consistent with the protocol that will make α true.

Reasoning about Abilities under Imperfect Information

Knowledge/belief:

- $\mathcal{M}, w \models \Box\varphi$ iff for all v, if $w \rightsquigarrow v$ then $\mathcal{M}, v \models \varphi$.

Abilities:

- $M, u \models \langle\pi\rangle^{\exists}\alpha$ iff $\exists(u, X) \in R_{\pi}^{\exists}$, $\exists w \in X$ such that $M, w \models \alpha$.
- $M, u \models \langle\pi\rangle^{\forall}\alpha$ iff $\exists(u, X) \in R_{\pi}^{\forall}$ such that $\forall w \in X$ we have $M, w \models \alpha$.

"Epistemized" versions of the above operators:

- $M, u \models \langle\pi\rangle^{\Box}\alpha$ iff $\exists(u, X) \in R_{\pi}^{\Box}$ such that $\forall w \in X$ we have $M, w \models \alpha$.
- $M, u \models \langle\pi\rangle^{\Diamond}\alpha$ iff $\exists(u, X) \in R_{\pi}^{\Diamond}$ such that $\forall w \in X$ we have $M, w \models \alpha$.

Reasoning about Abilities under Imperfect Information

Knowledge/belief:

- $\mathcal{M}, w \models \Box\varphi$ iff for all v, if $w \rightsquigarrow v$ then $\mathcal{M}, v \models \varphi$.

Abilities:

- $M, u \models \langle\pi\rangle^{\exists}\alpha$ iff $\exists(u, X) \in R_{\pi}^{\exists}$, $\exists w \in X$ such that $M, w \models \alpha$.
- $M, u \models \langle\pi\rangle^{\forall}\alpha$ iff $\exists(u, X) \in R_{\pi}^{\forall}$ such that $\forall w \in X$ we have $M, w \models \alpha$.

"Epistemized" versions of the above operators:

- $M, u \models \langle\pi\rangle^{\Box}\alpha$ iff $\exists(u, X) \in R_{\pi}^{\Box}$ such that $\forall w \in X$ we have $M, w \models \alpha$.
- $M, u \models \langle\pi\rangle^{\Diamond}\alpha$ iff $\exists(u, X) \in R_{\pi}^{\Diamond}$ such that $\forall w \in X$ we have $M, w \models \alpha$.

Reasoning about Abilities under Imperfect Information

Knowledge/belief:

- $\mathcal{M}, w \models \Box\varphi$ iff for all v, if $w \rightsquigarrow v$ then $\mathcal{M}, v \models \varphi$.

Abilities:

- $M, u \models \langle\pi\rangle^{\exists}\alpha$ iff $\exists(u, X) \in R_\pi^{\exists}$, $\exists w \in X$ such that $M, w \models \alpha$.
- $M, u \models \langle\pi\rangle^{\forall}\alpha$ iff $\exists(u, X) \in R_\pi^{\forall}$ such that $\forall w \in X$ we have $M, w \models \alpha$.

"Epistemized" versions of the above operators:

- $M, u \models \langle\pi\rangle^{\Box}\alpha$ iff $\exists(u, X) \in R_\pi^{\Box}$ such that $\forall w \in X$ we have $M, w \models \alpha$.
- $M, u \models \langle\pi\rangle^{\Diamond}\alpha$ iff $\exists(u, X) \in R_\pi^{\Diamond}$ such that $\forall w \in X$ we have $M, w \models \alpha$.

Reasoning about Abilities under Imperfect Information

- $R_t^{\mathfrak{F}} = \{(u, X) \mid \mathit{enabled}(T_t, u)$ and $\mathit{frontier}(T_u \upharpoonright T_t)) = X\}$ (for $\mathfrak{F} \in \{\exists, \forall\}$).

- $R_t^{\Box} = \{(u, X) \mid \mathit{s\text{-}enabled}(T_t, u)$ and $\mathit{frontier}((\mathcal{G}, u) \upharpoonright_s T_t)) = X\}$.

- $R_t^{\Diamond} = \{(u, X) \mid \mathit{s\text{-}enabled}(T_t, u)$ and $\exists \rho \in \mathit{Paths}(T_t)$ with $\mathfrak{S}(\rho, u) = Z_0 Z_1 \ldots Z_k$ and $X = \mathit{last}(Z_k)\}$.

Logics of Information and Abilities

- Combining PDL with epistemic/doxastic logics

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

Logics of Information and Abilities

- Combining PDL with epistemic/doxastic logics

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

- Knowing *how* to execute a plan/win a game

A. Herzig and N. Troquard. *Knowing how to play: uniform choices in logics of agency*. Proceedings of AAMAS 2006, pgs. 209 - 216.

Y. Lesperance, H. Levesque, F. Lin and R. Scherl. *Ability and Knowing How in the Situation Calculus*. Studia Logica 65, pgs. 165 - 186, 2000.

EP and S. Simon. *Reasoning about protocols under imperfect information*. manuscript, 2010.

Merging Logics of Rational Agency

- ✓ Entangling Knowledge/Beliefs and Preferences
- ✓ Reasoning with *protocols*
- ✓ "Epistemizing" Logics of Action and Ability: *knowing how to achieve* φ vs. *knowing that you can achieve* φ
- Plans Change (dynamic BDI - belief desire intention - logics)

The Revision Problem

$$State * Input = State'$$

The Revision Problem

$$\textcolor{red}{State} * Input = \textcolor{red}{State'}$$

1. How should we *describe* the "mental state"?
2.
3.
4.

The Revision Problem

$$\color{red}{State} * Input = \color{red}{State'}$$

1. How should we *describe* the “mental state”?

2. What is a *coherent* description?

3.

4.

The Revision Problem

$$State * Input = State'$$

1. How should we *describe* the "mental state"?

2. What is a *coherent* description?

3. What are the inputs?

4.

The Revision Problem

$$State * Input = State'$$

1. How should we *describe* the "mental state"?

2. What is a *coherent* description?

3. What are the inputs?

4. How should we characterize the revision operator?

The Revision Problem

$$\mathcal{B} * \varphi = \mathcal{B}'$$

1. How should we *describe* the "mental state"?
   Propositional formulas
2. What is a *coherent* description?
   A belief state is a consistent and deductively closed set of propositional formulas
3. What are the inputs?
   Consistent propositional formula, other belief sets
4. How should we characterize the revision operator?
   AGM postulates, Dynamic Doxastic Logic

The Revision Problem

$$State * Input = State'$$

What about other aspects of rational agency?

1. How should we *describe* the "mental state"?
   Propositional formulas
2. What is a *coherent* description?
   A belief state is a consistent and deductively closed set of propositional formulas
3. What are the inputs?
   Consistent propositional formula, other belief sets
4. How should we characterize the revision operator?
   AGM postulates, Dynamic Doxastic Logic

Revising Mental Attitudes

✓ Preference change

T. Grüne-Yanoff and S. Ove Hansen (eds.). *Preference Change*. Vol. 42, Theory and Decision Library (2009).

Revising Mental Attitudes

✓ Preference change

T. Grüne-Yanoff and S. Ove Hansen (eds.). *Preference Change*. Vol. 42, Theory and Decision Library (2009).

✓ General model of "attitude" change

C. List and F. Dietrich. *The aggregation of propositional attitudes: towards a general theory*. Oxford Studies in Epistemology 3 (forthcoming).

Revising Mental Attitudes

✓ Preference change

T. Grüne-Yanoff and S. Ove Hansen (eds.). *Preference Change*. Vol. 42, Theory and Decision Library (2009).

✓ General model of "attitude" change

C. List and F. Dietrich. *The aggregation of propositional attitudes: towards a general theory*. Oxford Studies in Epistemology 3 (forthcoming).

✓ Goal dynamics

C. Castelgranchi and F. Paglieri. *The role of beliefs in goal dynamics: Prolegomena to a constructive theory of intentions*. Synthese 155: 237 - 263 (2007).

Revising Mental Attitudes

✓ Preference change

T. Grüne-Yanoff and S. Ove Hansen (eds.). *Preference Change*. Vol. 42, Theory and Decision Library (2009).

✓ General model of "attitude" change

C. List and F. Dietrich. *The aggregation of propositional attitudes: towards a general theory*. Oxford Studies in Epistemology 3 (forthcoming).

✓ Goal dynamics

C. Castelgranchi and F. Paglieri. *The role of beliefs in goal dynamics: Prolegomena to a constructive theory of intentions*. Synthese 155: 237 - 263 (2007).

Today: Intention dynamics

Conceptual Background: Intentions

Important distinctions:

1. (Present-directed) The intention with which someone acts
2. (Present-directed) Intentional action
3. (Future-directed) Intending to do some action

Conceptual Background: Intentions

Important distinctions:

1. (Present-directed) The intention with which someone acts
2. (Present-directed) Intentional action
3. (Future-directed) Intending to do some action

Some issues:

- Unifying account of *intentions*

  "Where we are tempted to speak of 'different senses' of a word which is clearly not equivocal, we may infer that we are pretty much in the dark about the character of the concept which it represents"
  - G.E.M. Anscombe, *Intention*, pg. 1

Conceptual Background: Intentions

Important distinctions:

1. (Present-directed) The intention with which someone acts
2. (Present-directed) Intentional action
3. (Future-directed) Intending to do some action

Some issues:

- Unifying account of *intentions*
- Intention as a *mental state*

  pro-attitude (vs. informational attitude), *world-to-mind* direction of fit, *conduct-controlling*

Conceptual Background: Intentions

Important distinctions:

1. (Present-directed) The intention with which someone acts
2. (Present-directed) Intentional action
3. (Future-directed) Intending to do some action

Some issues:

- Unifying account of *intentions*
- Intention as a *mental state*
- Intentions are (always) directed towards *actions*
  "Although we sometimes report intention as a propositional attitude — 'I intend that *p*' — such reports can always be recast as 'intending to ....' as when I intend to bring about that *p*. By contrast, it is difficult to rephrase such mundane expressions as 'I intend to walk home' in propositional terms"

Conceptual Background: Intentions

Important distinctions:

1. (Present-directed) The intention with which someone acts
2. (Present-directed) Intentional action
3. (Future-directed) Intending to do some action

Some issues:

- Unifying account of *intentions*
- Intention as a *mental state*
- Intentions are (always) directed towards *actions*

An extensive literature:

K. Setiya. *Intention*. Stanford Encyclopedia of Philosophy (2010).

Conceptual Background: Intentions

Important distinctions:

1. (Present-directed) The intention with which someone acts
2. (Present-directed) Intentional action
3. (Future-directed) Intending to do some action

Some issues:

- Unifying account of *intentions*
- Intention as a *mental state*
- Intentions are (always) directed towards *actions*

An extensive literature:

K. Setiya. *Intention*. Stanford Encyclopedia of Philosophy (2010).

Functional Description of Intentions

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

"intention is a distinctive practical attitude marked by its pivotal role in planning for the future.

Functional Description of Intentions

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

"intention is a distinctive practical attitude marked by its pivotal role in planning for the future. Intention involves desire, but even predominant desire is insufficient for intention, since it need not involve a commitment to act:

Functional Description of Intentions

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

"intention is a distinctive practical attitude marked by its pivotal role in planning for the future. Intention involves desire, but even predominant desire is insufficient for intention, since it need not involve a commitment to act: intentions are conduct-controlling pro-attitudes, ones which we are disposed to retain without reconsideration, and which play a significant role as inputs to [means-end] reasoning" (pg. 20)

Functional Description of Intentions

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

Committing to an action in advance is crucial for

Functional Description of Intentions

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

Committing to an action in advance is crucial for

1. our capacity to make rational decisions (as a *bounded agent*)

Functional Description of Intentions

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

Committing to an action in advance is crucial for

1. our capacity to make rational decisions (as a *bounded agent*)
2. our capacity to engage in complex, temporally extended projects

Functional Description of Intentions

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

Committing to an action in advance is crucial for

1. our capacity to make rational decisions (as a *bounded agent*)
2. our capacity to engage in complex, temporally extended projects
3. our capacity to coordinate with others

Functional Description of Intentions

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

Committing to an action in advance is crucial for

1. our capacity to make rational decisions (as a *bounded agent*)

2. our capacity to engage in complex, temporally extended projects

3. our capacity to coordinate with others

Of course, this commitment is *defeasible*...

Stability of Plans

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

plans normally resist reconsideration:

Stability of Plans

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

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*".

Stability of Plans

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

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". .... "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*" (pg. 65)

Stability of Plans

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

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". .... "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*" (pg. 65)

What happens in "abnormal" or "surprising" situations? This points to a theory of (rational) *intention/plan revision*...

Conceptual Issue: intentions and beliefs are *entangled*

Conceptual Issue: intentions and beliefs are *entangled*

1. Intending to act just *is* a special kind of belief that one will;

Conceptual Issue: intentions and beliefs are *entangled*

1. Intending to act just *is* a special kind of belief that one will;

2. Intending to act *involves* a belief that one will so act;

Conceptual Issue: intentions and beliefs are *entangled*

1. Intending to act just *is* a special kind of belief that one will;
2. Intending to act *involves* a belief that one will so act;
3. Intending to act involves a belief that it is *possible* that one will so act.

Conceptual Issue: intentions and beliefs are *entangled*

1. Intending to act just *is* a special kind of belief that one will;
2. Intending to act *involves* a belief that one will so act;
3. Intending to act involves a belief that it is *possible* that one will so act.

Conceptual Issue: rationality constraints on intentions

Conceptual Issue: rationality constraints on intentions

1. *Consistency*: “one’s intentions, taken together with one’s beliefs fit together into a consistent model of one’s future”

Conceptual Issue: rationality constraints on intentions

1. *Consistency*: "one's intentions, taken together with one's beliefs fit together into a consistent model of one's future"

2. *Means-ends consistency*: "it is irrational that one intends E, believes that E requires that one intend means M and yet not intend M"

Conceptual Issue: rationality constraints on intentions

1. *Consistency*: "one's intentions, taken together with one's beliefs fit together into a consistent model of one's future"

2. *Means-ends consistency*: "it is irrational that one intends E, believes that E requires that one intend means M and yet not intend M"

3. *Agglomeration*: "Intending A and Intending B implies Intending (A and B)"

M. Bratman. *Intention, Belief, Practical, Theoretical*. in *Spheres of Reason* (2009).

Intention Dynamics

Intention Dynamics

E. Lorini and A. Herzig. *A logic of intention and attempt*. Synthese 163, pp. 45 - 77 (2008).

Intention Dynamics

1. **Intentional Action Execution**: precise characterization under which an agent's intention *transforms* into an action. (trying, attempting)

E. Lorini and A. Herzig. *A logic of intention and attempt*. Synthese 163, pp. 45 - 77 (2008).

Intention Dynamics

1. **Intentional Action Execution**: precise characterization under which an agent's intention *transforms* into an action. (trying, attempting)

2. **Intention Generation**: model appropriate principles of intention generation (practical or instrumental reasoning)

E. Lorini and A. Herzig. *A logic of intention and attempt*. Synthese 163, pp. 45 - 77 (2008).

Intention Dynamics

1. **Intentional Action Execution**: precise characterization under which an agent's intention *transforms* into an action. (trying, attempting)

2. **Intention Generation**: model appropriate principles of intention generation (practical or instrumental reasoning)

3. **Intention Persistence**: intentions normally *resist* reconsideration (bounded agents)

E. Lorini and A. Herzig. *A logic of intention and attempt*. Synthese 163, pp. 45 - 77 (2008).

Belief *and* Intention Revision

Belief *and* Intention Revision

1. Strong analogy between the problem of plan/intention change and belief change.

Belief *and* Intention Revision

1. Strong analogy between the problem of plan/intention change and belief change. *When an agent learns that his plans are no longer viable, he must change them so that they account for the new information.*

Belief *and* Intention Revision

1. Strong analogy between the problem of plan/intention change and belief change. *When an agent learns that his plans are no longer viable, he must change them so that they account for the new information.*

2. There are important differences between the two problems:

Belief *and* Intention Revision

1. Strong analogy between the problem of plan/intention change and belief change. *When an agent learns that his plans are no longer viable, he must change them so that they account for the new information.*

2. There are important differences between the two problems: *If the agent has committed to doing action a at time i (perhaps on condition φ), there is always a specific reason for this commitment. In contrast, standard models of belief change do not record specific reasons for each belief.*

Belief *and* Intention Revision

1. Strong analogy between the problem of plan/intention change and belief change. *When an agent learns that his plans are no longer viable, he must change them so that they account for the new information.*

2. There are important differences between the two problems: *If the agent has committed to doing action a at time i (perhaps on condition φ), there is always a specific reason for this commitment. In contrast, standard models of belief change do not record specific reasons for each belief.*

3. Finally, the two problems are *intertwined*:

Belief *and* Intention Revision

1. Strong analogy between the problem of plan/intention change and belief change. *When an agent learns that his plans are no longer viable, he must change them so that they account for the new information.*

2. There are important differences between the two problems: *If the agent has committed to doing action a at time i (perhaps on condition φ), there is always a specific reason for this commitment. In contrast, standard models of belief change do not record specific reasons for each belief.*

3. Finally, the two problems are *intertwined*: *This is because we must not only maintain that the agent's beliefs and plans are individually consistent, but also that they are jointly coherent.*

Belief *and* Intention Revision

1. How should we *describe* the "mental state"?

2. What is a *coherent* description?

3. What are the inputs?

4. How should we characterize the revision operator?

Belief *and* Intention Revision

1. How should we *describe* the "mental state"?

2. What is a *coherent* description?

3. What are the inputs?

4. How should we characterize the revision operator?

Goal: Find a logic where *consistency* represents

Goal: Find a logic where *consistency* represents

1. the agent's beliefs are consistent,

Goal: Find a logic where *consistency* represents

1. the agent's beliefs are consistent,
2. the agent's future plans are consistent and

Goal: Find a logic where *consistency* represents

1. the agent's beliefs are consistent,
2. the agent's future plans are consistent and
3. the beliefs and intentions together form a coherent picture of what may happen (and how the agent's own actions will play a role in what may happen).

Background: BDI Logics

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

- Cohen and Levesque ▸ Example; Rao and Georgeff (BDI); Meyer and van der Hoek (KARO); and many others.

Background: BDI Logics

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

- Cohen and Levesque ▸ Example; Rao and Georgeff (BDI); Meyer and van der Hoek (KARO); 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.

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 (eg., the "side-effect problem").

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 (eg., the "side-effect problem").

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 (eg., the "side-effect problem").

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.

Describing the Mental State

W. van der Hoek, W. Jamroga and M. Wooldridge. *Towards a theory of intention revision*. Synthese 155, pgs. 265 - 290 (2007).

Describing the Mental State

W. van der Hoek, W. Jamroga and M. Wooldridge. *Towards a theory of intention revision*. Synthese 155, pgs. 265 - 290 (2007).

$$\Gamma = \langle \mathcal{B}, \mathcal{D}, \langle \mathcal{I}, \mathcal{A} \rangle, \P \rangle$$

1. $\mathcal{B}$: beliefs

Describing the Mental State

W. van der Hoek, W. Jamroga and M. Wooldridge. *Towards a theory of intention revision*. Synthese 155, pgs. 265 - 290 (2007).

$$\Gamma = \langle \mathcal{B}, \mathcal{D}, \langle \mathcal{I}, \mathcal{A} \rangle, \P \rangle$$

1. $\mathcal{B}$: beliefs
2. $\mathcal{D}$: desires

Describing the Mental State

W. van der Hoek, W. Jamroga and M. Wooldridge. *Towards a theory of intention revision*. Synthese 155, pgs. 265 - 290 (2007).

$$\Gamma = \langle \mathcal{B}, \mathcal{D}, \langle \mathcal{I}, \mathcal{A} \rangle, \P \rangle$$

1. $\mathcal{B}$: beliefs
2. $\mathcal{D}$: desires
3. $\langle \mathcal{I}, \mathcal{A} \rangle$: intentions and active plans

Describing the Mental State

W. van der Hoek, W. Jamroga and M. Wooldridge. *Towards a theory of intention revision*. Synthese 155, pgs. 265 - 290 (2007).

$$\Gamma = \langle \mathcal{B}, \mathcal{D}, \langle \mathcal{I}, \mathcal{A} \rangle, \P \rangle$$

1. $\mathcal{B}$: beliefs
2. $\mathcal{D}$: desires
3. $\langle \mathcal{I}, \mathcal{A} \rangle$: intentions and active plans
4. ¶: practical reasoning rules ("know how")

Describing the Mental State

W. van der Hoek, W. Jamroga and M. Wooldridge. *Towards a theory of intention revision*. Synthese 155, pgs. 265 - 290 (2007).

$$\Gamma = \langle \mathcal{B}, \mathcal{D}, \langle \mathcal{I}, \mathcal{A} \rangle, \mathcal{P} \rangle$$

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

Describing the Mental State

W. van der Hoek, W. Jamroga and M. Wooldridge. *Towards a theory of intention revision*. Synthese 155, pgs. 265 - 290 (2007).

$$\Gamma = \langle \mathcal{B}, \mathcal{D}, \langle \mathcal{I}, \mathcal{A} \rangle, \mathcal{P} \rangle$$

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

Describing the Mental State

W. van der Hoek, W. Jamroga and M. Wooldridge. *Towards a theory of intention revision*. Synthese 155, pgs. 265 - 290 (2007).

$$\Gamma = \langle \mathcal{B}, \mathcal{D}, \langle \mathcal{I}, \mathcal{A} \rangle, \mathcal{P} \rangle$$

- Beliefs are sets of Linear Temporal Logic formulas (eg., $\bigcirc\varphi$)
- Desires are (possibly inconsistent) sets of Linear Temporal Logic formulas
- Practical reasoning rules: $\alpha \leftarrow \alpha_1, \alpha_2, \ldots, \alpha_n$ ("if $T\alpha$ is an intention then $T(\alpha_1 \wedge \cdots \wedge \alpha_n)$ is a possible fulfillment...")

Describing the Mental State

W. van der Hoek, W. Jamroga and M. Wooldridge. *Towards a theory of intention revision*. Synthese 155, pgs. 265 - 290 (2007).

$$\Gamma = \langle \mathcal{B}, \mathcal{D}, \langle \mathcal{I}, \mathcal{A} \rangle, \mathcal{P} \rangle$$

- Beliefs are sets of Linear Temporal Logic formulas (eg., $\bigcirc\varphi$)
- Desires are (possibly inconsistent) sets of Linear Temporal Logic formulas
- Practical reasoning rules: $\alpha \leftarrow \alpha_1, \alpha_2, \ldots, \alpha_n$ ("if $T\alpha$ is an intention then $T(\alpha_1 \wedge \cdots \wedge \alpha_n)$ is a possible fulfillment...")
- Intentions are derived from the agents current active plans (trees of practical reasoning rules)

What is the appropriate underlying logic?

What is the appropriate underlying logic?

Many proposals, but no clear consensus...

What is the appropriate underlying logic?

Many proposals, but no clear consensus...

- **KD45** for B?

What is the appropriate underlying logic?

Many proposals, but no clear consensus...

- **KD45** for B?
- $B\varphi \rightarrow Goal\varphi$?

What is the appropriate underlying logic?

Many proposals, but no clear consensus...

- **KD45** for B?
- $B\varphi \rightarrow Goal\varphi$?
- $Goal\varphi \rightarrow \neg B \neg \varphi$?

What is the appropriate underlying logic?

Many proposals, but no clear consensus...

- **KD45** for B?
- $B\varphi \rightarrow Goal\varphi$?
- $Goal\varphi \rightarrow \neg B \neg \varphi$?
- $Goal\varphi \rightarrow BGoal\varphi$?

What is the appropriate underlying logic?

Many proposals, but no clear consensus...

- **KD45** for B?
- $B\varphi \rightarrow Goal\varphi$?
- $Goal\varphi \rightarrow \neg B \neg \varphi$?
- $Goal\varphi \rightarrow BGoal\varphi$?
- Temporal logic, action logic, doxastic logic, combinations, etc., etc.

Focusing the Discussion

Start from an explicit description of *what is being modeled*: eg., a "planner" using a "database" to maintain its current set of beliefs and plans.

Y. Shoham. *Logic of Intention and the Database Perspective*. JPL 2009.

Focusing the Discussion

Start from an explicit description of *what is being modeled*: eg., a "planner" using a "database" to maintain its current set of beliefs and plans.

Y. Shoham. *Logic of Intention and the Database Perspective*. JPL 2009.

How should we *describe* the "mental state"?
What type of entries are in the database?

Focusing the Discussion

Start from an explicit description of *what is being modeled*: eg., a "planner" using a "database" to maintain its current set of beliefs and plans.

Y. Shoham. *Logic of Intention and the Database Perspective*. JPL 2009.

How should we *describe* the "mental state"?
What type of entries are in the database?

1. Beliefs (about future states, which actions are available plus what the agent might *do*)
2. Current instructions from the planner

Contingent vs. Non-contingent Beliefs

Post-conditions of *intended actions* are justifiably believed *by the mere fact that the agent has committed to bringing them about.*

Contingent vs. Non-contingent Beliefs

Post-conditions of *intended actions* are justifiably believed *by the mere fact that the agent has committed to bringing them about.*

On the other hand, *pre-conditions* may still pose a practical problem yet to be solved.

Contingent vs. Non-contingent Beliefs

"My belief that I will be at Tanner Library this afternoon is based on my knowledge that I intend to go there.

Contingent vs. Non-contingent Beliefs

"My belief that I will be at Tanner Library this afternoon is based on my knowledge that I intend to go there. If I reconsider this intention, I must bracket the support it provides for this belief and others. I must take care not to keep assuming I will be at Tanner, even while reconsidering my intention to go there....

Contingent vs. Non-contingent Beliefs

"My belief that I will be at Tanner Library this afternoon is based on my knowledge that I intend to go there. If I reconsider this intention, I must bracket the support it provides for this belief and others. I must take care not to keep assuming I will be at Tanner, even while reconsidering my intention to go there....Keeping track of the ways in which one's beliefs depend on intentions being reconsidered may become a fairly complex matter, especially as one reconsiders more extensive elements in one's prior plans.

Contingent vs. Non-contingent Beliefs

"My belief that I will be at Tanner Library this afternoon is based on my knowledge that I intend to go there. If I reconsider this intention, I must bracket the support it provides for this belief and others. I must take care not to keep assuming I will be at Tanner, even while reconsidering my intention to go there....Keeping track of the ways in which one's beliefs depend on intentions being reconsidered may become a fairly complex matter, especially as one reconsiders more extensive elements in one's prior plans. *But this should not be taken to show that one may rationally proceed without adjusting one's beliefs as one reconsiders*.

Contingent vs. Non-contingent Beliefs

"My belief that I will be at Tanner Library this afternoon is based on my knowledge that I intend to go there. If I reconsider this intention, I must bracket the support it provides for this belief and others. I must take care not to keep assuming I will be at Tanner, even while reconsidering my intention to go there....Keeping track of the ways in which one's beliefs depend on intentions being reconsidered may become a fairly complex matter, especially as one reconsiders more extensive elements in one's prior plans. *But this should not be taken to show that one may rationally proceed without adjusting one's beliefs as one reconsiders*. Rather, it shows just how complicated — and so, costly — reconsideration of prior intentions can be."

[Bratman, pg. 63, my emphasis]

Belief-Intention States: Beliefs

$$\varphi \quad := \quad p_t \mid pre(a)_t \mid post(a)_t \mid Do(a)_t \mid \Box\varphi \mid \varphi \wedge \varphi \mid \neg\varphi$$

$p \in \text{At}$, $a \in \text{Act}$, $t \in \mathbb{Z}^+$

- p_t means p is true at time t
- For every a and time t associate $pre(a)_t$, $post(a)_{t+1}$, which we treat as distinguished propositional variables
- $Do(a)_t$ mean the agent does a at t units from now
- $\Box$ is historic necessity

Belief-Intention States: Beliefs

Beliefs are **consistent and deductively closed** sets of formulas:

Belief-Intention States: Beliefs

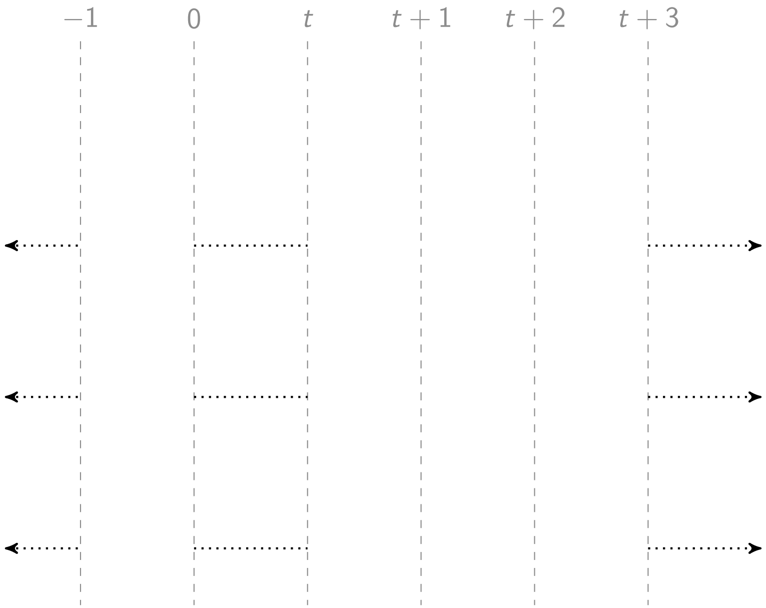
Beliefs are **consistent and deductively closed** sets of formulas:

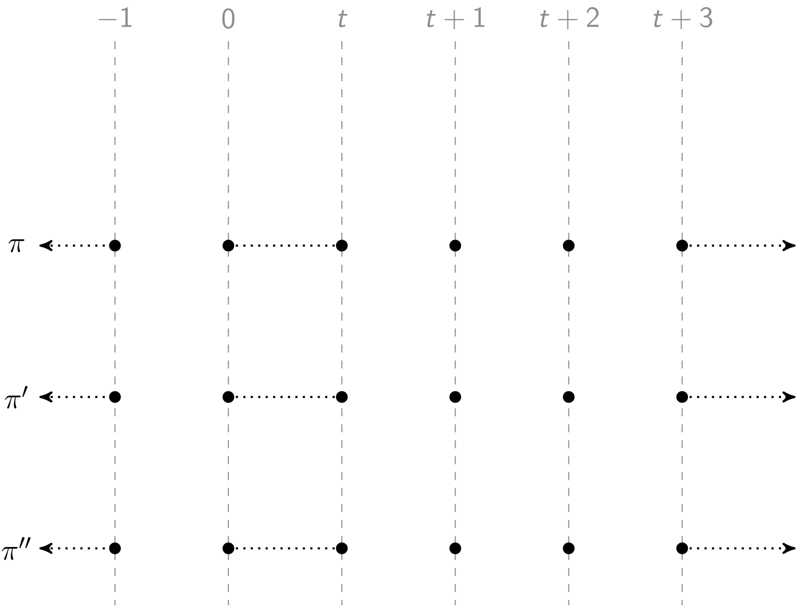
1. Propositional Tautologies;
2. **S5** axiom schemes for $\Box$: $(\Box(\varphi \to \psi) \to \Box\varphi \to \Box\psi$, $\Box\varphi \to \varphi$, $\Box\varphi \to \Box\Box\varphi$,$\Diamond\varphi \to \Box\Diamond\varphi$, Necessitation);
3. $\bigvee_{a \in \mathsf{Act}} Do(a)_t$;
4. $Do(a)_t \to \bigwedge_{b \neq a} \neg Do(b)_t$;
5. $Do(a)_t \to post(a)_{t+1}$;
6. $pre(a)_t \to \Diamond Do(a)_t$;
7. Modus Ponens

Belief-Intention States: Semantics

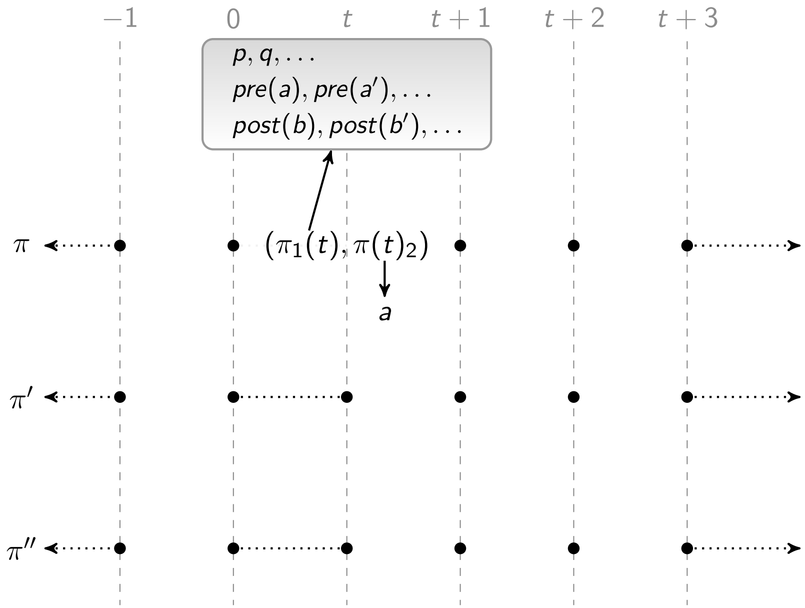
$P = \mathcal{P}(\mathsf{Prop} \cup \{pre(a), post(a) \mid a \in \mathsf{Act}\})$

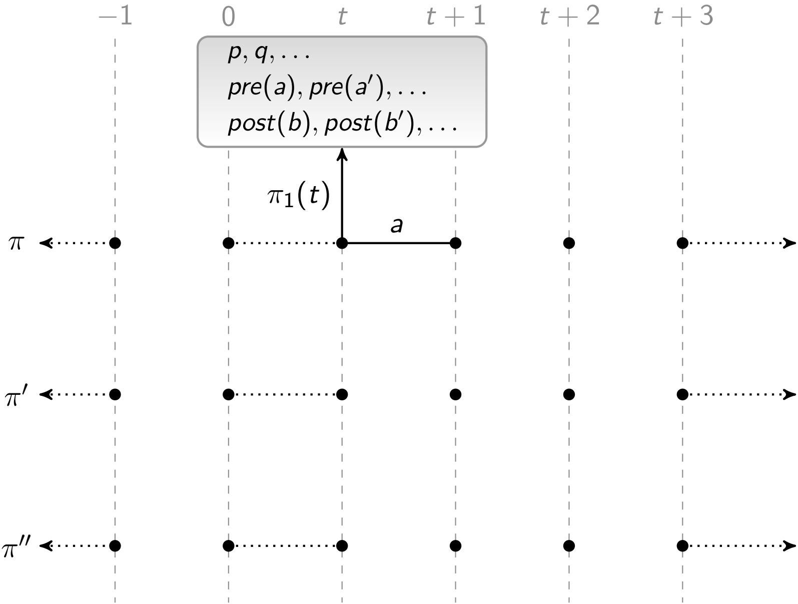
Path: $\pi : \mathbb{Z} \rightarrow (P \times \mathsf{Act})$


-1
0
t
$t+1$
$t+2$
$t+3$



-1
0
t
$t+1$
$t+2$
$t+3$
π
π'
π''


−1
0
t
t + 1
t + 2
t + 3
p, q, . . .
pre(a), pre(a′), . . .
post(b), post(b′), . . .
π
(π1(t), π(t)2)
a
π′
π′′


−1
0
t
t + 1
t + 2
t + 3
p, q, . . .
pre(a), pre(a′), . . .
post(b), post(b′), . . .
π1(t)
a
π
π′
π′′

Belief-Intention States: Semantics

$P = \mathcal{P}(\mathsf{Prop} \cup \{pre(a), post(a) \mid a \in \mathsf{Act}\})$

Path: $\pi : \mathbb{Z} \to (P \times \mathsf{Act})$

Belief-Intention States: Semantics

$P = \mathcal{P}(\mathsf{Prop} \cup \{pre(a), post(a) \mid a \in \mathsf{Act}\})$

Path: $\pi : \mathbb{Z} \rightarrow (P \times \mathsf{Act})$

Historic necessity: $\pi \sim_t \pi'$ iff for all $t' \leq t$, $\pi(t') = \pi'(t')$

Belief-Intention States: Semantics

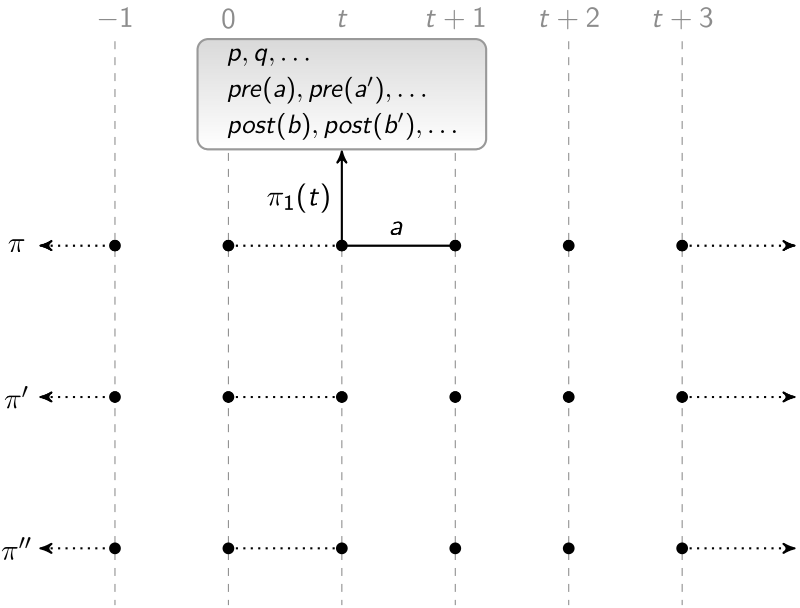
$P = \mathcal{P}(\text{Prop} \cup \{\mathit{pre}(a), \mathit{post}(a) \mid a \in \text{Act}\})$

Path: $\pi : \mathbb{Z} \to (P \times \text{Act})$

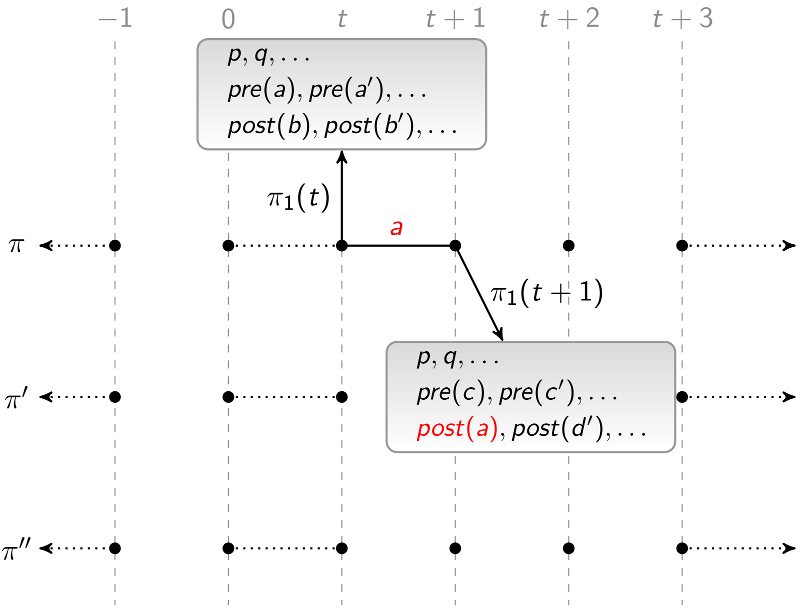
Historic necessity: $\pi \sim_t \pi'$ iff for all $t' \le t$, $\pi(t') = \pi'(t')$

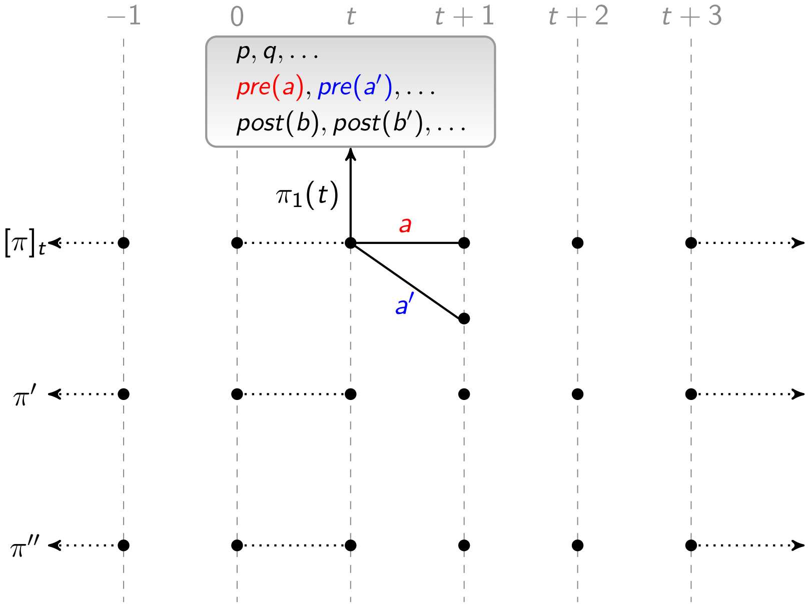
Appropriate path:

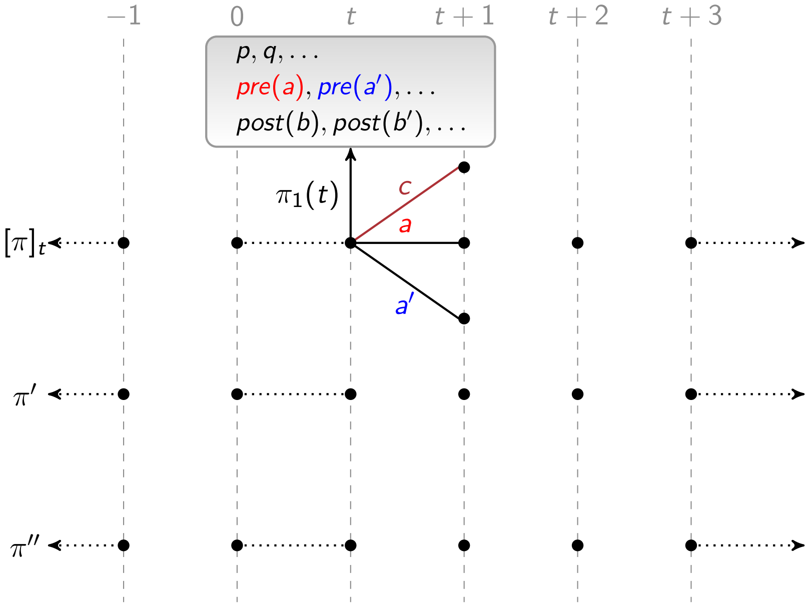
- If $\pi(t)_2 = a$ then $\mathit{post}(a) \in \pi(t+1)_1$
- If $\mathit{pre}(a) \in \pi(t)_1$ then there is some $\pi' \sim_t \pi$ such that $\pi'(t)_2 = a$.



−1
0
t
t + 1
t + 2
t + 3
p, q, . . .
pre(a), pre(a′), . . .
post(b), post(b′), . . .
π₁(t)
a
π
π′
π′′


−1
0
t
t + 1
t + 2
t + 3
p, q, ...
pre(a), pre(a′), ...
post(b), post(b′), ...
π1(t)
a
π1(t + 1)
π
π′
π′′
p, q, ...
pre(c), pre(c′), ...
post(a), post(d′), ...


-1
0
t
$t+1$
$t+2$
$t+3$
$p, q, \ldots$
$pre(a), pre(a'), \ldots$
$post(b), post(b'), \ldots$
$\pi_1(t)$
a
a'
$[\pi]_t$
π'
π''



-1
0
t
$t+1$
$t+2$
$t+3$
$p, q, \ldots$
$pre(a), pre(a'), \ldots$
$post(b), post(b'), \ldots$
$\pi_1(t)$
c
a
a'
$[\pi]_t$
π'
π''

Belief-Intention States: Semantics

Truth:

- $\pi, t \models \alpha_{t'}$ iff $\alpha \in \pi(t')_1$ for $\alpha = p, pre(a), post(a)$
- $\pi, t \models Do(a)_{t'}$ iff $\pi(t')_2 = a$
- $\pi, t \models \Box\varphi$ iff for all $\pi' \sim_t \pi$, $\pi', t \models \varphi$
- ....

Belief-Intention States: Semantics

Truth:

- $\pi, t \models \alpha_{t'}$ iff $\alpha \in \pi(t')_1$ for $\alpha = p, pre(a), post(a)$
- $\pi, t \models Do(a)_{t'}$ iff $\pi(t')_2 = a$
- $\pi, t \models \Box\varphi$ iff for all $\pi' \sim_t \pi$, $\pi', t \models \varphi$
- ....

Completeness Theorem. The logic given earlier is sound and complete with respect to the class of all appropriate paths.

Proof. Standard modal reasoning.

Belief-Intention States: Instructions

There are "*instructions*" from the Planner about *future choices* that the agent *agrees* (promises, commits) to follow (if he can).

These instructions may

Belief-Intention States: Instructions

There are "*instructions*" from the Planner about *future choices* that the agent *agrees* (promises, commits) to follow (if he can).

These instructions may

1. be a **complete plan**: for each (future) moment specify a single action $a \in \text{Act}$ the agent *will* perform.

Belief-Intention States: Instructions

There are "*instructions*" from the Planner about *future choices* that the agent *agrees* (promises, commits) to follow (if he can).

These instructions may

1. be a **complete plan**: for each (future) moment specify a single action $a \in \mathrm{Act}$ the agent *will* perform.
2. be a **partial** plan: finite set of pairs (a, t) with $a \in \mathrm{Act}$, $t \in \mathbb{N}$.

Belief-Intention States: Instructions

There are "*instructions*" from the Planner about *future choices* that the agent *agrees* (promises, commits) to follow (if he can).

These instructions may

1. be a **complete plan**: for each (future) moment specify a single action $a \in \mathrm{Act}$ the agent *will* perform.
2. be a **partial** plan: finite set of pairs (a, t) with $a \in \mathrm{Act}$, $t \in \mathbb{N}$.
3. be a **conditional** plan: do a at time t provided φ is true.

Belief-Intention States: Instructions

There are "*instructions*" from the Planner about *future choices* that the agent *agrees* (promises, commits) to follow (if he can).

These instructions may

1. be a **complete plan**: for each (future) moment specify a single action $a \in \mathrm{Act}$ the agent *will* perform.
2. be a **partial** plan: finite set of pairs (a, t) with $a \in \mathrm{Act}$, $t \in \mathbb{N}$.
3. be a **conditional** plan: do a at time t provided φ is true.
4. *restrict available choices* (rather than instructing the agent to follow a specific plan), i.e., **disjunctive plans**.

Belief-Intention States: Instructions

There are "*instructions*" from the Planner about *future choices* that the agent *agrees* (promises, commits) to follow (if he can).

These instructions may

1. be a **complete plan**: for each (future) moment specify a single action $a \in \mathrm{Act}$ the agent *will* perform.
2. be a **partial** plan: finite set of pairs (a, t) with $a \in \mathrm{Act}$, $t \in \mathbb{N}$.
3. be a **conditional** plan: do a at time t provided φ is true.
4. *restrict available choices* (rather than instructing the agent to follow a specific plan), i.e., **disjunctive plans**.
5. be a more complicated structure (**subplans, goals, etc.**)

Belief-Intention States: Instructions

There are "*instructions*" from the Planner about *future choices* that the agent *agrees* (promises, commits) to follow (if he can).

These instructions may

1. be a **complete plan**: for each (future) moment specify a single action $a \in \mathrm{Act}$ the agent *will* perform.
2. be a **partial** plan: finite set of pairs (a, t) with $a \in \mathrm{Act}$, $t \in \mathbb{N}$.
3. be a **conditional** plan: do a at time t provided φ is true.
4. *restrict available choices* (rather than instructing the agent to follow a specific plan), i.e., **disjunctive plans**.
5. be a more complicated structure (**subplans, goals, etc.**)

Belief-Intention Base: Coherence

A **belief-intention base** is a pair (B, I) where 1. B is a set of beliefs and 2. I is a finite set of pairs (a, t)

Belief-Intention Base: Coherence

A **belief-intention base** is a pair (B, I) where 1. B is a set of beliefs and 2. I is a finite set of pairs (a, t)

Coherence: $Cohere_I \;:=\; \Diamond \bigwedge_{(a,t)\in I} pre(a)_t$

- (B, I) is *coherent* if $\neg Cohere_I \not\in B$.
- (Π, I) is *coherent* if there is some $\pi \in \Pi$ such that $\pi, 0 \models Cohere_I$.

1. How should we *describe* the "mental state"?
   Sufficiently rich structure/(modal) language

2. What is a *coherent* description?
   What properties/logic characterize a *rational balance*?

3. What are the inputs?

4. How should we characterize the revision operator?

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).

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).

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.

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.

Typically only doing an action moves "time" forward. However, all three may change the agent's beliefs and current instructions.

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.

Typically only doing an action moves "time" forward. However, all three may change the agent's beliefs and current instructions.

The Revision Problem

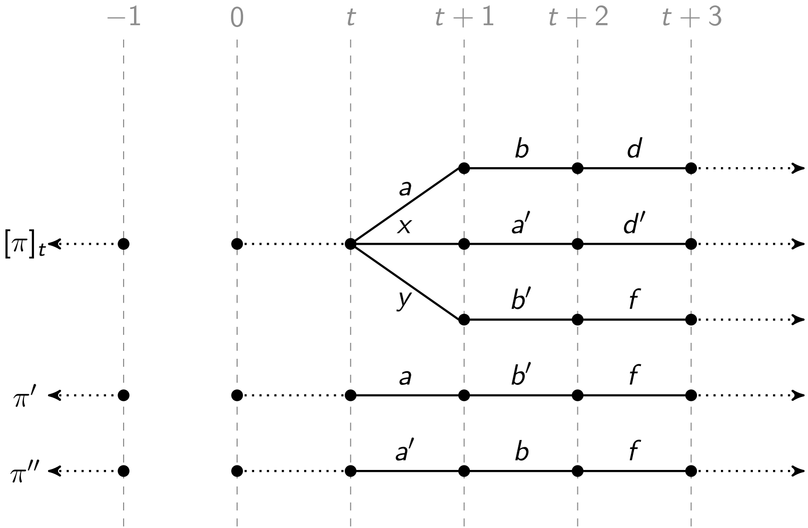
1. How should we *describe* the "mental state"?
   Sufficiently rich structure/(modal) language

2. What is a *coherent* description?
   What properties/logic characterize a rational balance*?*

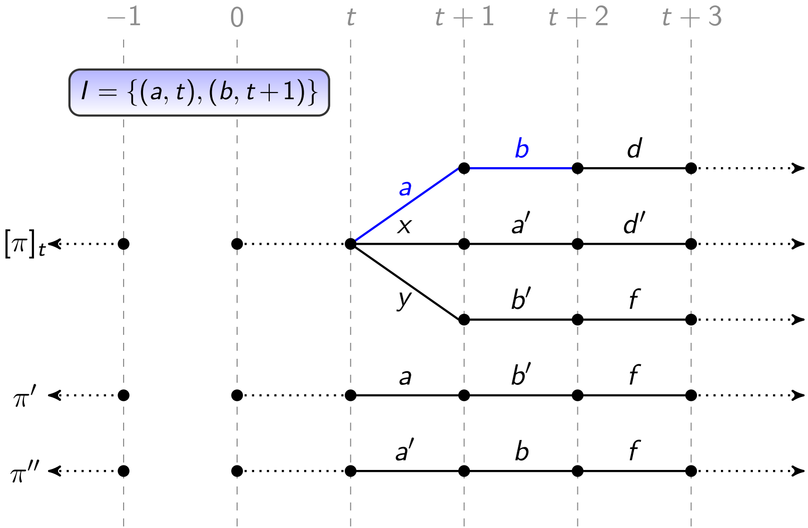
3. What are the inputs?
   Formulas and new plans

4. How should we characterize the revision operator?

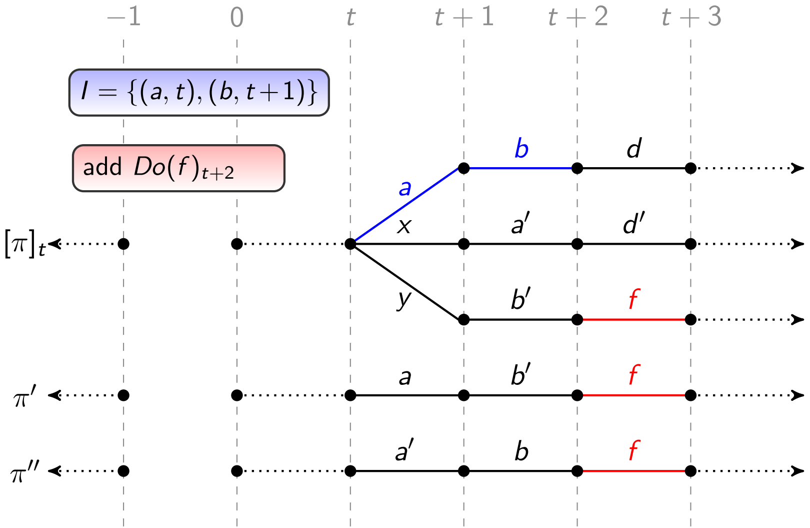
The Revision Problem



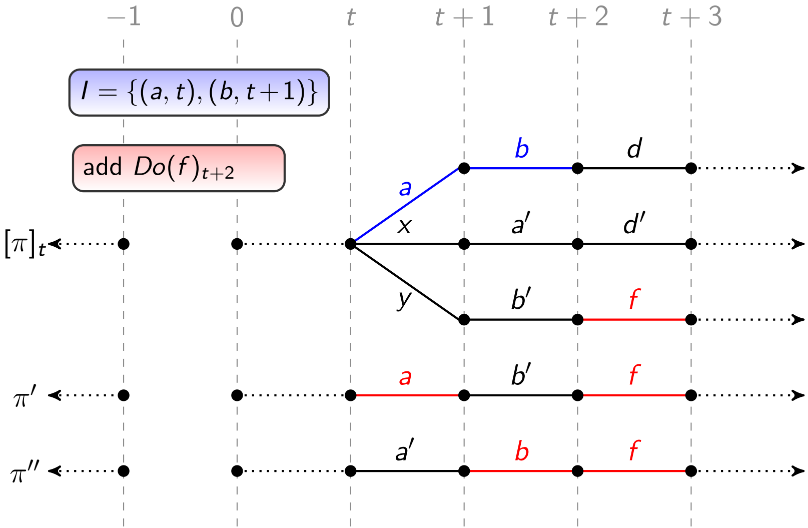
The Revision Problem



The Revision Problem



The Revision Problem



The Revision Problem

Let (B, I) be a coherent belief-intention base.

The Revision Problem

Let (B, I) be a coherent belief-intention base. In general, after revising by φ, the constraint of coherence may force a choice between *any* subset of I (including $\emptyset$).

The Revision Problem

Let (B, I) be a coherent belief-intention base. In general, after revising by φ, the constraint of coherence may force a choice between *any* subset of I (including $\emptyset$).

Which element of $\mathcal{P}(I)$ "should" be the new plan?

The Revision Problem

Let (B, I) be a coherent belief-intention base. In general, after revising by φ, the constraint of coherence may force a choice between *any* subset of I (including $\emptyset$).

Which element of $\mathcal{P}(I)$ "should" be the new plan? *Depends on many features of the plan not represented in the current framework: subplan structure, goals, costs, etc.*

The Revision Problem

Let (B, I) be a coherent belief-intention base. In general, after revising by φ, the constraint of coherence may force a choice between *any* subset of I (including $\emptyset$).

Which element of $\mathcal{P}(I)$ "should" be the new plan? *Depends on many features of the plan not represented in the current framework: subplan structure, goals, costs, etc.*

Intention revision: what is the difference between "add $Do(a)_t$" and "add (a, t) to I"?

The Revision Problem

$$(B, I) \quad \star \quad x \quad = \quad (B', I')$$

The Revision Problem

$$\underset{\text{Coherent}}{(B, I)} \quad \star \quad x \quad = \quad \underset{\text{Coherent}}{(B', I')}$$

The Revision Problem

$$\underset{\text{Coherent}}{(B, I)} \quad \star \quad x \quad = \quad \underset{\text{Coherent}}{(B', I')}$$

$$(a, t) \qquad \varphi$$

The Revision Problem

AGM Revision: $*$ Intention Revision: $\circ$

$$(B, I) \star x = (B', I')$$

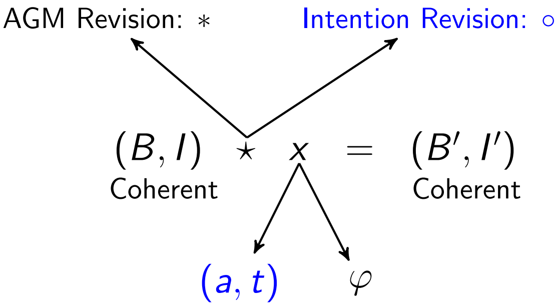
Coherent Coherent

(a, t) φ

The Revision Problem

$$
\begin{tikzcd}
\text{AGM Revision: } * & & & \text{Intention Revision: } \circ & \\
\underset{\text{Coherent}}{(B, I)} & \star \arrow[ul] \arrow[urr] & x \arrow[dl] \arrow[dr] & = & \underset{\text{Coherent}}{(B', I')} \\
& (a, t) & & \varphi &
\end{tikzcd}
$$

The Revision Problem



Postulates: Adding an Intention

Suppose (B, I) is coherent and (a, t) and action/time pair.

Intention Revision Operator: $(B, I) \circ (a, t) = (B', I')$ where

Postulates: Adding an Intention

Suppose (B, I) is coherent and (a, t) and action/time pair.

Intention Revision Operator: $(B, I) \circ (a, t) = (B', I')$ where

1. (B', I') is coherent
2. If $(B, \{(a, t)\})$ is coherent then $(a, t) \in I'$
3. If $(B, I \cup \{(a, t)\})$ is coherent then $I \cup \{(a, t)\} \subseteq I'$
4. $B' = B$

Postulates: AGM Revision

Suppose (B, I) is coherent and φ a consistent formula.

Postulates: AGM Revision

Suppose (B, I) is coherent and φ a consistent formula.
Let $(B, I) \star \varphi = (B', I')$ where

Postulates: AGM Revision

Suppose (B, I) is coherent and φ a consistent formula.
Let $(B, I) \star \varphi = (B', I')$ where

1. $(B', I') = (B', I) \circ \epsilon$ where $\circ$ satisfies the earlier postulates
2. $\varphi \in B'$
3. If $\neg\varphi \notin B$ then $Cl(B \cup \{\varphi\}) = B'$
4. If φ and ψ are *equivalent* and $(B, I) * \psi = (B'', I'')$ then $B' = B''$
5. $B' = Cl(B')$
6. If $\neg\psi \notin B'$ and $(B, I) * \psi = (B'', I'')$ then $Cl(B' \cup \{\psi\}) \subseteq B''$
7. If $(B, I'') * \varphi = (B'', I''')$ then $B' = B''$.

Modeling Revision

$(\Pi, I, \leq, \gamma)$ is a belief intention model where

- Π is an appropriate set of paths
- $\leq$ is a total preorder on Π
- $(min_{\leq}(\Pi), I)$ is coherent
- γ is a **selection function** mapping triples $(\Pi', I', (a, t))$ to J where
  - (Π', J) is coherent
  - If $(\Pi', \{(a, t)\})$ is coherent then $(a, t) \in J$
  - If $(\Pi', I' \cup \{(a, t)\})$ is coherent then $J = I' \cup \{(a, t)\}$
  - $J \subseteq I' \cup \{(a, t)\}$

Modeling Revision

$(\Pi, I, \leq, \gamma)$ is a belief intention model where

Modeling Revision

$(\Pi, I, \leq, \gamma)$ is a belief intention model where

Adding an intention: $(\Pi, I, \leq, \gamma) \circ (a, t) = (\Pi, I', \leq, \gamma')$ where $\gamma = \gamma'$ and $I' = \gamma(min(\Pi), I, (a, t))$

Modeling Revision

$(\Pi, I, \leq, \gamma)$ is a belief intention model where

Adding an intention: $(\Pi, I, \leq, \gamma) \circ (a, t) = (\Pi, I', \leq, \gamma')$ where $\gamma = \gamma'$ and $I' = \gamma(min(\Pi), I, (a, t))$

Adding a belief: $(\Pi, I, \leq, \gamma) * \varphi = (\Pi, I', \leq', \gamma')$ where $\gamma = \gamma'$, $\leq'$ is the **lexicographic** re-ordering of $\leq$ by φ and $I' = \gamma(min_{\leq'}(\Pi), I, \epsilon)$.

Conclusions

1. How should we *describe* the "mental state"?
   Sufficiently rich structure/(modal) language

2. What is a *coherent* description?
   What properties/logic characterize a rational balance*?*

3. What are the inputs?
   Formulas and new plans

4. How should we characterize the revision operator?
   general postulates, dynamic modal operators

Conclusions

- Contingent beliefs: $B^I = Cl(B \cup \{Do(a)_t \mid (a, t) \in I\})$. We can read off postulates.

Conclusions

- Contingent beliefs: $B^I = Cl(B \cup \{Do(a)_t \mid (a,t) \in I\})$. We can read off postulates.
- What is the precise relationship with existing BDI logics (eg., DEL-style dynamics)? ▸ Example

Conclusions

- Contingent beliefs: $B^I = Cl(B \cup \{Do(a)_t \mid (a,t) \in I\})$. We can read off postulates.
- What is the precise relationship with existing BDI logics (eg., DEL-style dynamics)? ▸ Example
- Complex plans (eg., conditional, disjunctive plans): what is the appropriate notion of *coherence*?

Conclusions

- Contingent beliefs: $B^I = Cl(B \cup \{Do(a)_t \mid (a,t) \in I\})$. We can read off postulates.
- What is the precise relationship with existing BDI logics (eg., DEL-style dynamics)? Example
- Complex plans (eg., conditional, disjunctive plans): what is the appropriate notion of *coherence*?
- Iterated revision...

Conclusions

- Contingent beliefs: $B^I = Cl(B \cup \{Do(a)_t \mid (a,t) \in I\})$. We can read off postulates.
- What is the precise relationship with existing BDI logics (eg., DEL-style dynamics)? Example
- Complex plans (eg., conditional, disjunctive plans): what is the appropriate notion of *coherence*?
- Iterated revision...
- Other mental attitudes...

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: Conclusions

- Special issue of JOLLI: Temporal Logics of Agency (eds. J. van Benthem and EP)
- 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)

Thank You!

Dynamic Operator

[Recall a BDI state: $\Gamma_0 = \langle \mathcal{B}_0, \mathcal{D}_0, \langle \mathcal{I}_0, \mathcal{A}_0 \rangle, \mathcal{P}_0 \rangle$
$\Gamma_1 = \langle \mathcal{B}_1, \mathcal{D}_1, \langle \mathcal{I}_1, \mathcal{A}_1 \rangle, \mathcal{P}_1 \rangle$]

Let Ω be a finite set of *observations* (ground or temporal formulas).

Dynamic Operator

[Recall a BDI state: $\Gamma_0 = \langle \mathcal{B}_0, \mathcal{D}_0, \langle \mathcal{I}_0, \mathcal{A}_0 \rangle, \mathcal{P}_0 \rangle$
$\Gamma_1 = \langle \mathcal{B}_1, \mathcal{D}_1, \langle \mathcal{I}_1, \mathcal{A}_1 \rangle, \mathcal{P}_1 \rangle$]

Let Ω be a finite set of *observations* (ground or temporal formulas).

$\Gamma_0 \ \longrightarrow_{\Omega} \ \Gamma_1$ iff:

1. $\mathcal{B}_1 = \mathcal{B}_0 *_{AGM} \bigwedge_{\omega \in \Omega} \omega$ (and not inconsistent)

Dynamic Operator

[Recall a BDI state: $\Gamma_0 = \langle \mathcal{B}_0, \mathcal{D}_0, \langle \mathcal{I}_0, \mathcal{A}_0 \rangle, \mathcal{P}_0 \rangle$
$\Gamma_1 = \langle \mathcal{B}_1, \mathcal{D}_1, \langle \mathcal{I}_1, \mathcal{A}_1 \rangle, \mathcal{P}_1 \rangle$]

Let Ω be a finite set of *observations* (ground or temporal formulas).

$\Gamma_0 \longrightarrow_\Omega \Gamma_1$ iff:

1. $\mathcal{B}_1 = \mathcal{B}_0 *_{AGM} \bigwedge_{\omega \in \Omega} \omega$ (and not inconsistent)
2. $\mathcal{A}_1 = \mathit{activePlans}(\mathit{cleanup}(\mathcal{B}_1, \mathcal{A}_0, \mathcal{D}, \mathcal{P}), \mathcal{B}_1, \mathcal{D}, \mathcal{P})$

Dynamic Operator

[Recall a BDI state: $\Gamma_0 = \langle \mathcal{B}_0, \mathcal{D}_0, \langle \mathcal{I}_0, \mathcal{A}_0 \rangle, \mathcal{P}_0 \rangle$
$\Gamma_1 = \langle \mathcal{B}_1, \mathcal{D}_1, \langle \mathcal{I}_1, \mathcal{A}_1 \rangle, \mathcal{P}_1 \rangle$]

Let Ω be a finite set of *observations* (ground or temporal formulas).

$\Gamma_0 \longrightarrow_\Omega \Gamma_1$ iff:

1. $\mathcal{B}_1 = \mathcal{B}_0 *_{AGM} \bigwedge_{\omega \in \Omega} \omega$ (and not inconsistent)
2. $\mathcal{A}_1 = \mathit{activePlans}(\mathit{cleanup}(\mathcal{B}_1, \mathcal{A}_0, \mathcal{D}, \mathcal{P}), \mathcal{B}_1, \mathcal{D}, \mathcal{P})$
3. $\mathcal{I}_1 = \mathit{Int}(\mathcal{A}_1)$

Derived Properties

- Versions of AGM postulates for beliefs: eg., $[\Omega] \bigwedge_{w \in \Omega} B\omega$

Derived Properties

- Versions of AGM postulates for beliefs: eg., $[\Omega] \bigwedge_{w \in \Omega} B\omega$
- $[\Omega] B\varphi \rightarrow [\Omega] \neg I\varphi$

Derived Properties

- Versions of AGM postulates for beliefs: eg., $[\Omega] \bigwedge_{w \in \Omega} B\omega$
- $[\Omega]B\varphi \rightarrow [\Omega]\neg I\varphi$
- $I \bigwedge_{\omega \in \Omega} \omega \rightarrow \neg[\Omega]\bot$

Derived Properties

- Versions of AGM postulates for beliefs: eg., $[\Omega] \bigwedge_{w \in \Omega} B\omega$
- $[\Omega]B\varphi \rightarrow [\Omega]\neg I\varphi$
- $I \bigwedge_{\omega \in \Omega} \omega \rightarrow \neg[\Omega]\bot$
- $(D\varphi \wedge I\varphi \wedge \neg[\Omega]B\varphi \wedge \neg[\Omega]B\neg\varphi) \rightarrow [\Omega]I\varphi$

Back to conclusions

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

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

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

Back