Logic and Artificial Intelligence

Lecture 23

Eric Pacuit

Currently Visiting the Center for Formal Epistemology, CMU

Center for Logic and Philosophy of Science
Tilburg University

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

November 28, 2011

Merging Logics of Rational Agency

- ✓ Entangling Knowledge/Beliefs and Preferences
- "Epistemizing" Logics of Action and Ability
- BDI (Belief + Desires + Intentions) Logics

Knowing how to win

Consider the following game: Two cards, Ace and Joker, lie face down and the agent i must choose one. The Ace wins, the Joker loses.

Knowing how to win

Consider the following game: Two cards, Ace and Joker, lie face down and the agent i must choose one. The Ace wins, the Joker loses.

- Does the agent i have a strategy to win the game?

Knowing how to win

Consider the following game: Two cards, Ace and Joker, lie face down and the agent i must choose one. The Ace wins, the Joker loses.

- Does the agent i have a strategy to win the game?
- Does the agent i know that she has a strategy to win the game?

Knowing how to win

Consider the following game: Two cards, Ace and Joker, lie face down and the agent i must choose one. The Ace wins, the Joker loses.

- Does the agent i have a strategy to win the game?
- Does the agent i know that she has a strategy to win the game?
- Does the agent i know a strategy to win the game?

J. Fantl. *Knowing-how and knowing-that*. Philosophy Compass, 3 (2008), 451 470.

M.P. Singh. *Know-how*. In Foundations of Rational Agency (1999), M. Wooldridge and A. Rao, Eds., pp. 105 132.

Related Work: Knowing How to Execute a Plan

J. van Benthem. *Games in dynamic epistemic logic*. Bulletin of Economics Research 53, 4 (2001), 219 248..

J. Broersen. *A logical analysis of the interaction between obligation-to- do and knowingly doing*. In Proceedings of DEON 2008.

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.

W. Jamroga and T. Agotnes. *Constructive Knowledge: What Agents can Achieve under Imperfect Information*. Journal of Applied Non-Classical Logics 17(4):423–425, 2007.

The Logic of Know-How

The Logic of Know-How

- $K(R \vee B) \rightarrow K(R) \vee K(B)$: "If Ann knows that she can choose a red or blue card, then either she knows that she can choose a red card or she knows that she can choose a blue card."

The Logic of Know-How

- $K(R \vee B) \rightarrow K(R) \vee K(B)$: "If Ann knows that she can choose a red or blue card, then either she knows that she can choose a red card or she knows that she can choose a blue card."
- $C(R' \vee B') \rightarrow C(R') \vee C(B')$: "If Ann can choose either a red or blue card then either she can choose a red card or she can choose a black card."

The Logic of Know-How

- $K(R \vee B) \rightarrow K(R) \vee K(B)$: "If Ann knows that she can choose a red or blue card, then either she knows that she can choose a red card or she knows that she can choose a blue card."
- $C(R' \vee B') \rightarrow C(R') \vee C(B')$: "If Ann can choose either a red or blue card then either she can choose a red card or she can choose a black card."
- $Abl(R' \vee B') \rightarrow Abl(R') \vee Abl(B')$: "If Ann has the ability to select a red or blue card then either she has the ability to choose a red card or she has the ability to choose a black card."

The Logic of Know-How

- $K(R \vee B) \rightarrow K(R) \vee K(B)$: "If Ann knows that she can choose a red or blue card, then either she knows that she can choose a red card or she knows that she can choose a blue card."
- $C(R' \vee B') \rightarrow C(R') \vee C(B')$: "If Ann can choose either a red or blue card then either she can choose a red card or she can choose a black card."
- $Abl(R' \vee B') \rightarrow Abl(R') \vee Abl(B')$: "If Ann has the ability to select a red or blue card then either she has the ability to choose a red card or she has the ability to choose a black card."
- $Khow(R' \vee B') \rightarrow Khow(R') \vee Khow(B')$: "If Ann knows how to select a red or blue card then either she knows how to choose a red card or she knows how to choose a black card."

Grades of Know-How

i knows how to α only if:

1. it is possible that *i* α
2. were *i* to try to α, *i* would α
3. were *i* to try to α is a suitable context, *i* would α
4. *i* is able/has the ability to α particularly well
5. *i* knows that *w* is a way to α
6. *i* knows that *w* is a way for her to α
7. *i* knows why *w* is a way for her to α

J. Fantl. *Knowing-how and knowing-that*. Philosophy Compass 3, 3 (2008), 451 470.

Example

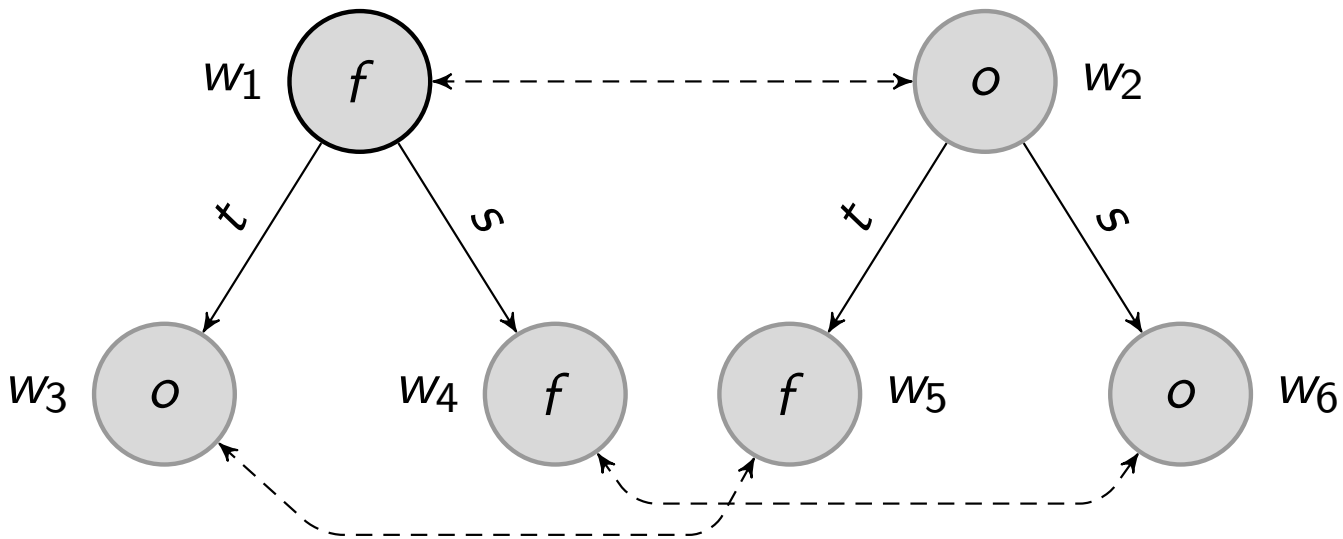
A. Herzig and N. Troquard. *Knowing how to play: uniform choices in logics of agency*. In Proceedings of AAMAS 2006.

Example

Ann, who is blind, is standing with her hand on a light switch. She has two options: toggle the switch (t) or do nothing (s):

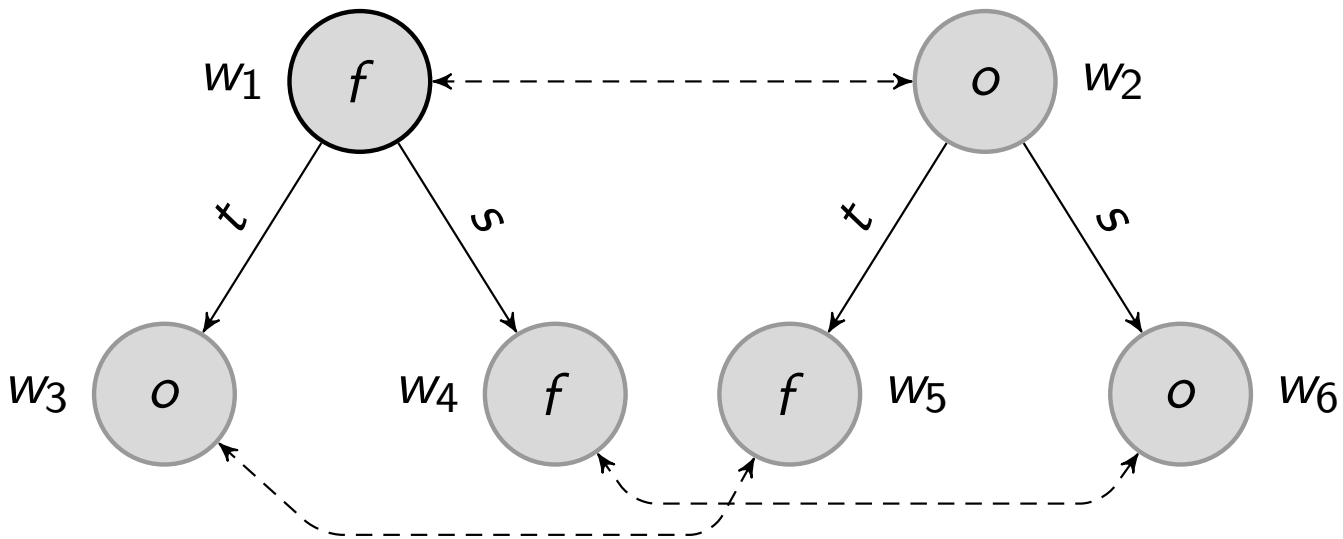
Example

Ann, who is blind, is standing with her hand on a light switch. She has two options: toggle the switch (t) or do nothing (s):



Example

Ann, who is blind, is standing with her hand on a light switch. She has two options: toggle the switch (t) or do nothing (s):



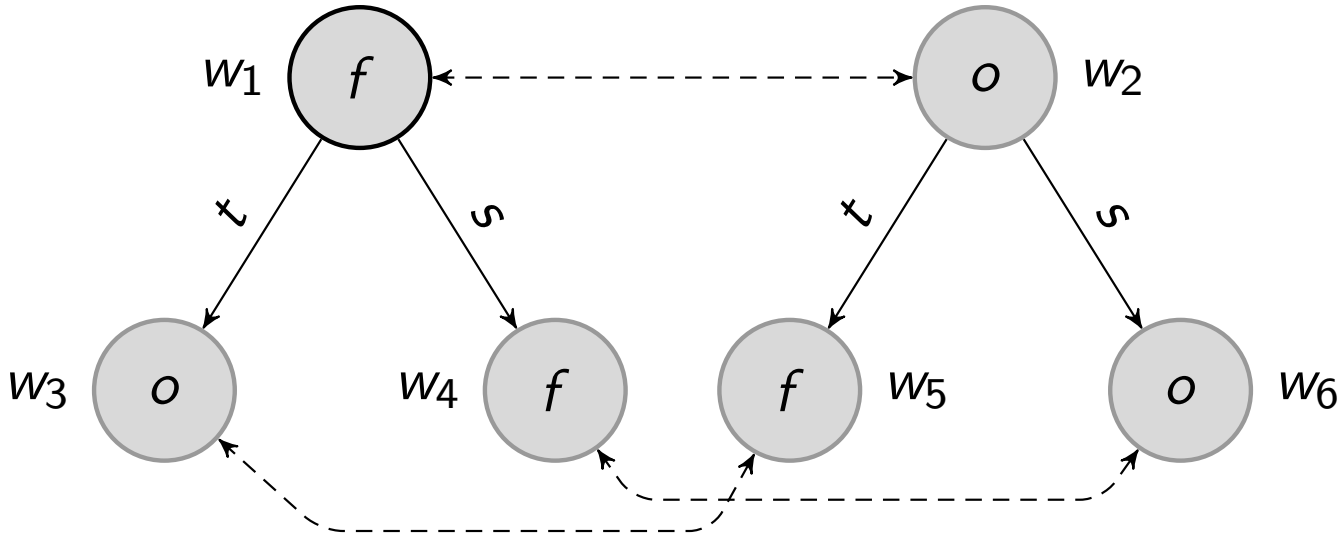
Does she have the *ability* to turn the light on? Is she *capable* of turning the light on? Does she *know how* to turn the light on?

Arena with Imperfect Information

$\mathcal{A}(w) = \{a \mid \text{there is a } v \text{ such that } w \longrightarrow_a v\}$

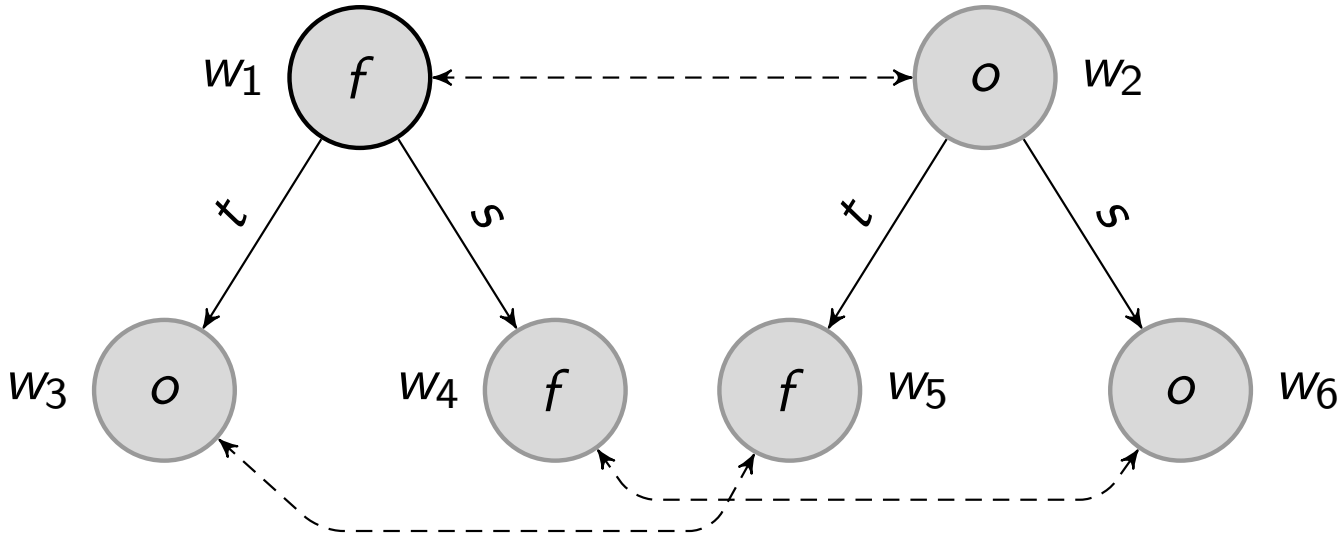
- **No Miracles**: for all $a \in \Sigma$ and all $w, v, w', v' \in W$, if $w \sim v$, $w \longrightarrow_a w'$, and $v \longrightarrow_a v'$, then $w' \sim v'$.
- **Success**: If $w \sim v$ then $\mathcal{A}(v) \subseteq \mathcal{A}(w)$
- **Awareness**: If $w \sim v$ then $\mathcal{A}(w) \subseteq \mathcal{A}(v)$
- **Certainty of available actions**: If $w \sim v$ and $w \sim v'$ then $\mathcal{A}(v) = \mathcal{A}(v')$

Example



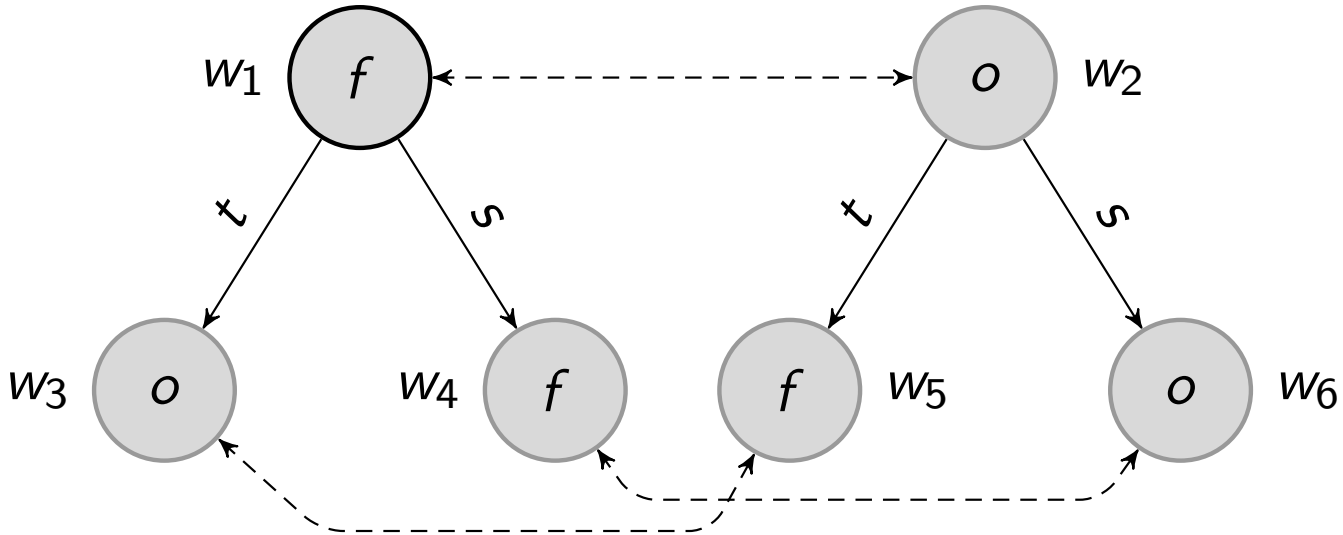
$w_1 \models \neg\Box f$: "Ann does not know the light is on"

Example



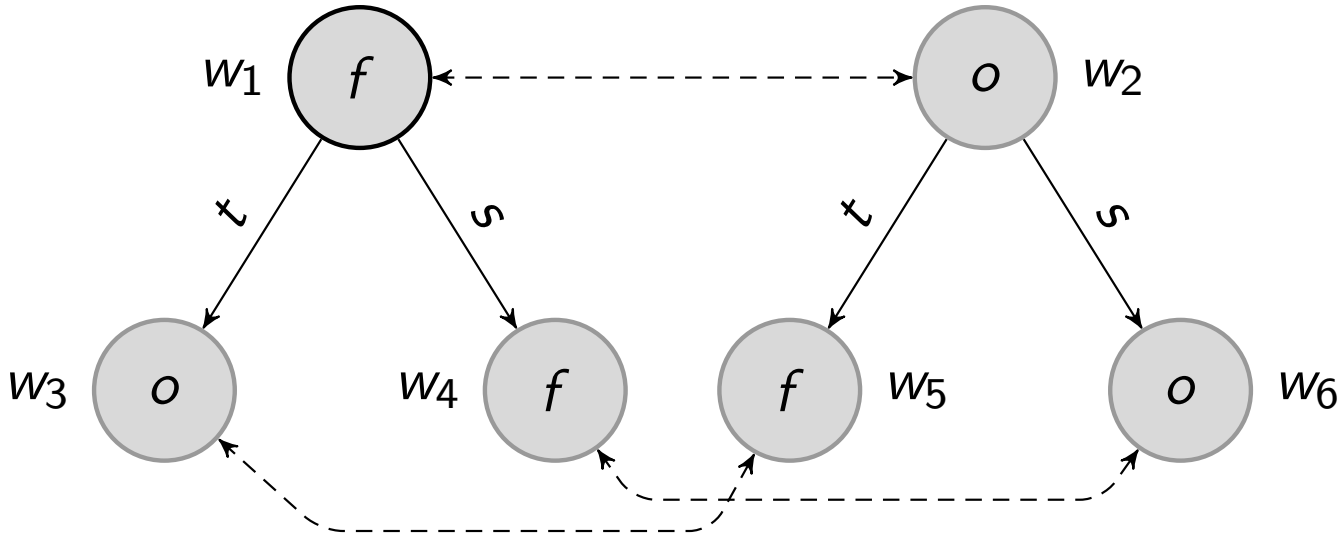
$w_1 \models \langle t \rangle o$ "after toggling the light switch, the light will be on"

Example



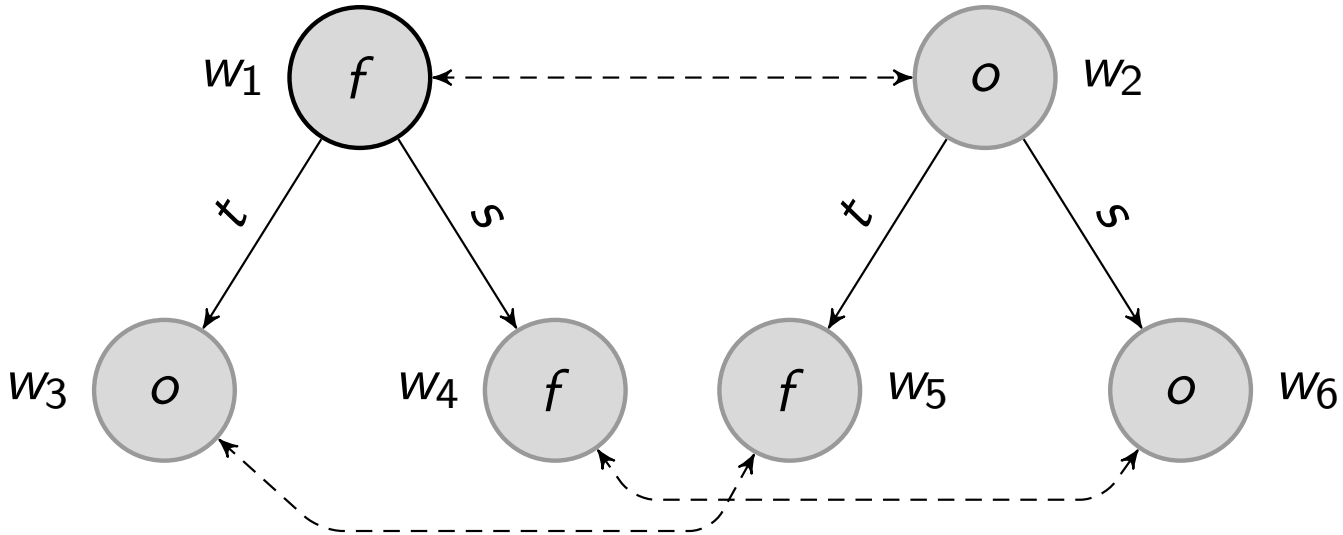
$w_1 \models \neg\Box\langle t\rangle o$: "Ann does not know that after toggling the light switch, the light will be on"

Example



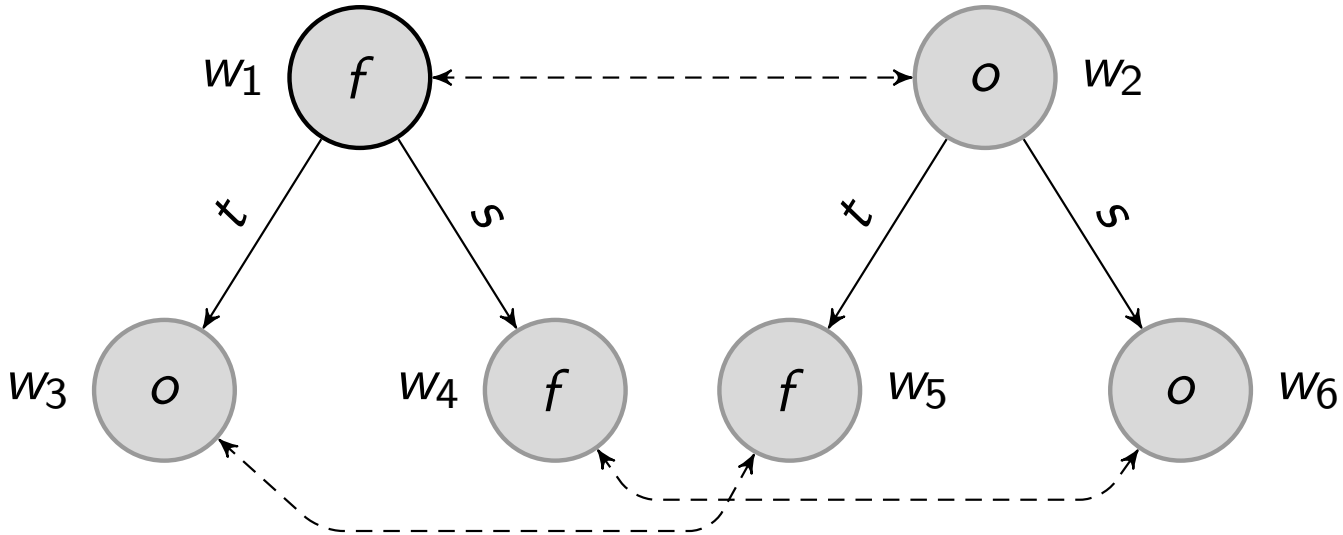
$w_1 \models \Box(\langle t \rangle \top \wedge \langle s \rangle \top)$: "Ann knows that she can toggle the switch and she can do nothing"

Example



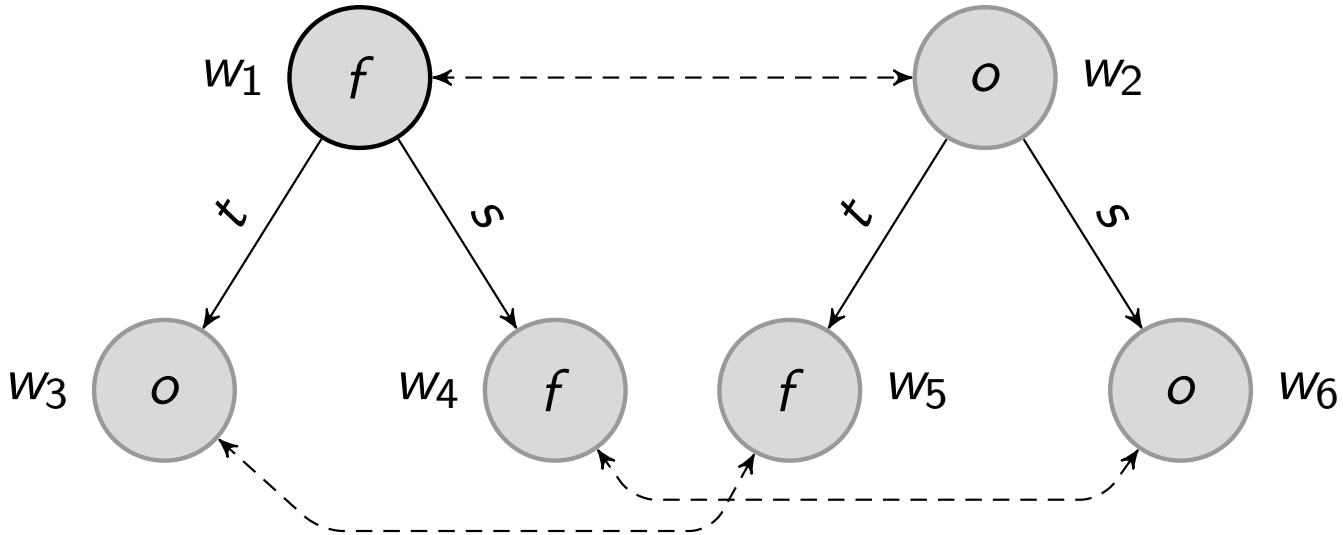
$w_1 \models \langle t \rangle \neg \Box o$: "after toggling the switch Ann does not know that the light is on"

Example



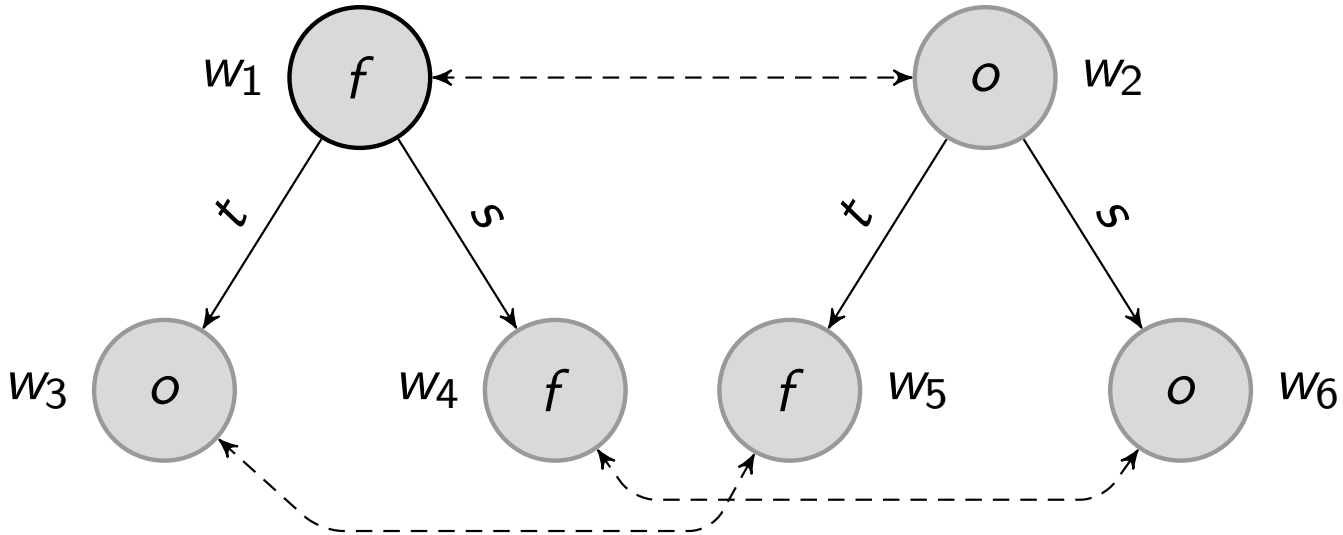
Let l be "turn the light on": a choice between t and s

Example



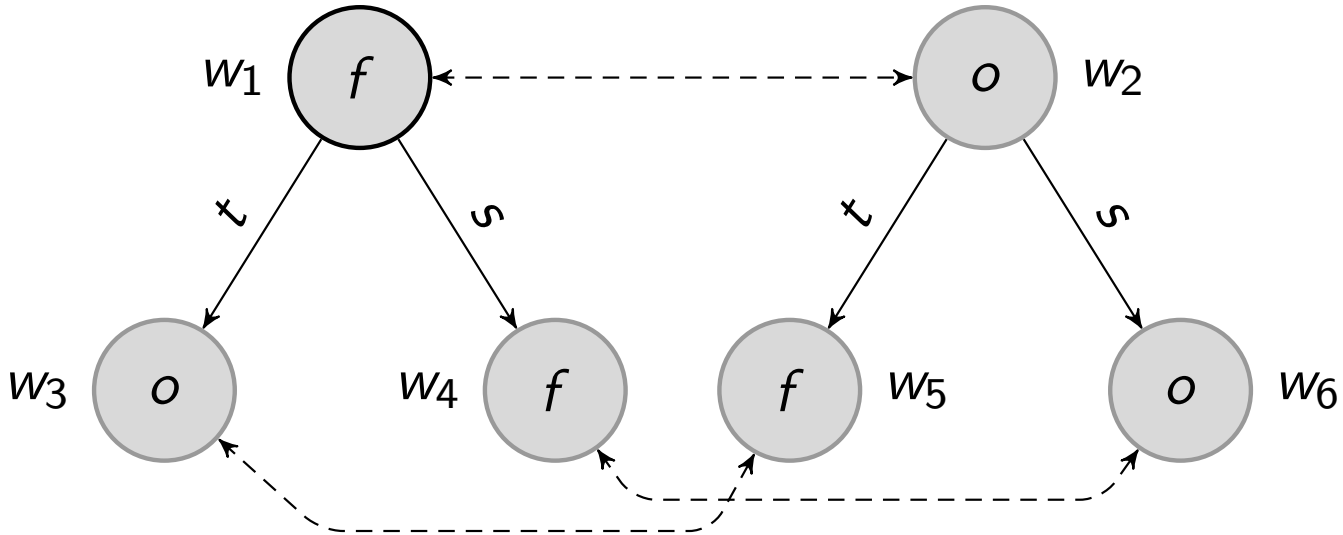
$w_1 \models \langle l \rangle^{\exists} o \wedge \neg \langle l \rangle^{\forall} o$: executing l can lead to a situation where the light is on, but this is not *guaranteed* (i.e., the plan may fail)

Example



$w_1 \models \Box\langle I \rangle^{\exists} o$: Ann knows that she is capable of turning the light on. She has *de re* knowledge that she can turn the light on.

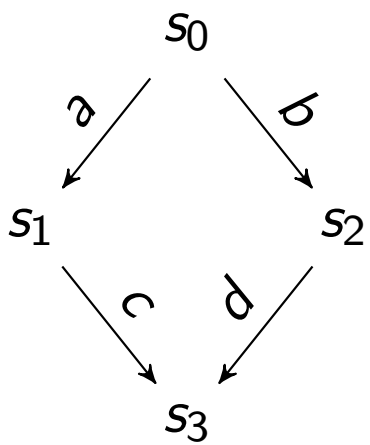
Example

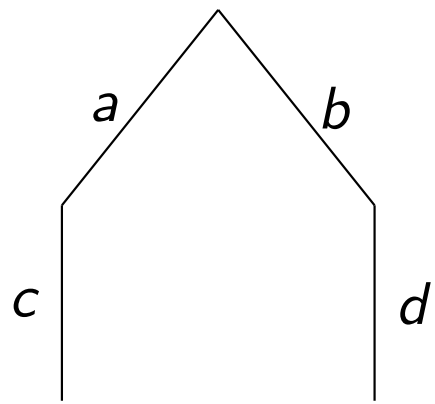


$w_1 \models \neg\langle I \rangle^{\Diamond} o$: Ann cannot knowingly turn on the light: there is no *subjective* path leading to states satisfying o (note that *all* elements of the last element of the subject path must satisfy o).

Enabled vs. Subjectively Enabled

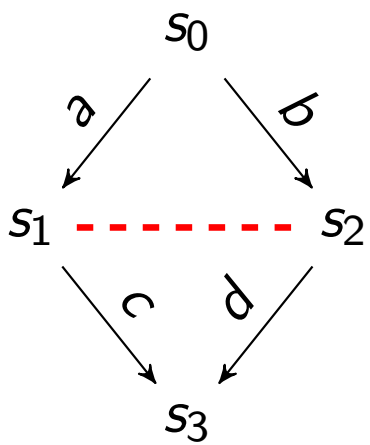
The protocol is **enabled**:

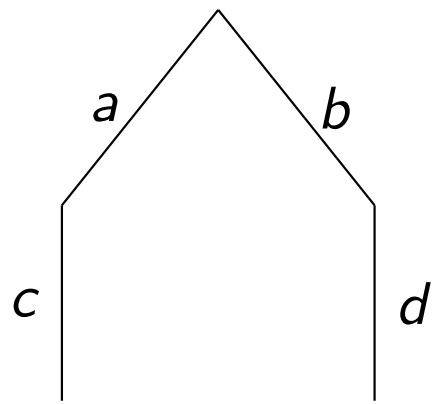




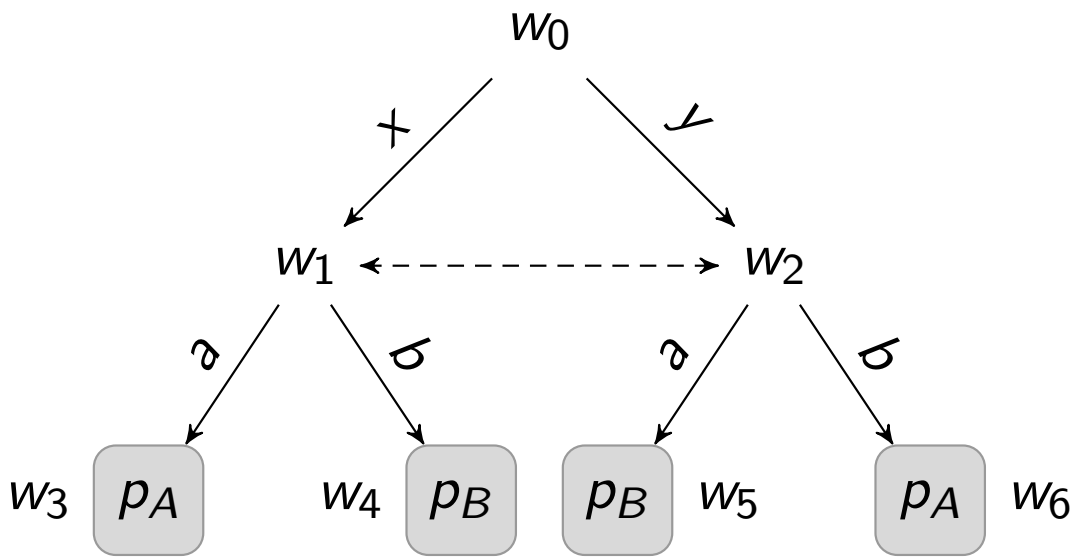
Enabled vs. Subjectively Enabled

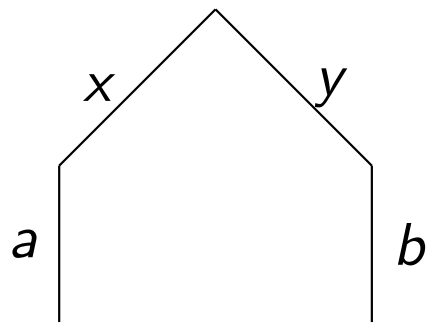
The protocol is not **enabled**:





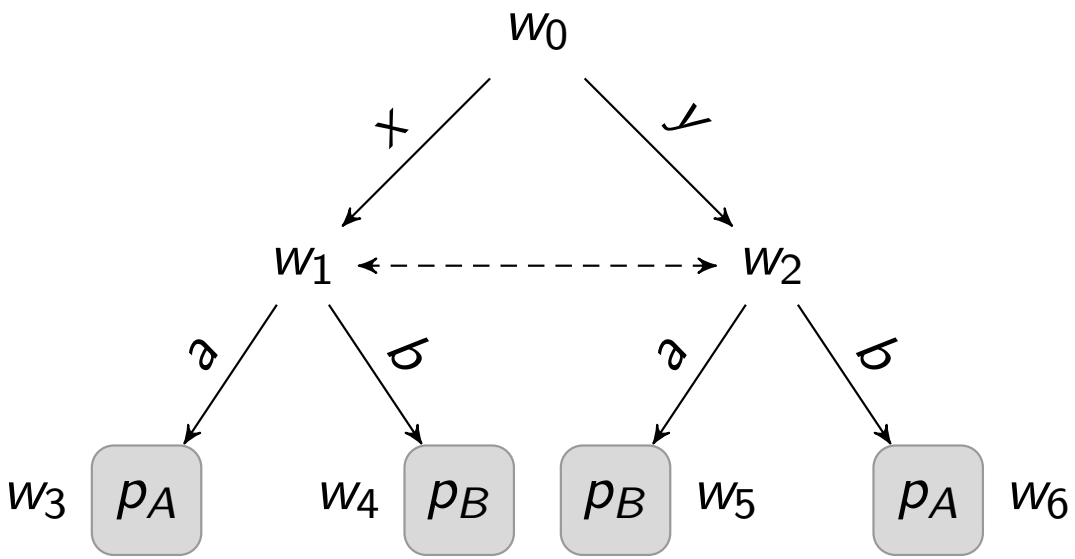
Knowing How to Win

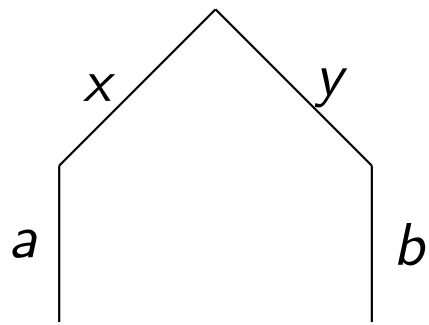




$w_0 \models \langle s \rangle^{\forall} p_A$: "$s$ is a winning strategy for Ann."

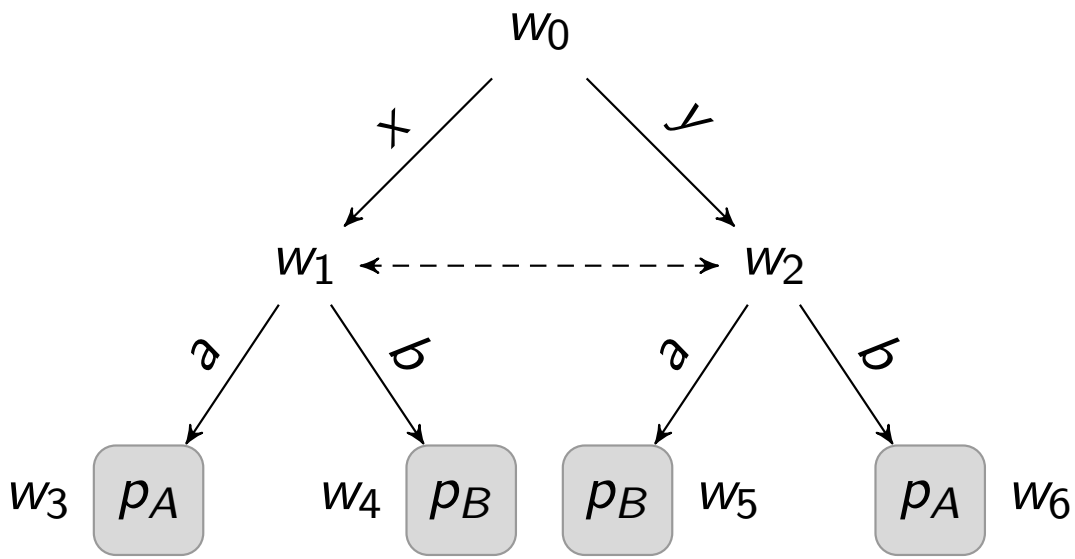
Knowing How to Win

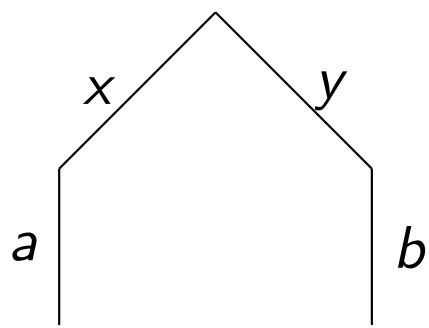




$w_0 \models \Box\langle s \rangle^{\forall} p_A$: "Ann knows that s is a winning strategy."

Knowing How to Win





$w_0 \models \langle s \rangle^{\Box} \top \wedge \neg \langle s \rangle^{\Box} p_A$: "$s$ is *subjectively enabled*, but Ann does not know how to use it to win."

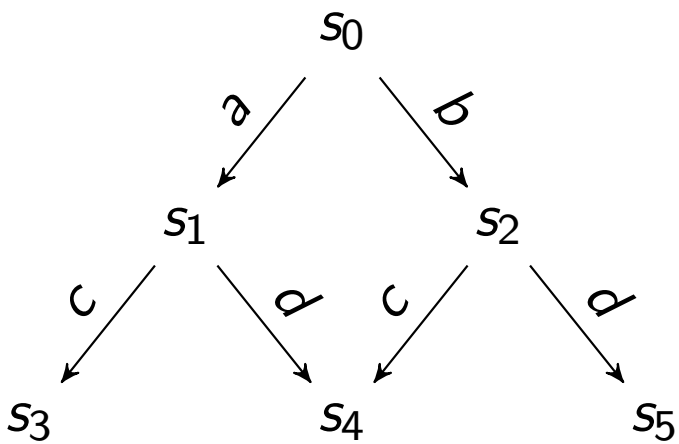
Committing to a *Plan*

Adopting a plan does not commit the agent to a single course of action, but, rather, focuses the agent's attention on the "relevant" decision problems.

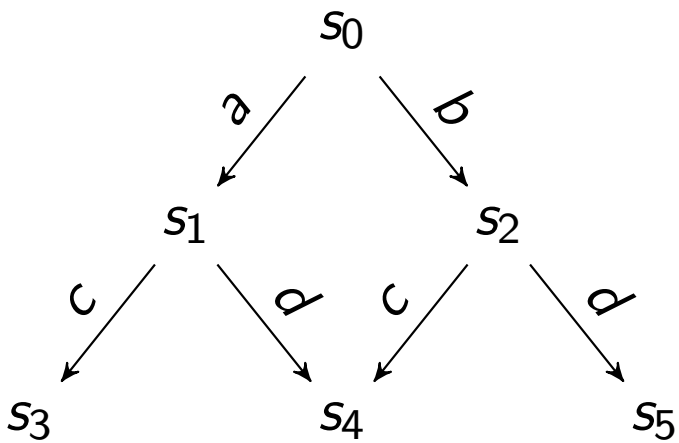
> *plans help make deliberation tractable for limited beings like us. They provide a clear, concrete purpose for deliberation, rather than merely a general injunction to do the best. They narrow the scope of the deliberation to a limited set of options. And they help answer a question that tends to remain unasked within traditional decision theory, namely; where do decision problems come from?*
>
> *(pg. 33)*

M. Bratman. *Intentions, Plans and Practical Reasons*. CSLI Publications, 1999.

Arena

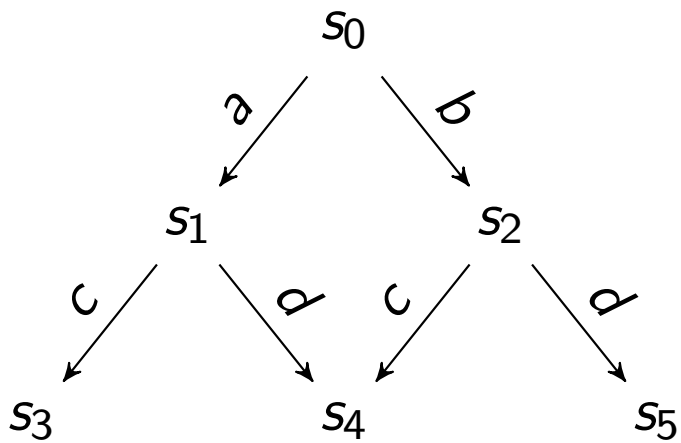


Committing to a choice



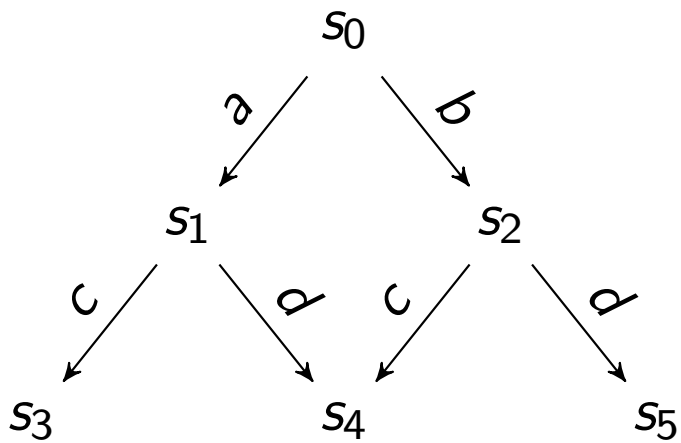
Committing to a choice

At s_0, the agent agrees to either choose c or choose d:



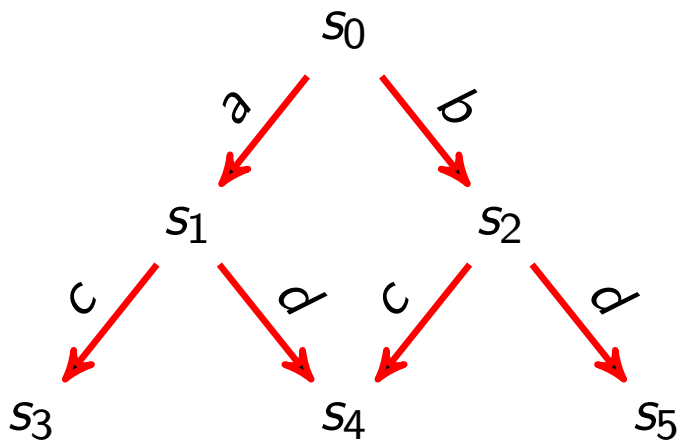
Committing to a choice

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



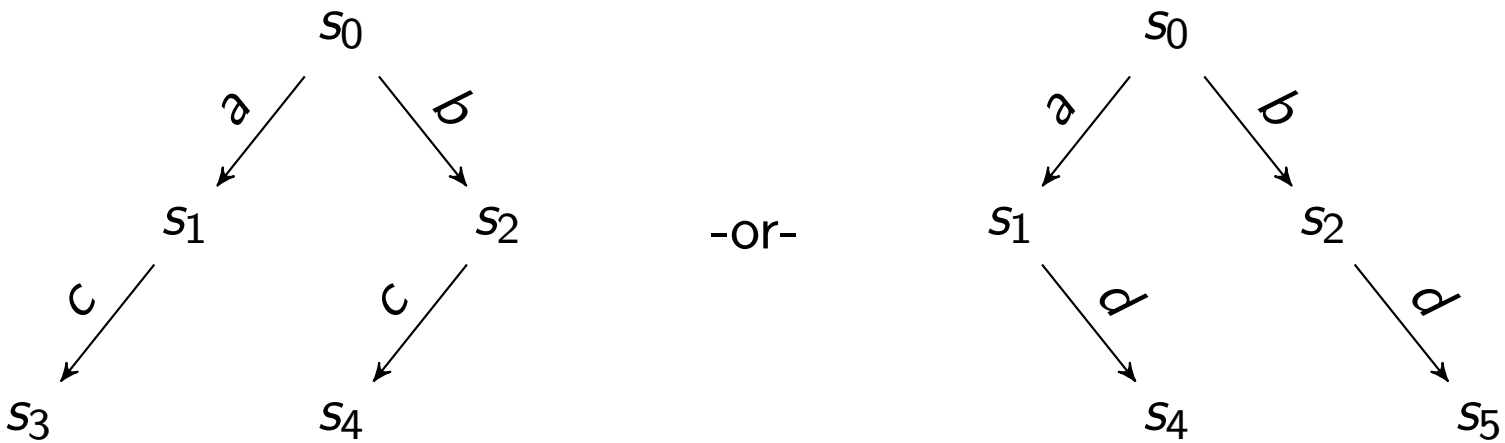
Committing to a choice

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



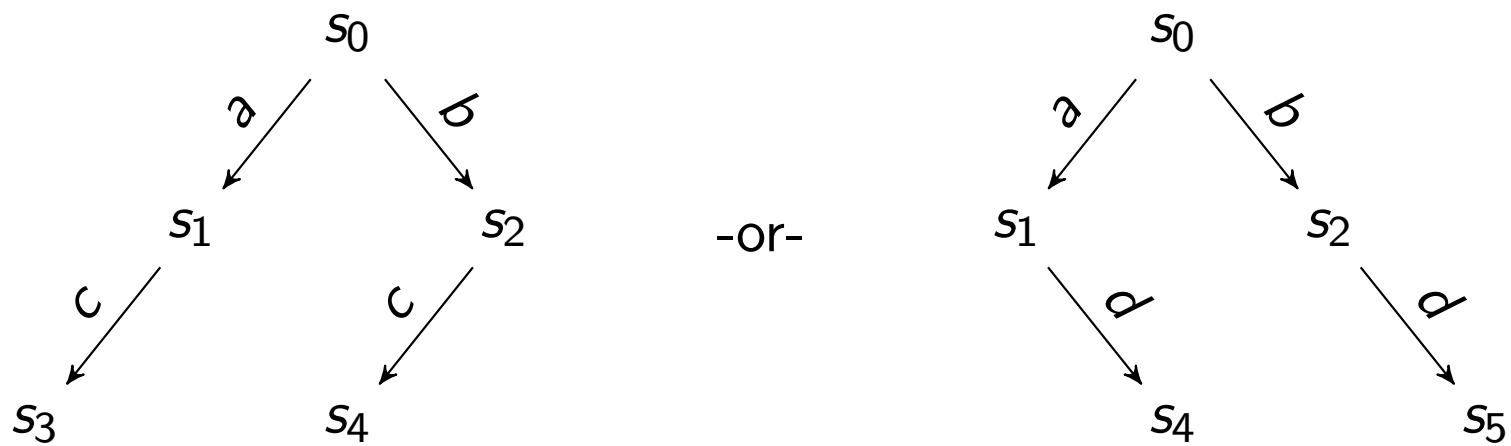
Committing to a choice

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



Committing to a choice

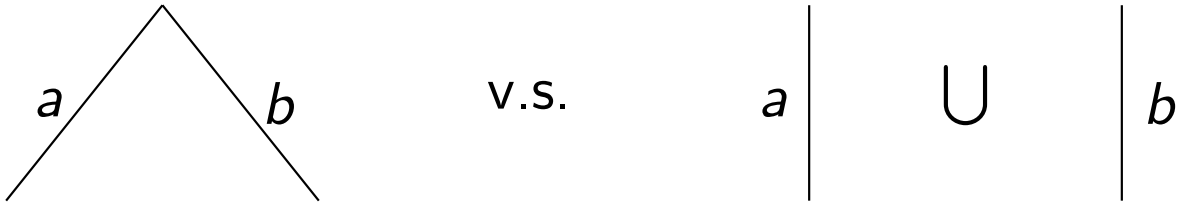
At s_0, the agent agrees to either choose c or choose d:
$(a \cup b); c \cup (a \cup b); d$



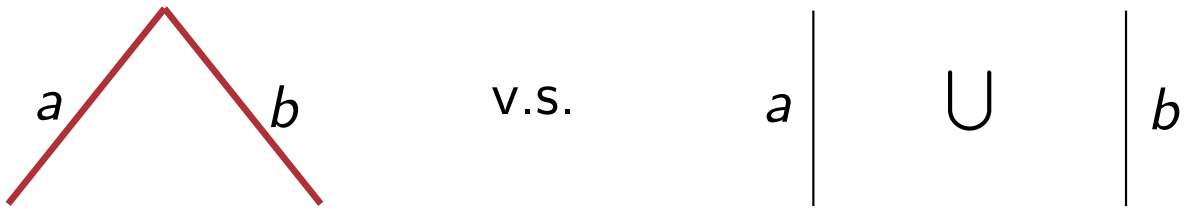
J. van Benthem. *Logical Dynamics of Information and Interaction*. Cambridge University Press, 2010.

Do *a* or Do *b*

Do a or Do b



Do *a* or Do *b*



1. the agent commits to choosing between actions *a* or *b* *when the time comes* (possibly ignoring the other options that may be available to the agent at that moment).

Do *a* or Do *b*

1. the agent commits to choosing between actions *a* or *b* *when the time comes* (possibly ignoring the other options that may be available to the agent at that moment).

2. the agent must choose between two future courses of actions: doing *a* or doing *b*. The point is that *a* and *b* each may lead to a different set of states.

(Partial) Plans

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

These instructions may

(Partial) Plans

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.

(Partial) Plans

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.
2. be a **partial** plan: finite set of pairs (a, t) with $a \in \text{Act}$, $t \in \mathbb{N}$.

(Partial) Plans

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.
2. be a **partial** plan: finite set of pairs (a, t) with $a \in \text{Act}$, $t \in \mathbb{N}$.
3. be a **conditional** plan: do a at time t provided φ is true.

(Partial) Plans

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.
2. be a **partial** plan: finite set of pairs (a, t) with $a \in \text{Act}$, $t \in \mathbb{N}$.
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**.

(Partial) Plans

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.
2. be a **partial** plan: finite set of pairs (a, t) with $a \in \text{Act}$, $t \in \mathbb{N}$.
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.**)

Merging Logics of Rational Agency

- ✓ Entangling Knowledge/Beliefs and Preferences
- ✓ "Epistemizing" Logics of Action and Ability
- ▶ BDI (Belief + Desires + Intentions) Logics

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).

Logics of Intentions: Key Issues

Logics of Intentions: Key Issues

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

Logics of Intentions: Key Issues

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).

Logics of Intentions: Key Issues

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).

Logics of Intentions: Key Issues

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).

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.

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.

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

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

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

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?
- $B\varphi \rightarrow \mathit{Goal}\varphi$?

What is the appropriate underlying logic?

Many proposals, but no clear consensus...

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

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.

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]

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

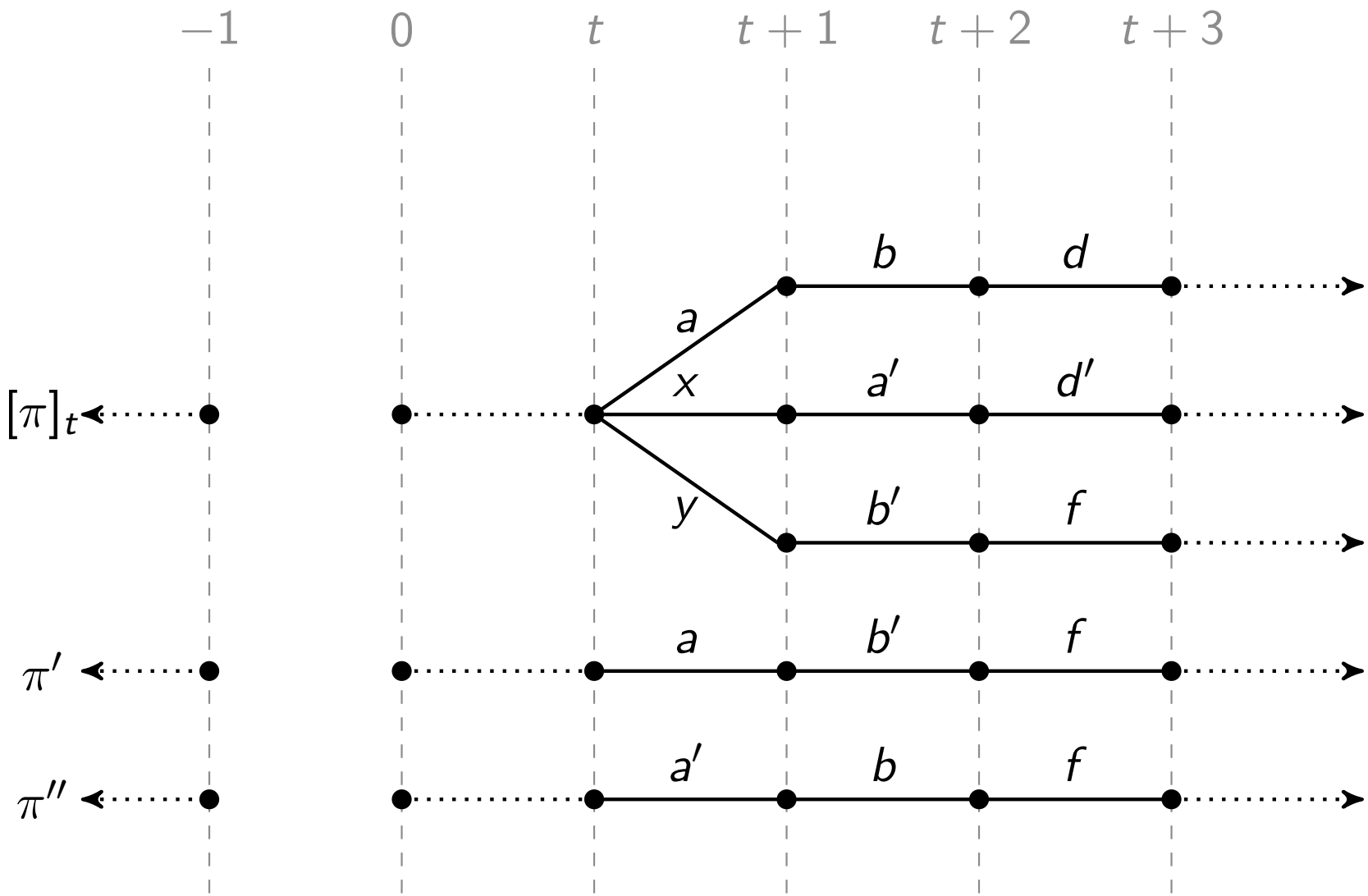
1. Nature can reveal (true) facts about the current choice situation (eg., facts that are true, choices that are available/not available in the future).

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

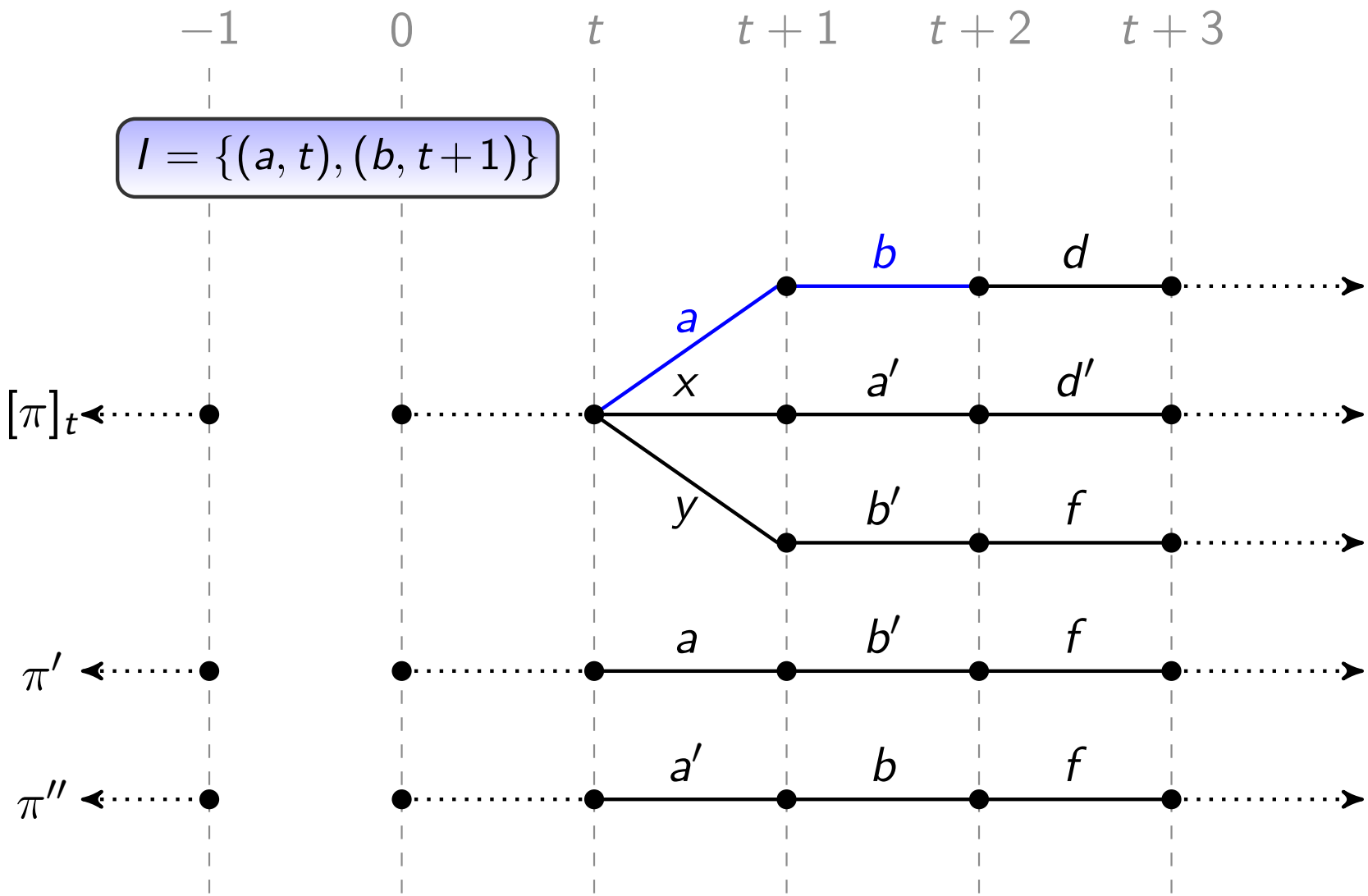
3. The Planner can amend the agent's current set of instructions.

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

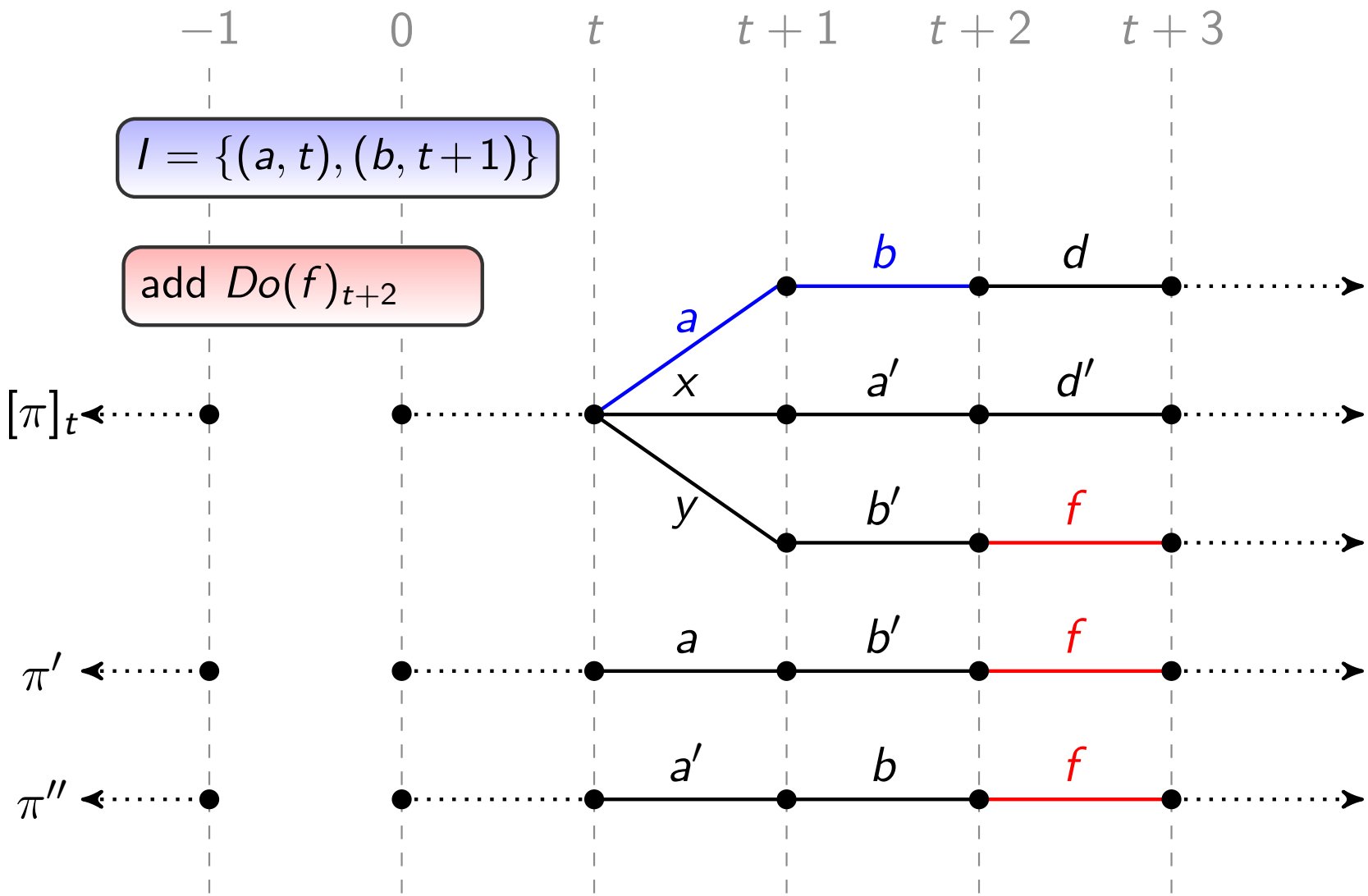
The Revision Problem



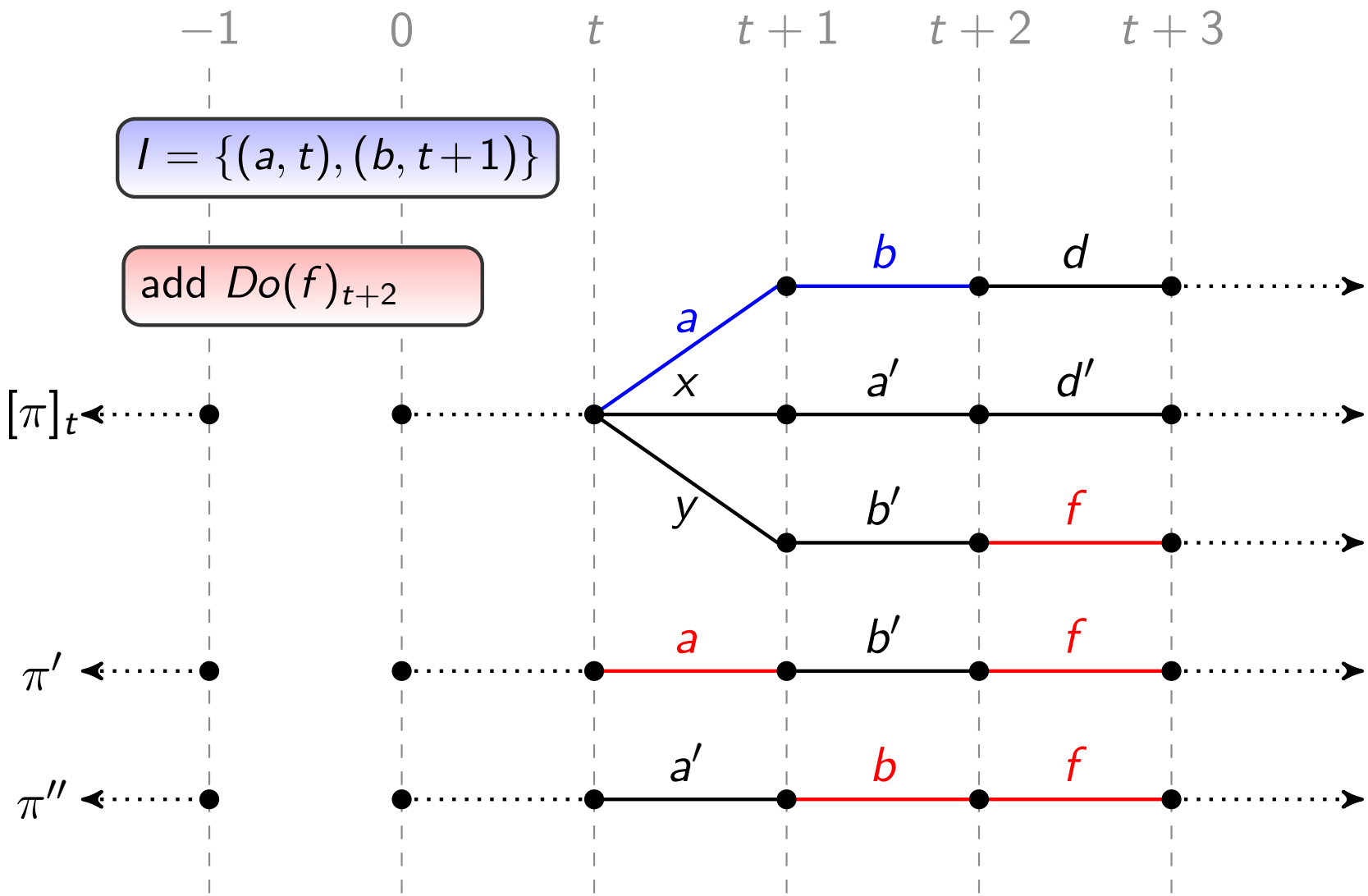
The Revision Problem



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 $\wp(I)$ "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 $\wp(I)$ "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 $\wp(I)$ "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"?

Revising Mental Attitudes

✓ Preference change

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

C. List and F. Dietrich. *A Model of Non-informational Preference Change*. Journal of Theoretical Politics 23(2): 145-164, 2011.

Revising Mental Attitudes

✓ Preference change

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

C. List and F. Dietrich. *A Model of Non-informational Preference Change*. Journal of Theoretical Politics 23(2): 145-164, 2011.

✓ 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).

C. List and F. Dietrich. *A Model of Non-informational Preference Change*. Journal of Theoretical Politics 23(2): 145-164, 2011.

✓ 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).

✓ Intention revision

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