Logics of Rational Agency

Eric Pacuit

Tilburg University

October 6, 2009

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

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

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

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

What is a rational agent?

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

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

There is a jungle of formal systems!

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

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

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

There is a jungle of formal systems!

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

and logical languages for reasoning about them)

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

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

What about *game-theoretic* analyses?

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

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

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

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

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

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

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

Logics of Rational Agency

Basic Ingredients

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

Basic Ingredients

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

Basic Ingredients

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

Basic Ingredients

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

Basic Ingredients

- informational attitudes
- time, actions and ability
- motivational attitudes

Single-Agent Epistemic Logic

Typically, we write KP when the intended interpretation is "P is *known*"

Single-Agent Epistemic Logic

Typically, we write KP when the intended interpretation is "P is *known*"

$K(P \rightarrow Q)$: "Ann *knows* that P implies Q"

Single-Agent Epistemic Logic

Typically, we write KP when the intended interpretation is "P is *known*"

$K(P \to Q)$: "Ann *knows* that P implies Q"

$KP \vee \neg KP$: "either Ann does or does not know P"

Single-Agent Epistemic Logic

Typically, we write KP when the intended interpretation is "P is *known*"

$K(P \to Q)$: "Ann *knows* that P implies Q"

$KP \vee \neg KP$: "either Ann does or does not know P"

$KP \vee K\neg P$: "Ann knows whether P is true"

Single-Agent Epistemic Logic

Typically, we write KP when the intended interpretation is "P is *known*"

$K(P \to Q)$: "Ann *knows* that P implies Q"

$KP \vee \neg KP$: "either Ann does or does not know P"

$KP \vee K\neg P$: "Ann knows whether P is true"

LP: "P is an epistemic possibility"

Single-Agent Epistemic Logic

Typically, we write KP when the intended interpretation is "P is *known*"

$K(P \to Q)$: "Ann *knows* that P implies Q"

$KP \vee \neg KP$: "either Ann does or does not know P"

$KP \vee K\neg P$: "Ann knows whether P is true"

LP: "P is an epistemic possibility"

KLP: "Ann knows that she thinks P is possible"

Example

Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, one of the cards is placed face down on the table and the third card is put back in the deck.

Example

Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, one of the cards is placed face down on the table and the third card is put back in the deck.

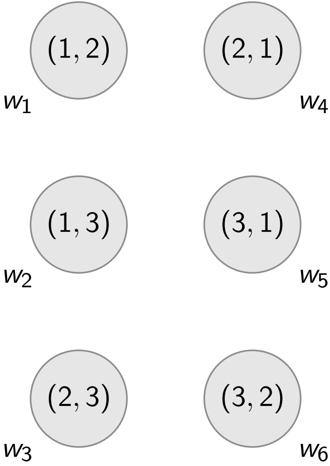
What are the relevant states?

Example

Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, one of the cards is placed face down on the table and the third card is put back in the deck.

What are the relevant states?

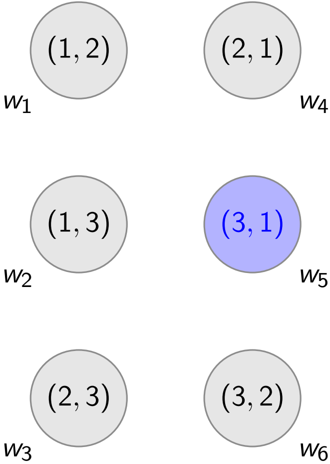


Example

Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, one of the cards is placed face down on the table and the third card is put back in the deck.

Ann receives card 3 and card 1 is put on the table

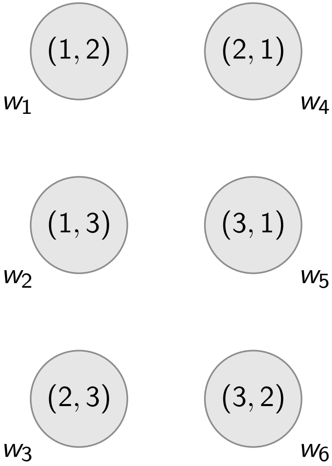


Example

Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, one of the cards is placed face down on the table and the third card is put back in the deck.

What information does Ann have?

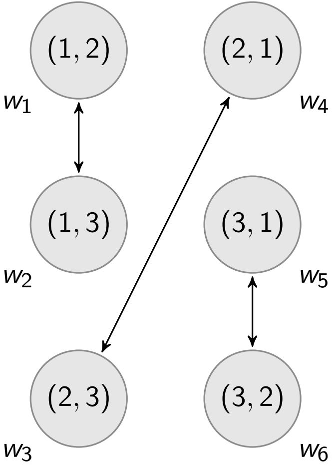


Example

Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, one of the cards is placed face down on the table and the third card is put back in the deck.

What information does Ann have?

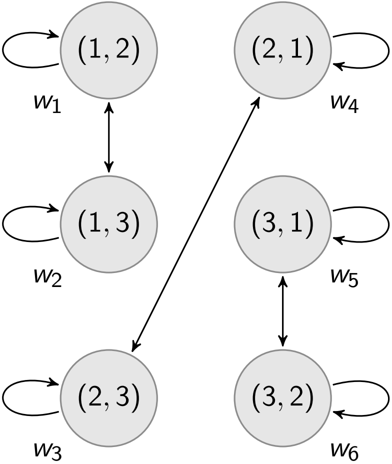


Example

Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, one of the cards is placed face down on the table and the third card is put back in the deck.

What information does Ann have?



Example

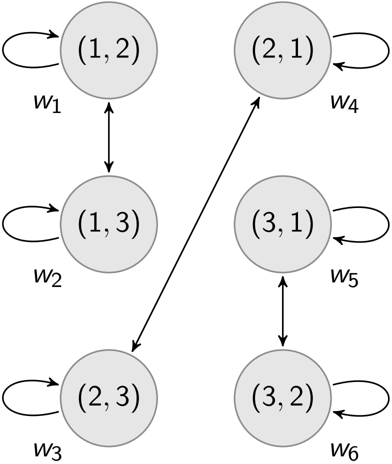
Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, one of the cards is placed face down on the table and the third card is put back in the deck.

Suppose H_i is intended to mean "Ann has card i"

T_i is intended to mean "card i is on the table"

Eg., $V(H_1) = \{w_1, w_2\}$



Example

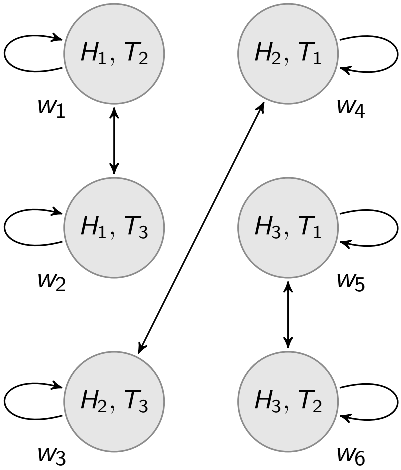
Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, one of the cards is placed face down on the table and the third card is put back in the deck.

Suppose H_i is intended to mean "Ann has card i"

T_i is intended to mean "card i is on the table"

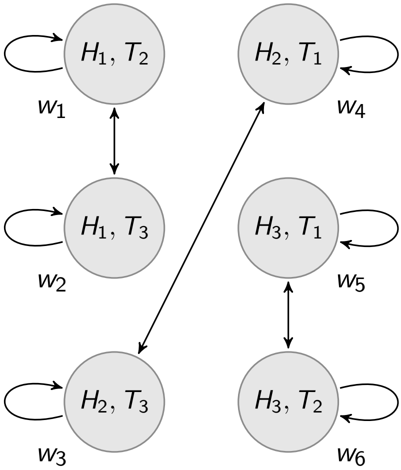
Eg., $V(H_1) = \{w_1, w_2\}$



Example

Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, one of the cards is placed face down on the table and the third card is put back in the deck.

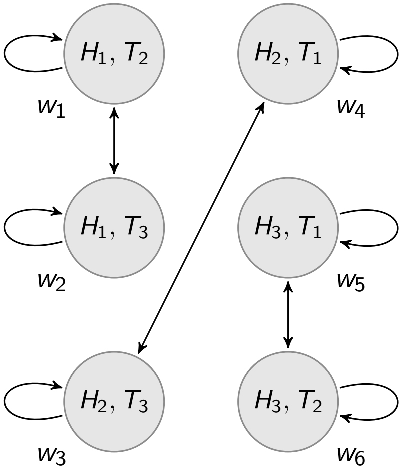


Example

Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, one of the cards is placed face down on the table and the third card is put back in the deck.

Suppose that Ann receives card 1 and card 2 is on the table.

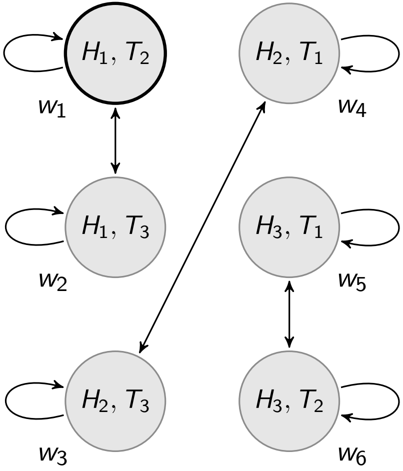


Example

Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, one of the cards is placed face down on the table and the third card is put back in the deck.

Suppose that Ann receives card 1 and card 2 is on the table.

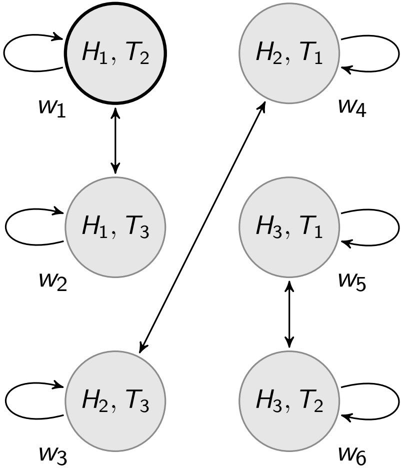


Example

Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, one of the cards is placed face down on the table and the third card is put back in the deck.

$\mathcal{M}, w_1 \models KH_1$

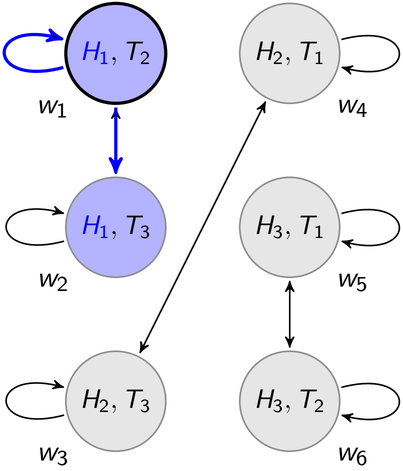


Example

Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, one of the cards is placed face down on the table and the third card is put back in the deck.

$\mathcal{M}, w_1 \models K H_1$



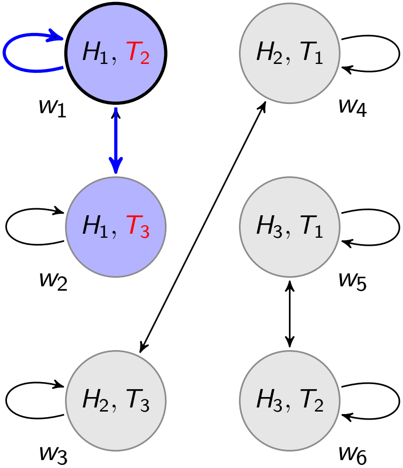
Example

Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, one of the cards is placed face down on the table and the third card is put back in the deck.

$\mathcal{M}, w_1 \models K H_1$

$\mathcal{M}, w_1 \models K \neg T_1$

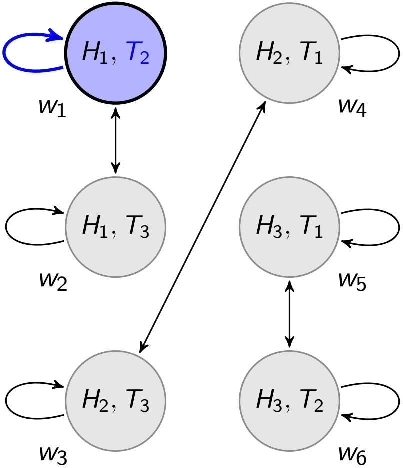


Example

Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, one of the cards is placed face down on the table and the third card is put back in the deck.

$\mathcal{M}, w_1 \models LT_2$

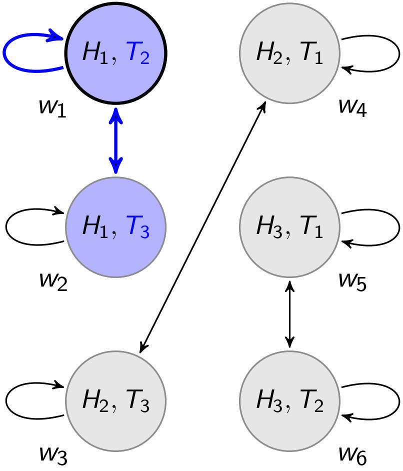


Example

Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, one of the cards is placed face down on the table and the third card is put back in the deck.

$\mathcal{M}, w_1 \models K(T_2 \vee T_3)$



The Language

: $\varphi \ := \ p \mid \neg\varphi \mid \varphi \wedge \psi \mid K\varphi$

Kripke Models: $\mathcal{M} = \langle W, R, V \rangle$ and $w \in W$

Truth: $\mathcal{M}, w \models \varphi$ is defined as follows:

- $\mathcal{M}, w \models p$ iff $w \in V(p)$ (with $p \in \text{At}$)
- $\mathcal{M}, w \models \neg\varphi$ if $\mathcal{M}, w \not\models \varphi$
- $\mathcal{M}, w \models \varphi \wedge \psi$ if $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models K\varphi$ if for each $v \in W$, if wRv, then $\mathcal{M}, v \models \varphi$

Some Questions

Should we make additional assumptions about R (i.e., reflexive, transitive, etc.)

What idealizations have we made?

Modal Formula	Property	Philosophical Assumption

Modal Formula	Property	Philosophical Assumption
$K(\varphi \to \psi) \to (K\varphi \to K\psi)$	—	Logical Omniscience

Modal Formula	Property	Philosophical Assumption
$K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$	—	Logical Omniscience
$K\varphi \rightarrow \varphi$	Reflexive	Truth

Modal Formula	Property	Philosophical Assumption
$K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$	—	Logical Omniscience
$K\varphi \rightarrow \varphi$	Reflexive	Truth
$K\varphi \rightarrow KK\varphi$	Transitive	Positive Introspection

Modal Formula	Property	Philosophical Assumption
$K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$	—	Logical Omniscience
$K\varphi \rightarrow \varphi$	Reflexive	Truth
$K\varphi \rightarrow KK\varphi$	Transitive	Positive Introspection
$\neg K\varphi \rightarrow K\neg K\varphi$	Euclidean	Negative Introspection

Modal Formula	Property	Philosophical Assumption
$K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$	—	Logical Omniscience
$K\varphi \rightarrow \varphi$	Reflexive	Truth
$K\varphi \rightarrow KK\varphi$	Transitive	Positive Introspection
$\neg K\varphi \rightarrow K\neg K\varphi$	Euclidean	Negative Introspection
$\neg K\bot$	Serial	Consistency

Multi-agent Epistemic Logic

The Language: $\varphi \ := \ p \mid \neg\varphi \mid \varphi \wedge \psi \mid K\varphi$

Kripke Models: $\mathcal{M} = \langle W, R, V \rangle$ and $w \in W$

Truth: $\mathcal{M}, w \models \varphi$ is defined as follows:

- $\mathcal{M}, w \models p$ iff $w \in V(p)$ (with $p \in \mathsf{At}$)
- $\mathcal{M}, w \models \neg\varphi$ if $\mathcal{M}, w \not\models \varphi$
- $\mathcal{M}, w \models \varphi \wedge \psi$ if $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models K\varphi$ if for each $v \in W$, if wRv, then $\mathcal{M}, v \models \varphi$

Multi-agent Epistemic Logic

The Language: $\varphi \; := \; p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_i\varphi$ with $i \in \mathcal{A}$

Kripke Models: $\mathcal{M} = \langle W, \{R_i\}_{i\in\mathcal{A}}, V\rangle$ and $w \in W$

Truth: $\mathcal{M}, w \models \varphi$ is defined as follows:

- $\mathcal{M}, w \models p$ iff $w \in V(p)$ (with $p \in \mathrm{At}$)
- $\mathcal{M}, w \models \neg\varphi$ if $\mathcal{M}, w \not\models \varphi$
- $\mathcal{M}, w \models \varphi \wedge \psi$ if $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models K_i\varphi$ if for each $v \in W$, if wR_iv, then $\mathcal{M}, v \models \varphi$

Multi-agent Epistemic Logic

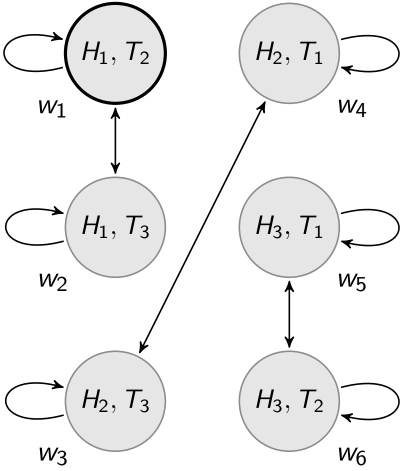
- $K_A K_B \varphi$: "Ann knows that Bob knows φ"
- $K_A(K_B\varphi \vee K_B\neg\varphi)$: "Ann knows that Bob knows whether φ
- $\neg K_B K_A K_B(\varphi)$: "Bob does not know that Ann knows that Bob knows that φ"

Example

Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, one of the cards is placed face down on the table and the third card is put back in the deck.

Suppose that Ann receives card 1 and card 2 is on the table.

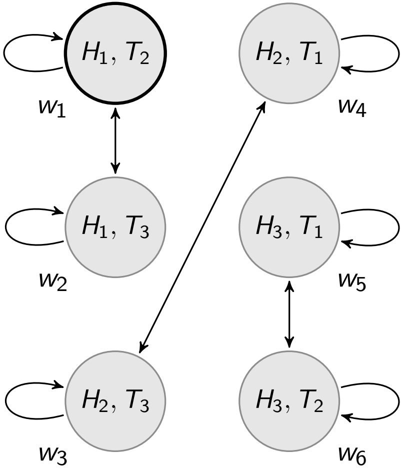


Example

Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, Bob is given one of the cards and the third card is put back in the deck.

Suppose that Ann receives card 1 and Bob receives card 2.

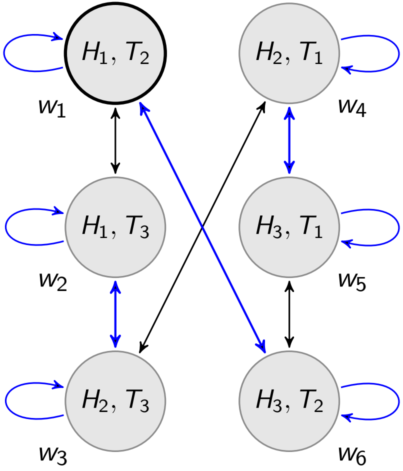


Example

Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, Bob is given one of the cards and the third card is put back in the deck.

Suppose that Ann receives card 1 and Bob receives card 2.

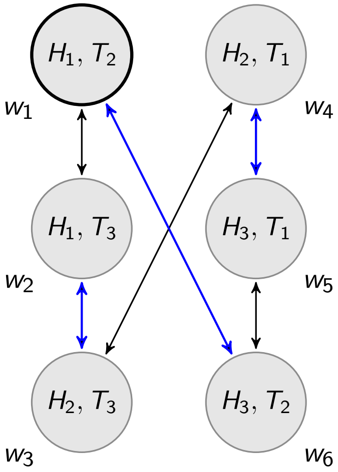


Example

Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, Bob is given one of the cards and the third card is put back in the deck.

Suppose that Ann receives card 1 and Bob receives card 2.



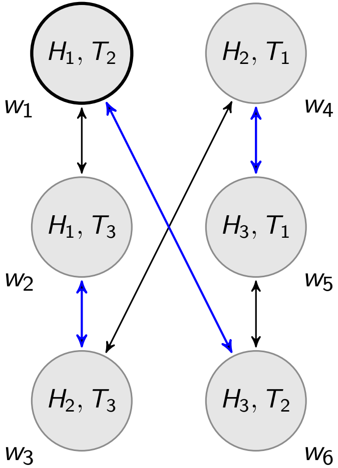
Example

Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, Bob is given one of the cards and the third card is put back in the deck.

Suppose that Ann receives card 1 and Bob receives card 2.

$\mathcal{M}, w \models K_B(K_A H_1 \vee K_A \neg H_1)$



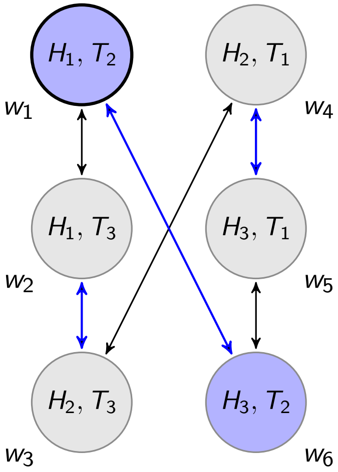
Example

Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, Bob is given one of the cards and the third card is put back in the deck.

Suppose that Ann receives card 1 and Bob receives card 2.

$\mathcal{M}, w \models K_B(K_A H_1 \vee K_A \neg H_1)$



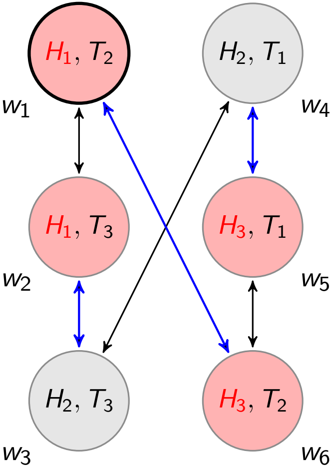
Example

Suppose there are three cards: 1, 2 and 3.

Ann is dealt one of the cards, Bob is given one of the cards and the third card is put back in the deck.

Suppose that Ann receives card 1 and Bob receives card 2.

$\mathcal{M}, w \models K_B(K_A H_1 \vee K_A \neg H_1)$



Group Knowledge

$K_A P$: "Ann knows that P"

Group Knowledge

$K_A P$: "Ann knows that P"

$K_B P$: "Bob knows that P"

Group Knowledge

$K_A P$: "Ann knows that P"

$K_B P$: "Bob knows that P"

$K_A K_B P$: "Ann knows that Bob knows that P"

Group Knowledge

$K_A P$: "Ann knows that P"

$K_B P$: "Bob knows that P"

$K_A K_B P$: "Ann knows that Bob knows that P"

$K_A P \wedge K_B P$: "Every one knows P".

Group Knowledge

$K_A P$: "Ann knows that P"

$K_B P$: "Bob knows that P"

$K_A K_B P$: "Ann knows that Bob knows that P"

$K_A P \wedge K_B P$: "Every one knows P". let $EP := K_A P \wedge K_B P$

Group Knowledge

$K_A P$: "Ann knows that P"

$K_B P$: "Bob knows that P"

$K_A K_B P$: "Ann knows that Bob knows that P"

$K_A P \wedge K_B P$: "Every one knows P". let $EP := K_A P \wedge K_B P$

$K_A EP$: "Ann knows that everyone knows that P".

Group Knowledge

$K_A P$: "Ann knows that P"

$K_B P$: "Bob knows that P"

$K_A K_B P$: "Ann knows that Bob knows that P"

$K_A P \wedge K_B P$: "Every one knows P". let $EP := K_A P \wedge K_B P$

$K_A EP$: "Ann knows that everyone knows that P".

EEP: "Everyone knows that everyone knows that P".

Group Knowledge

$K_A P$: "Ann knows that P"

$K_B P$: "Bob knows that P"

$K_A K_B P$: "Ann knows that Bob knows that P"

$K_A P \wedge K_B P$: "Every one knows P". let $EP := K_A P \wedge K_B P$

$K_A EP$: "Ann knows that everyone knows that P".

EEP: "Everyone knows that everyone knows that P".

$EEEP$: "Everyone knows that everyone knows that everyone knows that P."

Common Knowledge

CP: "It is common knowledge that P"

Common Knowledge

CP: "It is common knowledge that P" — "Everyone knows that everyone knows that everyone knows that $\cdots$ P".

Common Knowledge

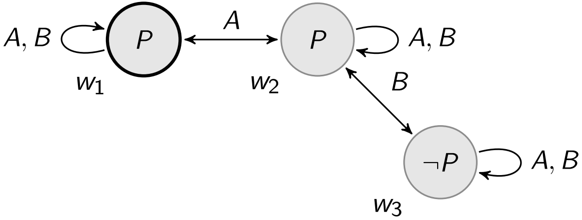
CP: "It is common knowledge that *P*" — "Everyone knows that everyone knows that everyone knows that $\cdots$ *P*".

Is *common knowledge* different from *everyone knows*?

Common Knowledge

CP: "It is common knowledge that P" — "Everyone knows that everyone knows that everyone knows that $\cdots$ P".

Is *common knowledge* different from *everyone knows*?

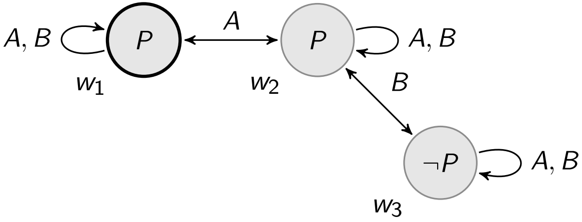


$$w_1 \models EP \wedge \neg CP$$

Common Knowledge

CP: "It is common knowledge that P" — "Everyone knows that everyone knows that everyone knows that $\cdots$ P".

Is *common knowledge* different from *everyone knows*?

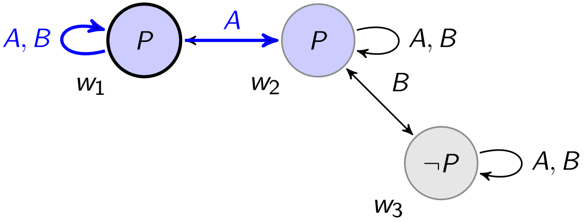


$$w_1 \models EP \wedge \neg CP$$

Common Knowledge

CP: "It is common knowledge that P" — "Everyone knows that everyone knows that everyone knows that $\cdots$ P".

Is *common knowledge* different from *everyone knows*?

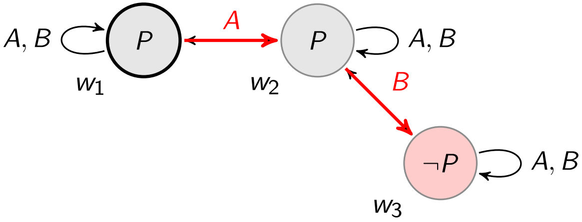


$$w_1 \models EP \wedge \neg CP$$

Common Knowledge

CP: "It is common knowledge that P" — "Everyone knows that everyone knows that everyone knows that $\cdots$ P".

Is *common knowledge* different from *everyone knows*?



$$w_1 \models EP \wedge \neg CP$$

Common Knowledge

The operator "everyone knows P", denoted EP, is defined as follows

$$EP \quad := \quad \bigwedge_{i \in \mathcal{A}} K_i P$$

$w \models CP$ iff every finite path starting at w ends with a state satisfying P.

$$CP \longrightarrow ECP$$

$$CP \rightarrow ECP$$

Suppose you are told "Ann and Bob are going together,"' and respond "sure, that's common knowledge." What you mean is not only that everyone knows this, but also that the announcement is pointless, occasions no surprise, reveals nothing new; in effect, that the situation after the announcement does not differ from that before. ...the event "Ann and Bob are going together" — call it P — is common knowledge if and only if some event — call it Q — happened that entails P and also entails all players' knowing Q (like all players met Ann and Bob at an intimate party). *(Robert Aumann)*

$$P \wedge C(P \rightarrow EP) \rightarrow CP$$

An Example

Two players Ann and Bob are told that the following will happen. Some positive integer n will be chosen and *one* of n, $n+1$ will be written on Ann's forehead, the other on Bob's. Each will be able to see the other's forehead, but not his/her own.

An Example

Two players Ann and Bob are told that the following will happen. Some positive integer n will be chosen and *one* of n, $n+1$ will be written on Ann's forehead, the other on Bob's. Each will be able to see the other's forehead, but not his/her own.

Suppose the number are (2,3).

An Example

Two players Ann and Bob are told that the following will happen. Some positive integer n will be chosen and *one* of n, $n+1$ will be written on Ann's forehead, the other on Bob's. Each will be able to see the other's forehead, but not his/her own.

Suppose the number are (2,3).

Do the agents know there numbers are less than 1000?

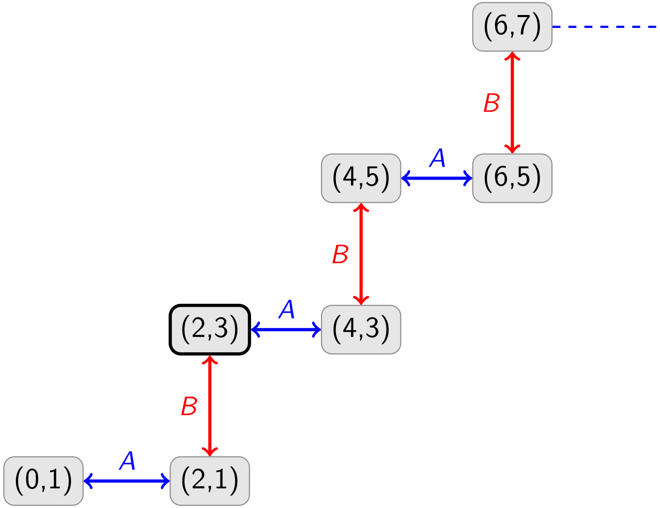
An Example

Two players Ann and Bob are told that the following will happen. Some positive integer n will be chosen and *one* of n, $n+1$ will be written on Ann's forehead, the other on Bob's. Each will be able to see the other's forehead, but not his/her own.

Suppose the number are (2,3).

Do the agents know there numbers are less than 1000?

Is it common knowledge that their numbers are less than 1000?


(6,7)
B
(4,5)
A
(6,5)
B
(2,3)
A
(4,3)
B
(0,1)
A
(2,1)

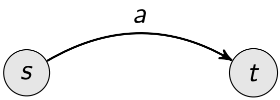
Basic Ingredients

- ✓ informational attitudes (knowledge, group knowledge, belief, certainty, etc.)
- time, actions and ability
- motivational attitudes

Actions: Two Views

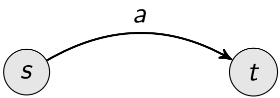
Actions: Two Views

1. Actions *transition between states, or situations*

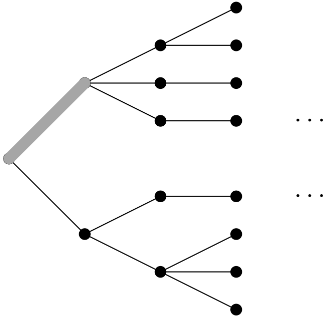


Actions: Two Views

1. Actions *transition between states, or situations*



2. Actions restrict the set of *possible future histories*



Propositional Dynamic Logic

Semantics: $\mathcal{M} = \langle W, \{R_a \mid a \in \mathrm{P}\}, V\rangle$ where for each $a \in \mathrm{P}$, $R_a \subseteq W \times W$ and $V : \mathrm{At} \to \wp(W)$

- $R_{\alpha\cup\beta} := R_\alpha \cup R_\beta$
- $R_{\alpha;\beta} := R_\alpha \circ R_\beta$
- $R_{\alpha^*} := \cup_{n\geq 0} R_\alpha^n$
- $R_{\varphi?} = \{(w, w) \mid \mathcal{M}, w \models \varphi\}$

$$\mathcal{M}, w \models [\alpha]\varphi \text{ iff for each } v\text{, if } wR_\alpha v \text{ then } \mathcal{M}, v \models \varphi$$

Background: Propositional Dynamic Logic

1. Axioms of propositional logic
2. $[\alpha](\varphi \rightarrow \psi) \rightarrow ([\alpha]\varphi \rightarrow [\alpha]\psi)$
3. $[\alpha \cup \beta]\varphi \leftrightarrow [\alpha]\varphi \wedge [\beta]\varphi$
4. $[\alpha; \beta]\varphi \leftrightarrow [\alpha][\beta]\varphi$
5. $[\psi?]\varphi \leftrightarrow (\psi \rightarrow \varphi)$
6. $\varphi \wedge [\alpha][\alpha^*]\varphi \leftrightarrow [\alpha^*]\varphi$
7. $\varphi \wedge [\alpha^*](\varphi \rightarrow [\alpha]\varphi) \rightarrow [\alpha^*]\varphi$
8. Modus Ponens and Necessitation (for each program α)

Background: Propositional Dynamic Logic

1. Axioms of propositional logic
2. $[\alpha](\varphi \rightarrow \psi) \rightarrow ([\alpha]\varphi \rightarrow [\alpha]\psi)$
3. $[\alpha \cup \beta]\varphi \leftrightarrow [\alpha]\varphi \wedge [\beta]\varphi$
4. $[\alpha; \beta]\varphi \leftrightarrow [\alpha][\beta]\varphi$
5. $[\psi?]\varphi \leftrightarrow (\psi \rightarrow \varphi)$
6. $\varphi \wedge [\alpha][\alpha^*]\varphi \leftrightarrow [\alpha^*]\varphi$ (Fixed-Point Axiom)
7. $\varphi \wedge [\alpha^*](\varphi \rightarrow [\alpha]\varphi) \rightarrow [\alpha^*]\varphi$ (Induction Axiom)
8. Modus Ponens and Necessitation (for each program α)

Propositional Dynamic Logic

Theorem PDL is sound and weakly complete with respect to the Segerberg Axioms.

Theorem The satisfiability problem for **PDL** is decidable (EXPTIME-Complete).

D. Kozen and R. Parikh. *A Completeness proof for Propositional Dynamic Logic*. .

D. Harel, D. Kozen and Tiuryn. *Dynamic Logic*. 2001.

Actions and Ability

An early approach to interpret PDL as logic of actions was put forward by Krister Segerberg.

Segerberg adds an "agency" program to the PDL language δA where A is a formula.

K. Segerberg. *Bringing it about*. JPL, 1989.

Actions and Agency

The intended meaning of the program 'δA' is that the agent "brings it about that A': *formally*, δA is the set of all paths p such that

Actions and Agency

The intended meaning of the program 'δA' is that the agent "brings it about that A': *formally*, δA is the set of all paths p such that

1. p is the computation according to some program α, and
2. α only terminates at states in which it is true that A

Actions and Agency

The intended meaning of the program 'δA' is that the agent "brings it about that A': *formally*, δA is the set of all paths p such that

1. p is the computation according to some program α, and
2. α only terminates at states in which it is true that A

Interestingly, Segerberg also briefly considers a third condition:

3. p is optimal (in some sense: shortest, maximally efficient, most convenient, etc.) in the set of computations satisfying conditions (1) and (2).

Actions and Agency

The intended meaning of the program 'δA' is that the agent "brings it about that A': *formally*, δA is the set of all paths p such that

1. p is the computation according to some program α, and
2. α only terminates at states in which it is true that A

Interestingly, Segerberg also briefly considers a third condition:

3. p is optimal (in some sense: shortest, maximally efficient, most convenient, etc.) in the set of computations satisfying conditions (1) and (2).

The axioms:

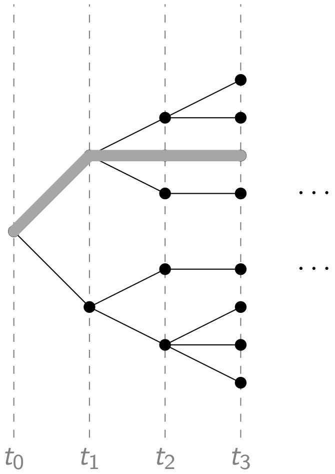
1. $[\delta A]A$
2. $[\delta A]B \rightarrow ([\delta B]C \rightarrow [\delta A]C)$

Actions and Agency

J. Horty. *Agency and Deontic Logic*. 2001.

Logics of Action and Agency

Alternative accounts of agency do not include explicit description of the actions:



STIT

- Each node represents a choice point for the agent.
- A **history** is a maximal branch in the above tree.
- Formulas are interpreted at history moment pairs.
- At each moment there is a choice available to the agent (partition of the histories through that moment)
- The key modality is $[stit]\varphi$ which is intended to mean that the agent i can "see to it that φ is true".
  - $[stit]\varphi$ is true at a history moment pair provided the agent can choose a (set of) branch(es) such that every future history-moment pair satisfies φ

STIT

We use the modality '$\Diamond$' to mean historic possibility.

$\Diamond[stit]\varphi$: "the agent has the ability to bring about φ".

STIT

We use the modality '$\Diamond$' to mean historic possibility.

$\Diamond[\mathit{stit}]\varphi$: "the agent has the ability to bring about φ".

Example Consider the example of an agent (call her Ann) throwing a dart. Suppose Ann is not a very good dart player, but she just happens to throw a bull's eye.

STIT

We use the modality '$\Diamond$' to mean historic possibility.

$\Diamond[\mathit{stit}]\varphi$: "the agent has the ability to bring about φ".

Example Consider the example of an agent (call her Ann) throwing a dart. Suppose Ann is not a very good dart player, but she just happens to throw a bull's eye. Intuitively, we do not want to say that Ann has the *ability* to throw a bull's eye even though it happens to be true. That is, the following principle should be falsifiable:

$$\varphi \rightarrow \Diamond[\mathit{stit}]\varphi$$

STIT

Example Continuing with this example, suppose that Ann has the ability to hit the dart board, but has no other control over the placement of the dart. Now, when she throws the dart, as a matter of fact, it will either hit the top half of the board or the bottom half of the board. Since, Ann has the ability to hit the dart board, she has the ability to either hit the top half of the board or the bottom half of the board.

STIT

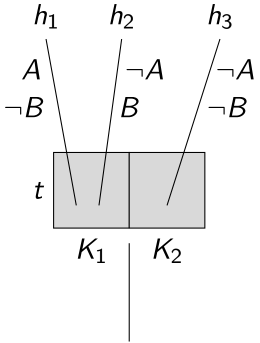
Example Continuing with this example, suppose that Ann has the ability to hit the dart board, but has no other control over the placement of the dart. Now, when she throws the dart, as a matter of fact, it will either hit the top half of the board or the bottom half of the board. Since, Ann has the ability to hit the dart board, she has the ability to either hit the top half of the board or the bottom half of the board.

However, intuitively, it seems true that Ann does not have the ability to hit the top half of the dart board, and also she does not have the ability to hit the bottom half of the dart board. Thus, the following principle should be falsifiable:

$$\Diamond[\mathit{stit}](\varphi \vee \psi) \rightarrow \Diamond[\mathit{stit}]\varphi \vee \Diamond[\mathit{stit}]\psi$$

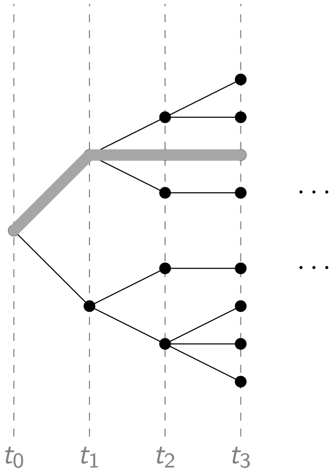
STIT

The following model will falsify both of the above formulas:

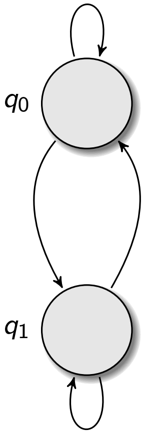


J. Horty. *Agency and Deontic Logic*. 2001.

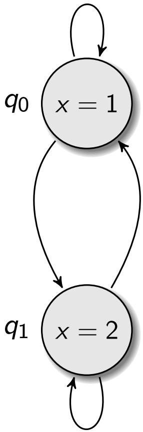
Temporal Logics



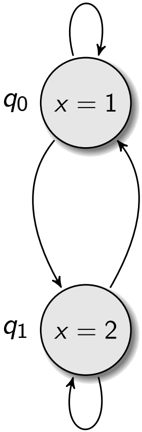
Computational vs. Behavioral Structures

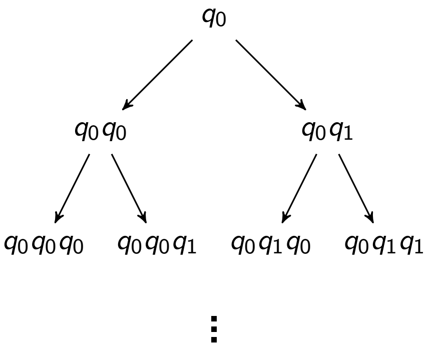


Computational vs. Behavioral Structures



Computational vs. Behavioral Structures





Temporal Logics

Temporal Logics

- *Linear Time Temporal Logic*: Reasoning about computation paths:

  $\Diamond\varphi$: φ is true some time in *the* future.

A. Pnuelli. *A Temporal Logic of Programs*. in *Proc. 18th IEEE Symposium on Foundations of Computer Science* (1977).

Temporal Logics

- *Linear Time Temporal Logic*: Reasoning about computation paths:

  $\Diamond\varphi$: φ is true some time in *the* future.

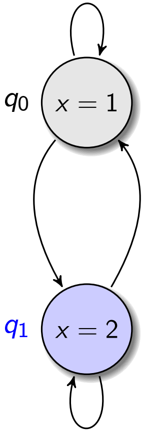
A. Pnuelli. *A Temporal Logic of Programs*. in *Proc. 18th IEEE Symposium on Foundations of Computer Science* (1977).

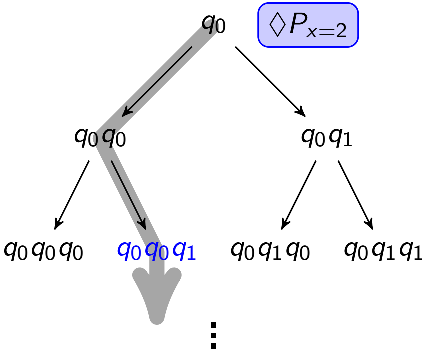
- *Branching Time Temporal Logic*: Allows quantification over paths:

  $\exists\Diamond\varphi$: there is a path in which φ is eventually true.

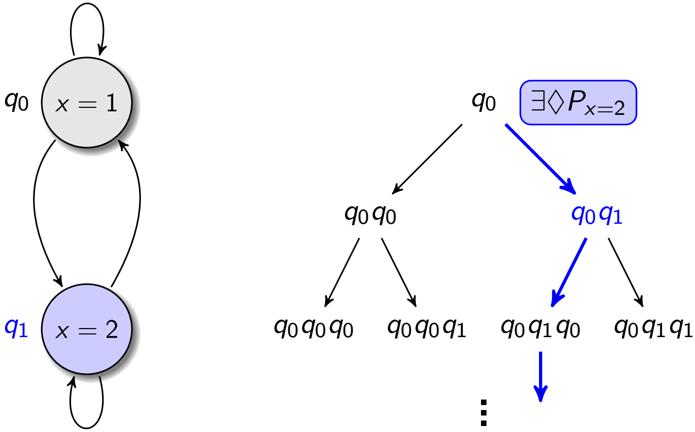
E. M. Clarke and E. A. Emerson. *Design and Synthesis of Synchronization Skeletons using Branching-time Temproal-logic Specifications*. In *Proceedings Workshop on Logic of Programs*, LNCS (1981).

Temporal Logics

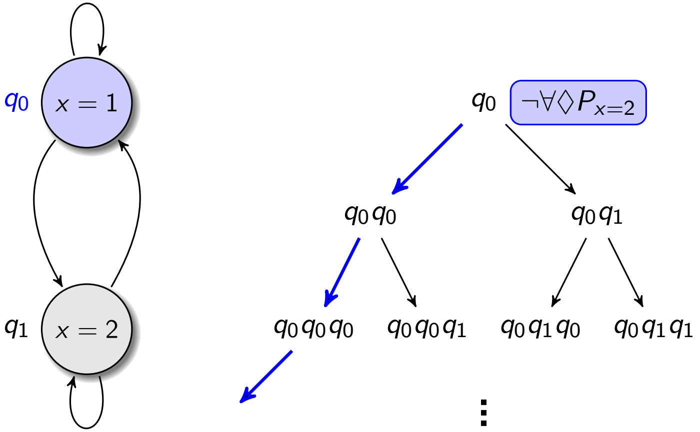




Temporal Logics



Temporal Logics



Many Agents

The previous model assumes there is *one* agent that “controls” the transition system.

Many Agents

The previous model assumes there is *one* agent that "controls" the transition system.

What if there is more than one agent?

Many Agents

The previous model assumes there is *one* agent that "controls" the transition system.

What if there is more than one agent?

Example: Suppose that there are two agents: a server (s) and a client (c). The client asks to set the value of x and the server can either grant or deny the request. Assume the agents make simultaneous moves.

Many Agents

The previous model assumes there is *one* agent that "controls" the transition system.

What if there is more than one agent?

Example: Suppose that there are two agents: a server (s) and a client (c). The client asks to set the value of x and the server can either grant or deny the request. Assume the agents make simultaneous moves.

	deny	*grant*
*set*1		
*set*2		

Many Agents

The previous model assumes there is *one* agent that "controls" the transition system.

What if there is more than one agent?

Example: Suppose that there are two agents: a server (s) and a client (c). The client asks to set the value of x and the server can either grant or deny the request. Assume the agents make simultaneous moves.

	deny	*grant*
*set*1		$q_0 \Rightarrow q_0,\ q_1 \Rightarrow q_0$
*set*2		$q_0 \Rightarrow q_1,\ q_1 \Rightarrow q_1$

Many Agents

The previous model assumes there is *one* agent that "controls" the transition system.

What if there is more than one agent?

Example: Suppose that there are two agents: a server (s) and a client (c). The client asks to set the value of x and the server can either grant or deny the request. Assume the agents make simultaneous moves.

	deny	*grant*
*set*1	$q \Rightarrow q$	$q_0 \Rightarrow q_0,\ q_1 \Rightarrow q_0$
*set*2	$q \Rightarrow q$	$q_0 \Rightarrow q_1,\ q_1 \Rightarrow q_1$

From Temporal Logic to Strategy Logic

- *Coalitional Logic*: Reasoning about (local) group power.

  $[C]\varphi$: coalition C has a **joint action** to bring about φ.

M. Pauly. *A Modal Logic for Coalition Powers in Games*. *Journal of Logic and Computation* **12** (2002).

From Temporal Logic to Strategy Logic

- *Coalitional Logic*: Reasoning about (local) group power.

  $[C]\varphi$: coalition C has a **joint action** to bring about φ.

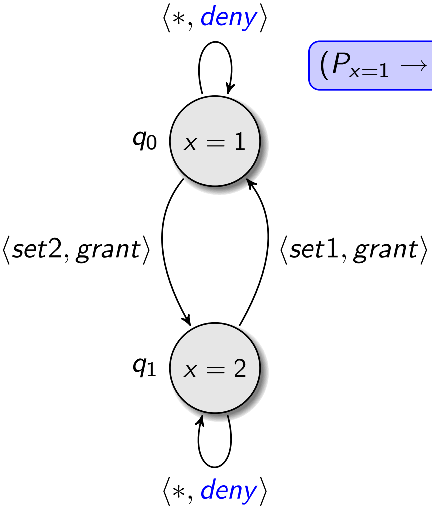
M. Pauly. *A Modal Logic for Coalition Powers in Games*. *Journal of Logic and Computation* **12** (2002).

- *Alternating-time Temporal Logic*: Reasoning about (local and global) group power:

  $\langle\langle A\rangle\rangle\Box\varphi$: The coalition A has a **joint action** to ensure that φ will remain true.

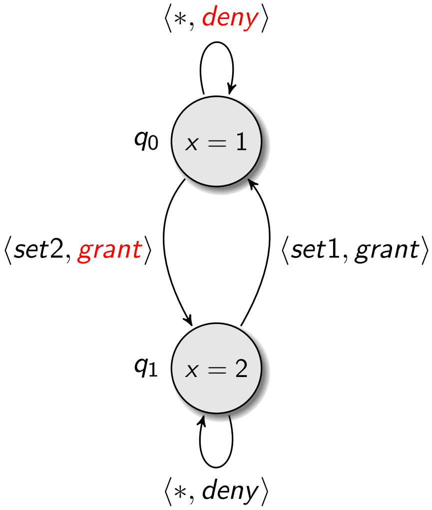
R. Alur, T. Henzinger and O. Kupferman. *Alternating-time Temproal Logic*. *Jouranl of the ACM* (2002).

Multi-agent Transition Systems



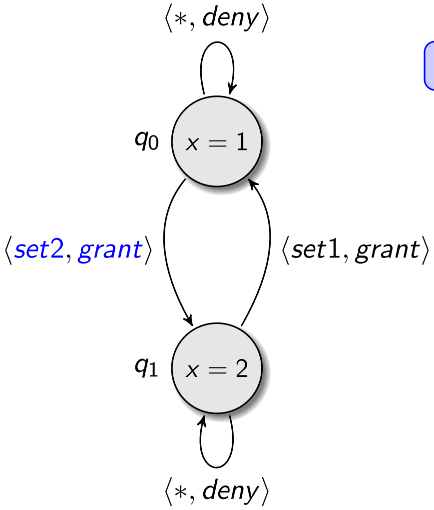
$$(P_{x=1} \rightarrow [s]P_{x=1}) \wedge (P_{x=2} \rightarrow [s]P_{x=2})$$

Multi-agent Transition Systems



$P_{x=1} \rightarrow \neg[s]P_{x=2}$

Multi-agent Transition Systems



$P_{x=1} \longrightarrow \neg[s, c]P_{x=2}$

Basic Ingredients

- ✓ informational attitudes (knowledge, group knowledge, belief, certainty, etc.)
- ✓ time, actions and ability (individual and coalitional ability)
- motivational attitudes

Preference (Modal) Logics

x, y objects

$x \succeq y$: x is at least as good as y

Preference (Modal) Logics

x, y objects

$x \succeq y$: x is at least as good as y

1. $x \succeq y$ and $y \not\succeq x$ ($x \succ y$)
2. $x \not\succeq y$ and $y \succeq x$ ($y \succ x$)
3. $x \succeq y$ and $y \succeq x$ ($x \sim y$)
4. $x \not\succeq y$ and $y \not\succeq x$ ($x \perp y$)

Preference (Modal) Logics

x, y objects

$x \succeq y$: x is at least as good as y

1. $x \succeq y$ and $y \not\succeq x$ ($x \succ y$)
2. $x \not\succeq y$ and $y \succeq x$ ($y \succ x$)
3. $x \succeq y$ and $y \succeq x$ ($x \sim y$)
4. $x \not\succeq y$ and $y \not\succeq x$ ($x \perp y$)

Properties: transitivity, connectedness, etc.

Preference (Modal) Logics

Modal betterness model $\mathcal{M} = \langle W, \succeq, V \rangle$

Preference (Modal) Logics

Modal betterness model $\mathcal{M} = \langle W, \succeq, V \rangle$

Preference Modalities $\langle \succeq \rangle \varphi$: "there is a world at least as good (as the current world) satisfying φ"

$\mathcal{M}, w \models \langle \succeq \rangle \varphi$ iff there is a $v \succeq w$ such that $\mathcal{M}, v \models \varphi$

Preference (Modal) Logics

Modal betterness model $\mathcal{M} = \langle W, \succeq, V\rangle$

Preference Modalities $\langle\succeq\rangle\varphi$: "there is a world at least as good (as the current world) satisfying φ"

$\mathcal{M}, w \models \langle\succeq\rangle\varphi$ iff there is a $v \succeq w$ such that $\mathcal{M}, v \models \varphi$

$\mathcal{M}, w \models \langle\succ\rangle\varphi$ iff there is $v \succeq w$ and $w \not\succeq v$ such that $\mathcal{M}, v \models \varphi$

Preference (Modal) Logics

1. $\langle\succ\rangle\varphi \rightarrow \langle\succeq\rangle\varphi$
2. $\langle\succeq\rangle\langle\succ\rangle\varphi \rightarrow \langle\succ\rangle\varphi$
3. $\varphi \wedge \langle\succeq\rangle\psi \rightarrow (\langle\succ\rangle\psi \vee \langle\succeq\rangle(\psi \wedge \langle\succeq\rangle\varphi))$
4. $\langle\succ\rangle\langle\succeq\rangle\varphi \rightarrow \langle\succ\rangle\varphi$

Theorem The above logic (with Necessitation and Modus Ponens) is sound and complete with respect to the class of preference models.

J. van Benthem, O. Roy and P. Girard. *Everything else being equal: A modal logic approach to* ceteris paribus *preferences*. JPL, 2008.

Preference Modalities

$\varphi \geq \psi$: the state of affairs φ is at least as good as ψ (ceteris paribus)

G. von Wright. *The logic of preference*. Edinburgh University Press (1963).

From worlds to sets and back

Lifting

- $X \geq_{\forall\exists} Y$ if $\forall y \in Y\ \exists x \in X$: $x \succeq y$

From worlds to sets and back

Lifting

- $X \geq_{\forall\exists} Y$ if $\forall y \in Y\ \exists x \in X$: $x \succeq y$
  $A(\varphi \rightarrow \langle\succeq\rangle\psi)$

From worlds to sets and back

Lifting

- $X \geq_{\forall\exists} Y$ if $\forall y \in Y\ \exists x \in X$: $x \succeq y$
  $A(\varphi \rightarrow \langle\succeq\rangle\psi)$
- $X \geq_{\forall\forall} Y$ if $\forall y \in Y\ \forall x \in X$: $x \succeq y$
  $A(\varphi \rightarrow [\succ]\neg\psi)$

From worlds to sets and back

Lifting

- $X \geq_{\forall\exists} Y$ if $\forall y \in Y\ \exists x \in X$: $x \succeq y$
  $A(\varphi \rightarrow \langle\succeq\rangle\psi)$
- $X \geq_{\forall\forall} Y$ if $\forall y \in Y\ \forall x \in X$: $x \succeq y$
  $A(\varphi \rightarrow [\succ]\neg\psi)$

Deriving

$P_1 \gg P_2 \gg P_3 \gg \cdots \gg P_n$

$x > y$ iff x and y differ in at least one P_i and the first P_i where this happens is one with $P_i x$ and $\neg P_i y$

F. Liu and D. De Jongh. *Optimality, belief and preference*. 2006.

The Logic of Group Decisions

The Logic of Group Decisions

Fundamental Problem: groups are inconsistent!

The Logic of Group Decisions: The Doctrinal "Paradox" (Kornhauser and Sager 1993)

p: a valid contract was in place

q: there was a breach of contract

r: the court is required to find the defendant liable.

	p	q	$(p \wedge q) \leftrightarrow r$	r
1	yes	yes	yes	yes
2	yes	no	yes	no
3	no	yes	yes	no

The Logic of Group Decisions: The Doctrinal “Paradox” (Kornhauser and Sager 1993)

Should we accept r?

	p	q	$(p \wedge q) \leftrightarrow r$	r
1	yes	yes	yes	yes
2	yes	no	yes	no
3	no	yes	yes	no

The Logic of Group Decisions: The Doctrinal "Paradox" (Kornhauser and Sager 1993)

Should we accept r? No, a simple majority votes no.

	p	q	$(p \wedge q) \leftrightarrow r$	r
1	yes	yes	yes	yes
2	yes	no	yes	no
3	no	yes	yes	no

The Logic of Group Decisions: The Doctrinal "Paradox" (Kornhauser and Sager 1993)

Should we accept r? Yes, a majority votes yes for p and q and $(p \wedge q) \leftrightarrow r$ is a legal doctrine.

	p	q	$(p \wedge q) \leftrightarrow r$	r
1	yes	yes	yes	yes
2	yes	no	yes	no
3	no	yes	yes	no

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"
$a \rightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"

$a \rightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"

b "There will be global warming"

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"

$a \rightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"

b "There will be global warming"

	a	$a \rightarrow b$	b

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"

$a \rightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"

b "There will be global warming"

	a	$a \rightarrow b$	b
1	True	True	

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"
$a \rightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"
b "There will be global warming"

	a	$a \rightarrow b$	b
1	True	True	True

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"

$a \rightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"

b "There will be global warming"

	a	$a \rightarrow b$	b
1	True	True	True
2	True	False	

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"
$a \to b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"
b "There will be global warming"

	a	$a \to b$	b
1	True	True	True
2	True	False	False

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"

$a \rightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"

b "There will be global warming"

	a	$a \rightarrow b$	b
1	True	True	True
2	True	False	False
3	False	True	

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"

$a \rightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"

b "There will be global warming"

	a	$a \rightarrow b$	b
1	True	True	True
2	True	False	False
3	False	True	False
Majority			

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"
$a \longrightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"
b "There will be global warming"

	a	$a \longrightarrow b$	b
1	True	True	True
2	True	False	False
3	False	True	False
Majority	True		

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"

$a \rightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"

b "There will be global warming"

	a	$a \rightarrow b$	b
1	True	True	True
2	True	False	False
3	False	True	False
Majority	True	True	

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"

$a \rightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"

b "There will be global warming"

	a	$a \rightarrow b$	b
1	True	True	True
2	True	False	False
3	False	True	False
Majority	True	True	False

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"
$a \longrightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"
b "There will be global warming"

	a	$a \longrightarrow b$	b
1	True	True	True
2	True	False	False
3	False	True	False
Majority	True	True	False

Conclusion: Groups are inconsistent, difference between 'premise-based' and 'conclusion-based' decision making, ...

Group Preference Logics

H. Andréka, M. Ryan and P Yves Schobbens. *Operators and laws for combining preference relations*. Journal of Logic and Computation, 2002.

P. Girard. *Modal Logic for Lexicographic Preference Aggregation*. Manuscript, 2008.

Basic Ingredients

- ✓ informational attitudes (knowledge, group knowledge, belief, certainty, etc.)
- ✓ time, actions and ability (individual and coalitional ability)
- ✓ motivational attitudes (individual preferences, group preferences)

General Issues

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

General Issues

How should we *compare* the different logical systems?

- Embedding one logic in another:

General Issues

How should we *compare* the different logical systems?

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

General Issues

How should we *compare* the different logical systems?

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

General Issues

How should we *compare* the different logical systems?

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

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

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

General Issues

How should we *compare* the different logical systems?

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

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

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

- Comparing different frameworks:

General Issues

How should we *compare* the different logical systems?

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

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

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

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

General Issues

How should we *merge* the different logical systems?

General Issues

How should we *merge* the different logical systems?

- Combining logics is hard!

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

General Issues

How should we *merge* the different logical systems?

- Combining logics is hard!

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

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

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

Merging logics of rational agency

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

Merging logics of rational agency

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

Conclusions

Example

Ann would like Bob to attend her talk; however, she only wants Bob to attend if he is interested in the subject of her talk, not because he is just being polite.

There is a very simple procedure to solve Ann's problem: *have a (trusted) friend tell Bob the time and subject of her talk.*

Is this procedure correct?

Example

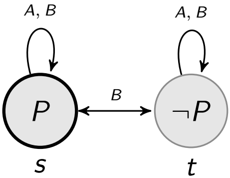
Ann would like Bob to attend her talk; however, she only wants Bob to attend if he is interested in the subject of her talk, not because he is just being polite.

There is a very simple procedure to solve Ann's problem: *have a (trusted) friend tell Bob the time and subject of her talk.*

Is this procedure correct? Yes, if

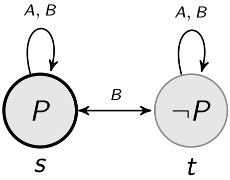
1. Ann knows about the talk.
2. Bob knows about the talk.
3. Ann knows that Bob knows about the talk.
4. Bob *does not* know that Ann knows that he knows about the talk.
5. *And nothing else.*

Example



P means "The talk is at 2PM".

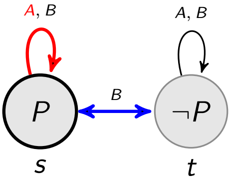
Example



P means "The talk is at 2PM".

$\mathcal{M}, s \models K_A P \wedge \neg K_B P$

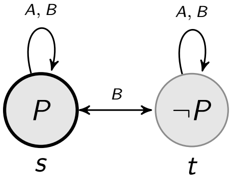
Example

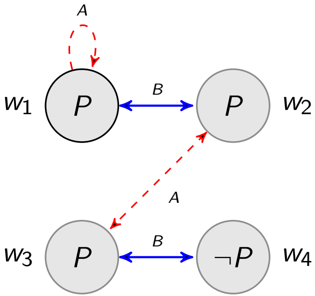


P means "The talk is at 2PM".

$\mathcal{M}, s \models K_A P \wedge \neg K_B P$

Example





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.

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.

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.

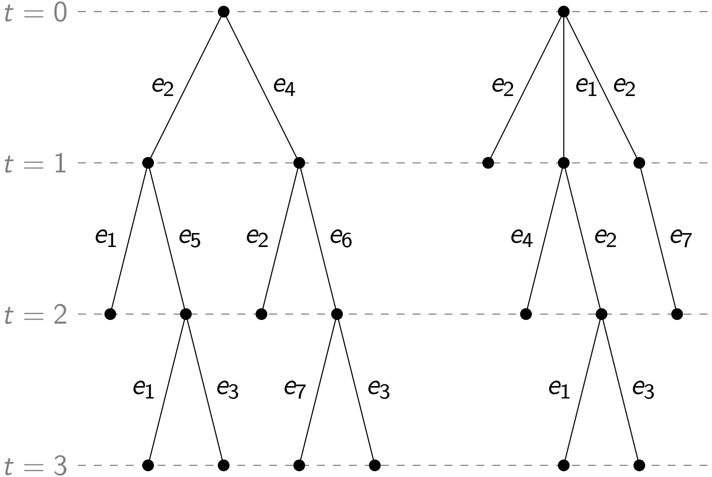
Dynamic Epistemic Logic

Epistemic Temporal Logic

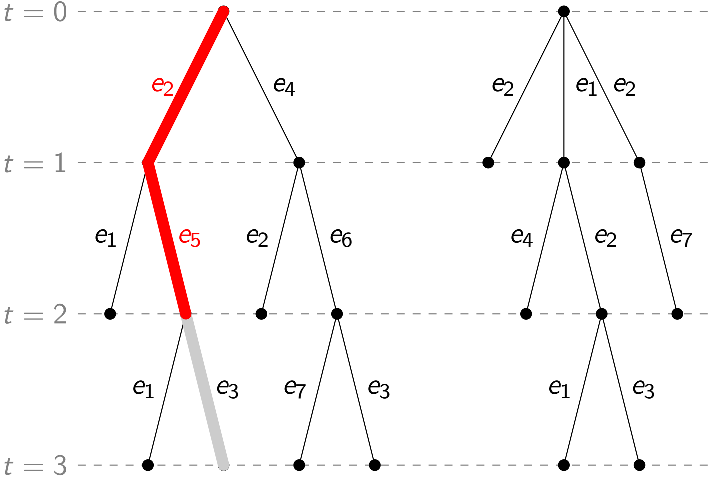
R. Parikh and R. Ramanujam. *A Knowledge Based Semantics of Messages*. *Journal of Logic, Language and Information*, 12: 453 – 467, 1985, 2003.

FHMV. Reasoning about Knowledge. MIT Press, 1995.

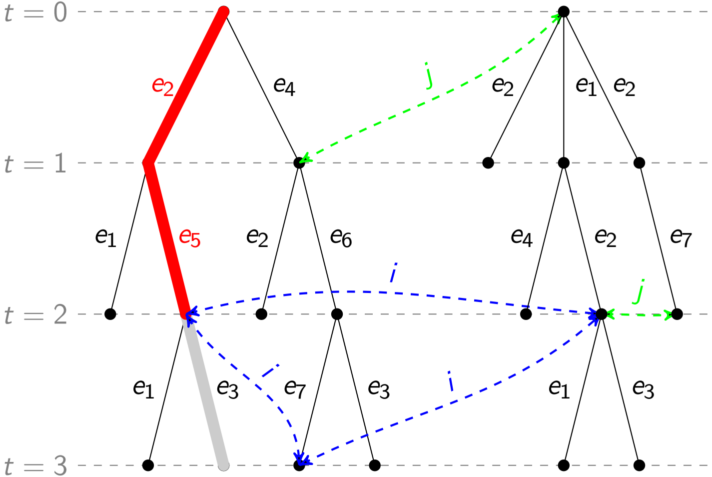
The 'Playground'



The ‘Playground’



The ‘Playground’



Formal Languages

- $P\varphi$ (φ is true *sometime* in the past),
- $F\varphi$ (φ is true *sometime* in the future),
- $Y\varphi$ (φ is true at *the* previous moment),
- $N\varphi$ (φ is true at *the* next moment),
- $N_e\varphi$ (φ is true after event e)
- $K_i\varphi$ (agent i knows φ) and
- $C_B\varphi$ (the group $B \subseteq \mathcal{A}$ commonly knows φ).

History-based Models

An ETL **model** is a structure $\langle \mathcal{H}, \{\sim_i\}_{i\in\mathcal{A}}, V\rangle$ where $\langle \mathcal{H}, \{\sim_i\}_{i\in\mathcal{A}}\rangle$ is an ETL frame and

$V : \mathsf{At} \rightarrow 2^{\mathsf{finite}(\mathcal{H})}$ is a valuation function.

Formulas are interpreted at pairs H, t:

$$H, t \models \varphi$$

Truth in a Model

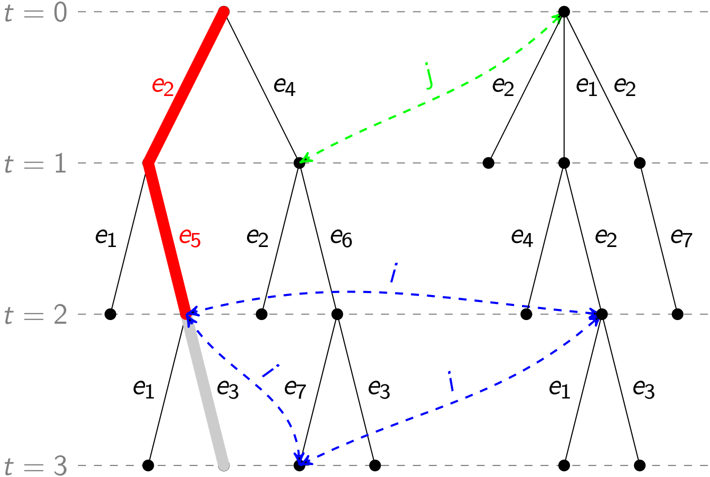
- $H, t \models P\varphi$ iff there exists $t' \leq t$ such that $H, t' \models \varphi$
- $H, t \models F\varphi$ iff there exists $t' \geq t$ such that $H, t' \models \varphi$
- $H, t \models N\varphi$ iff $H, t+1 \models \varphi$
- $H, t \models Y\varphi$ iff $t > 1$ and $H, t-1 \models \varphi$
- $H, t \models K_i\varphi$ iff for each $H' \in \mathcal{H}$ and $m \geq 0$ if $H_t \sim_i H'_m$ then $H', m \models \varphi$
- $H, t \models C\varphi$ iff for each $H' \in \mathcal{H}$ and $m \geq 0$ if $H_t \sim_* H'_m$ then $H', m \models \varphi$.

where $\sim_*$ is the reflexive transitive closure of the union of the $\sim_i$.

Truth in a Model

- $H, t \models P\varphi$ iff there exists $t' \leq t$ such that $H, t' \models \varphi$
- $H, t \models F\varphi$ iff there exists $t' \geq t$ such that $H, t' \models \varphi$
- $H, t \models N\varphi$ iff $H, t+1 \models \varphi$
- $H, t \models Y\varphi$ iff $t > 1$ and $H, t-1 \models \varphi$
- $H, t \models K_i\varphi$ iff for each $H' \in \mathcal{H}$ and $m \geq 0$ if $H_t \sim_i H'_m$ then $H', m \models \varphi$
- $H, t \models C\varphi$ iff for each $H' \in \mathcal{H}$ and $m \geq 0$ if $H_t \sim_* H'_m$ then $H', m \models \varphi$.

where $\sim_*$ is the reflexive transitive closure of the union of the $\sim_i$.


t = 0
t = 1
t = 2
t = 3
e2
e4
j
e2
e1
e2
e1
e5
e2
e6
e4
e2
e7
i
e1
e3
e7
e3
i
e1
e3

Returning to the Example

Ann would like Bob to attend her talk; however, she only wants Bob to attend if he is interested in the subject of her talk, not because he is just being polite.

Returning to the Example

Ann would like Bob to attend her talk; however, she only wants Bob to attend if he is interested in the subject of her talk, not because he is just being polite.

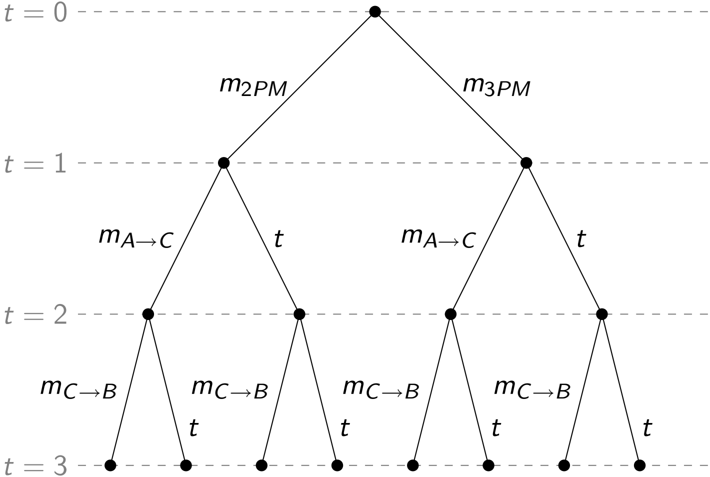
There is a very simple procedure to solve Ann's problem: *have a (trusted) friend tell Bob the time and subject of her talk.*

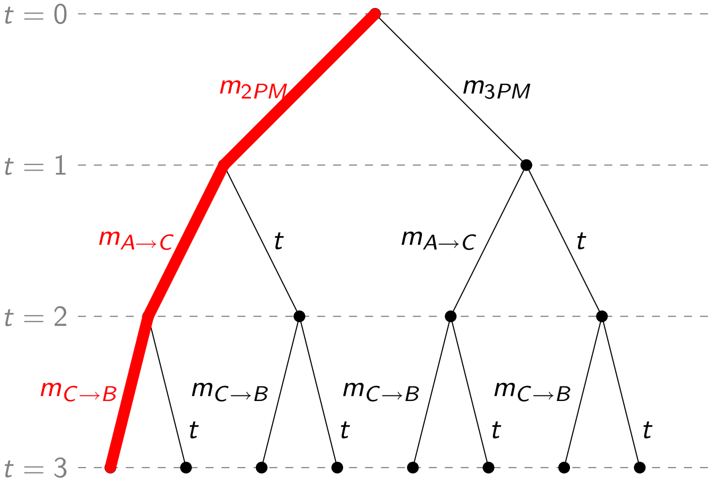
Returning to the Example

Ann would like Bob to attend her talk; however, she only wants Bob to attend if he is interested in the subject of her talk, not because he is just being polite.

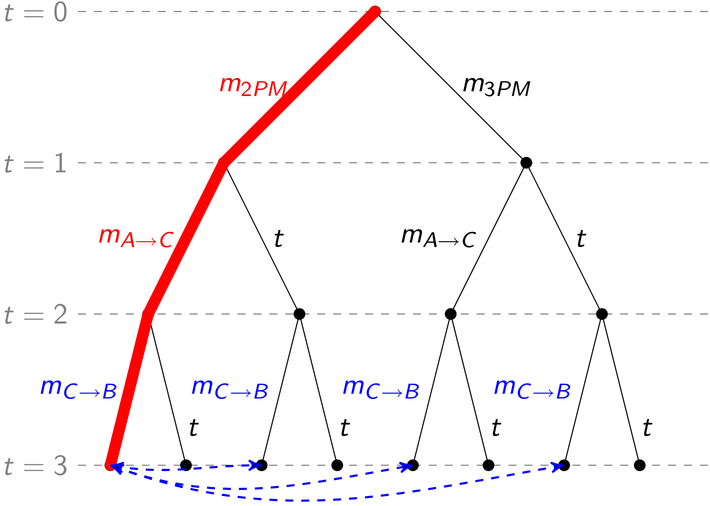
There is a very simple procedure to solve Ann's problem: *have a (trusted) friend tell Bob the time and subject of her talk.*

Is this procedure correct?

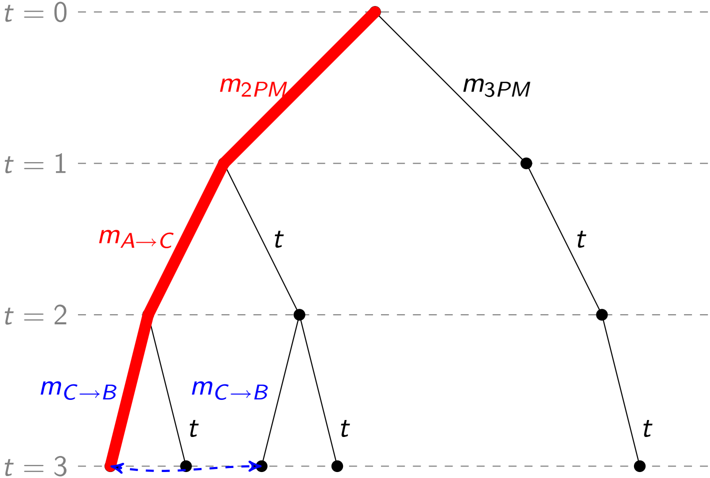

t = 0
t = 1
t = 2
t = 3
m2PM
m3PM
mA→C
t
mA→C
t
mC→B
t
mC→B
t
mC→B
t
mC→B
t



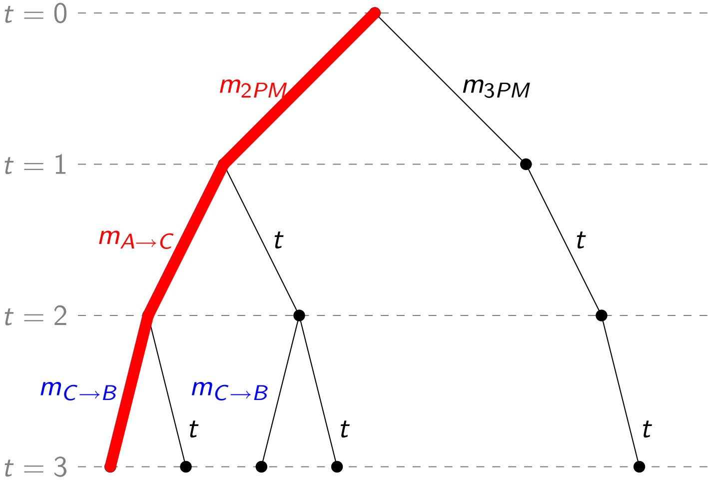
$$H, 3 \models \varphi$$



Bob's uncertainty: $H, 3 \models \neg K_B P_{2PM}$

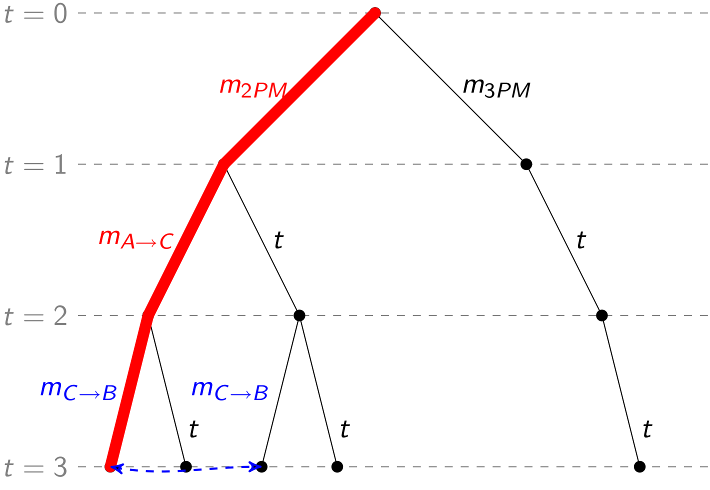


Bob's uncertainty + 'Protocol information': $H, 3 \models K_B P_{2PM}$



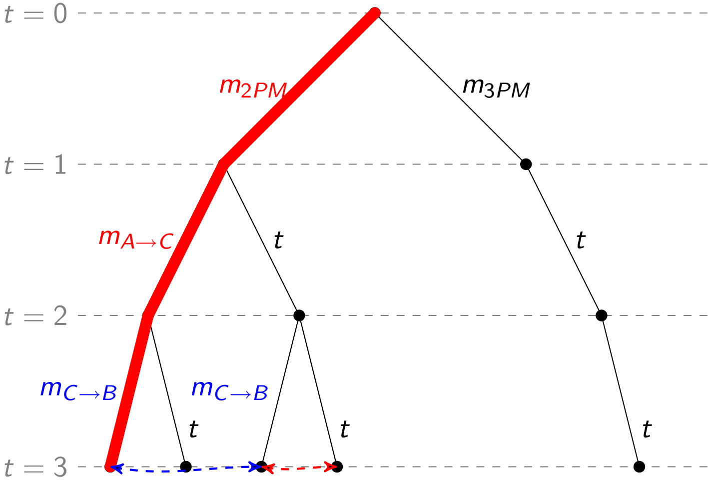
Bob's uncertainty + 'Protocol information':

$H, 3 \models \neg K_B K_A K_B P_{2PM}$



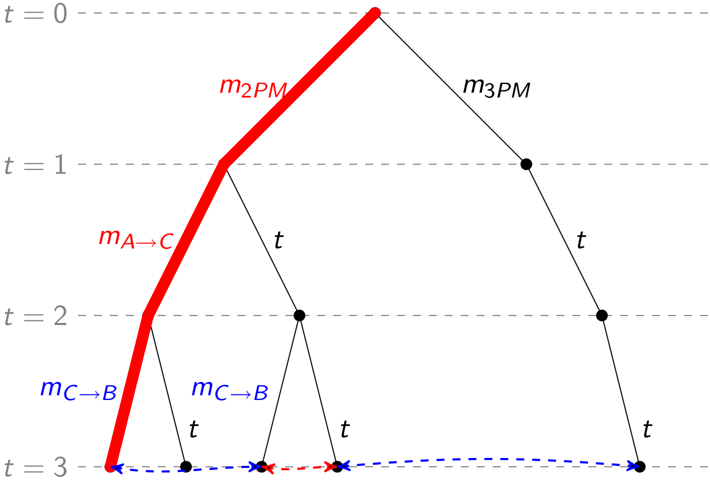
Bob's uncertainty + 'Protocol information':

$H, 3 \models \neg K_B K_A K_B P_{2PM}$



Bob's uncertainty + 'Protocol information':

$$H, 3 \models \neg K_B K_A K_B P_{2PM}$$



Bob's uncertainty + 'Protocol information':

$H, 3 \models \neg K_B K_A K_B P_{2PM}$

Parameters of the Logical Framework

Parameters of the Logical Framework

1. Expressivity of the formal language. Does the language include a common knowledge operator? A future operator? Both?

Parameters of the Logical Framework

1. Expressivity of the formal language. Does the language include a common knowledge operator? A future operator? Both?

2. Structural conditions on the underlying event structure. Do we restrict to protocol frames (finitely branching trees)? Finitely branching forests? Or, arbitrary ETL frames?

Parameters of the Logical Framework

1. Expressivity of the formal language. Does the language include a common knowledge operator? A future operator? Both?

2. Structural conditions on the underlying event structure. Do we restrict to protocol frames (finitely branching trees)? Finitely branching forests? Or, arbitrary ETL frames?

3. Conditions on the reasoning abilities of the agents. Do the agents satisfy perfect recall? No miracles? Do they agents' know what time it is?

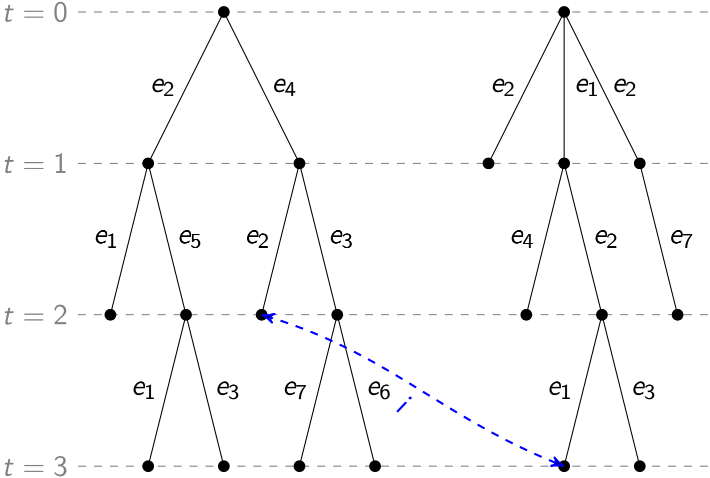
Agent Oriented Properties:

- **No Miracles**: For all finite histories $H, H' \in \mathcal{H}$ and events $e \in \Sigma$ such that $He \in \mathcal{H}$ and $H'e \in \mathcal{H}$, if $H \sim_i H'$ then $He \sim_i H'e$.

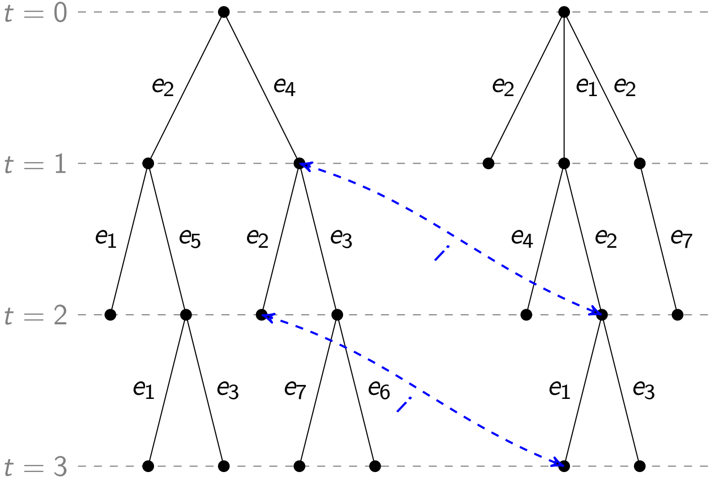
- **Perfect Recall**: For all finite histories $H, H' \in \mathcal{H}$ and events $e \in \Sigma$ such that $He \in \mathcal{H}$ and $H'e \in \mathcal{H}$, if $He \sim_i H'e$ then $H \sim_i H'$.

- **Synchronous**: For all finite histories $H, H' \in \mathcal{H}$, if $H \sim_i H'$ then $\text{len}(H) = \text{len}(H')$.

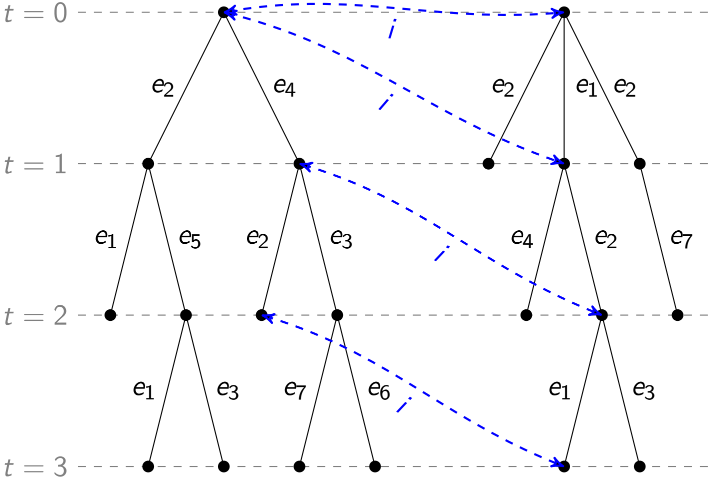
Perfect Recall



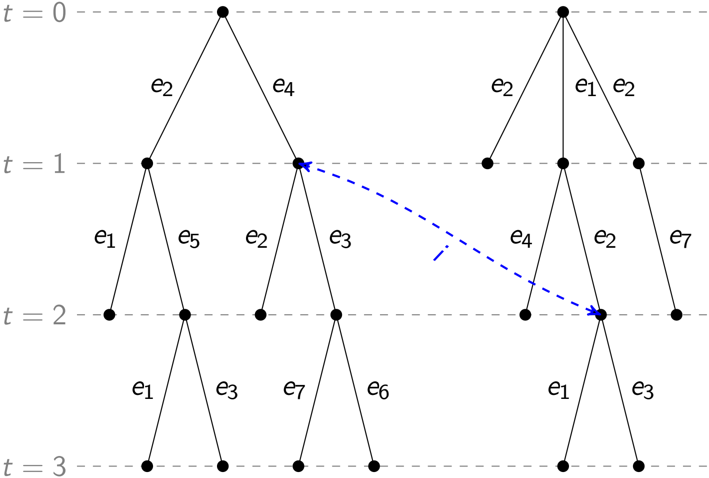
Perfect Recall



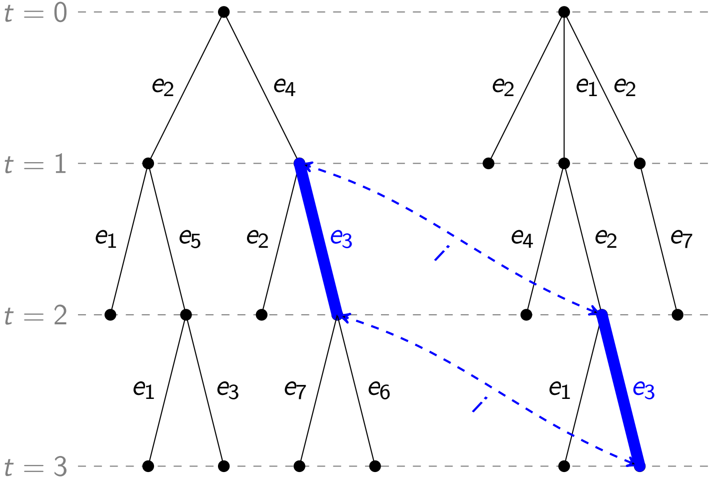
Perfect Recall



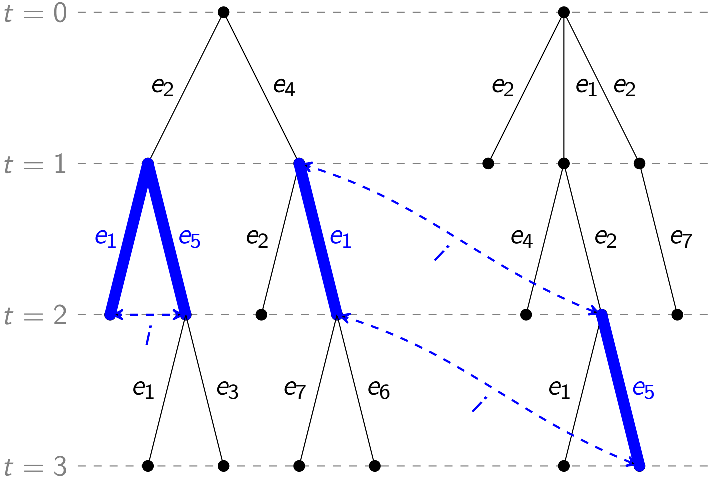
No Miracles



No Miracles



No Miracles



Ideal Agents

Assume there are two agents

Theorem

The logic of ideal *agents with respect to a language with common knowledge and future is highly undecidable (for example, by assuming perfect recall).*

J. Halpern and M. Vardi.. *The Complexity of Reasoning abut Knowledge and Time*. *J. Computer and Systems Sciences*, 38, 1989.

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

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}).

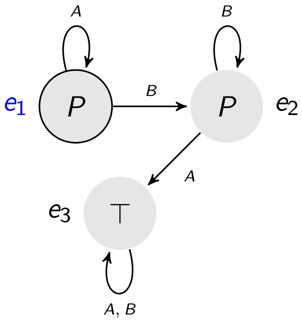
Private announcement to Ann about the talk.

Abstract Description of the Event

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

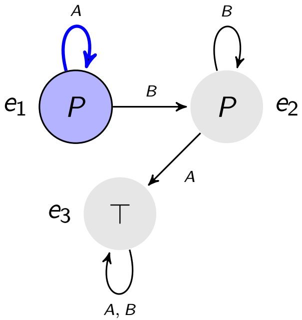
Abstract Description of the Event

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



Abstract Description of the Event

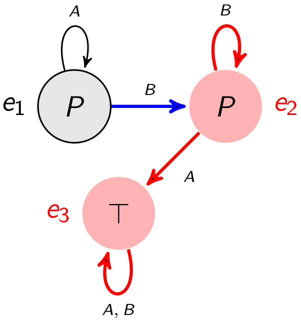
Recall the Ann and Bob example: Charles tells Bob that the talk is at 2PM.



Ann knows which event took place.

Abstract Description of the Event

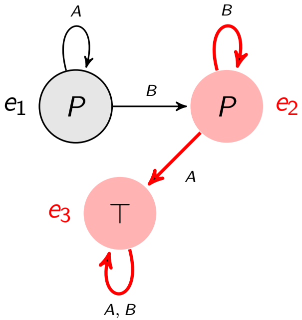
Recall the Ann and Bob example: Charles tells Bob that the talk is at 2PM.



Bob thinks a different event took place.

Abstract Description of the Event

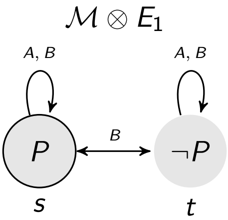
Recall the Ann and Bob example: Charles tells Bob that the talk is at 2PM.

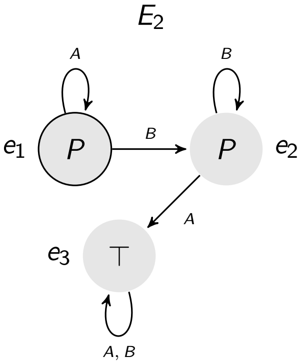


That is, Bob learns the time of the talk, but Ann learns nothing.

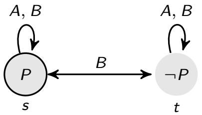
Product Update

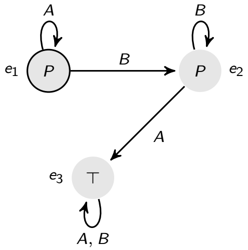
Product Update

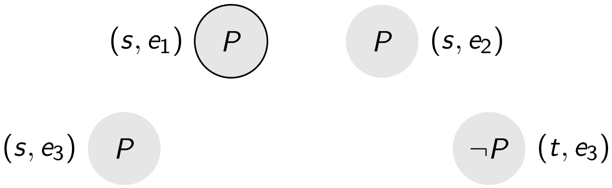




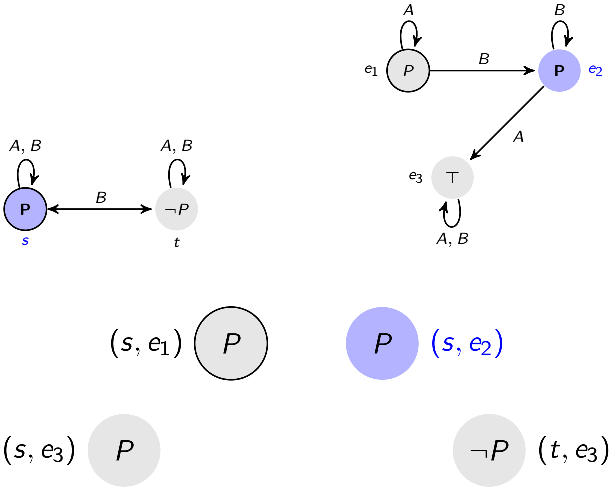
Product Update



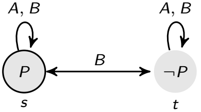


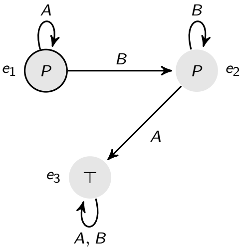


Product Update

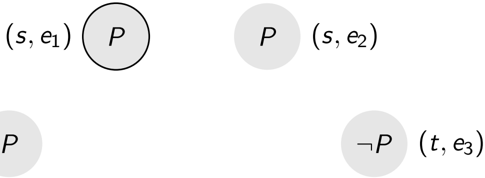


Product Update

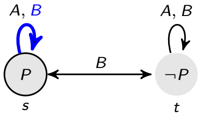


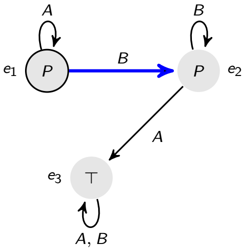


$(s, e_1) \models \neg K_B K_A K_B P$

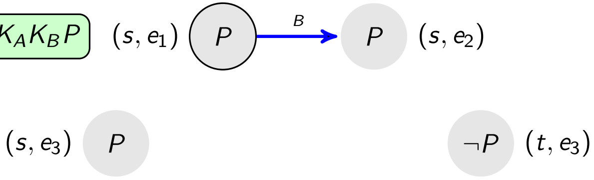


Product Update

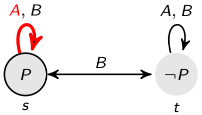


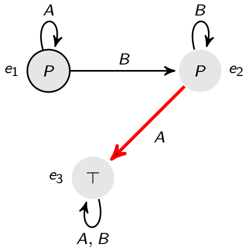


$(s, e_1) \models \neg K_B K_A K_B P$

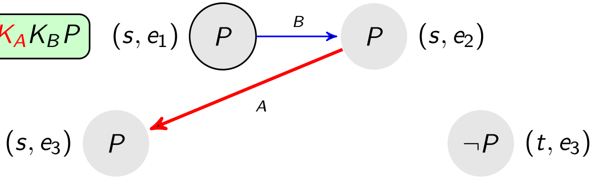


Product Update

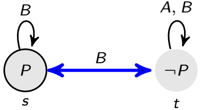


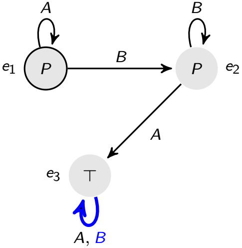


$(s, e_1) \models \neg K_B K_A K_B P$

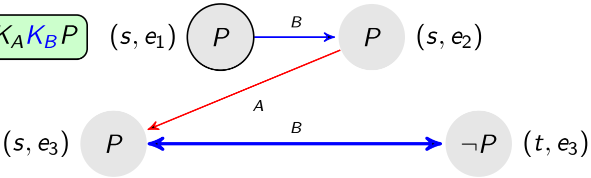


Product Update





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



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 $Pre : \mathcal{L} \to \wp(A)$.

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 $Pre : \mathcal{L} \rightarrow \wp(A)$.

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 $Pre : \mathcal{L} \to \wp(A)$.

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.

Some Questions

- How do we relate the ETL-style analysis with the DEL-style analysis?
- In the DEL setting, what are the underlying assumptions about the reasoning abilities of the agents?
- Can we axiomatize interesting subclasses of ETL frames?

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

Skip Details

DEL *and* ETL

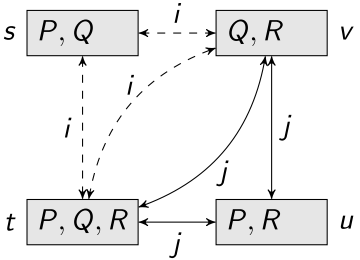
Observation: By repeatedly updating an epistemic model with event models, the machinery of DEL creates ETL models.

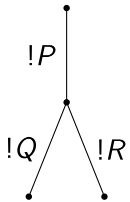
DEL *and* ETL

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

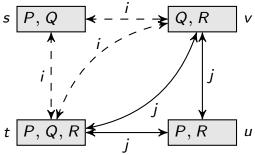
Let M be an epistemic model, and P a DEL protocol (tree of event models). The ETL model generated by M and P, forest(M, P), represents all possible evolutions of the system obtained by updating M with sequences from P.

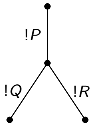
Example: Initial Model and Protocol



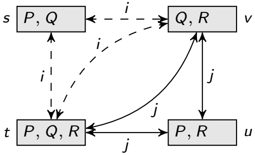


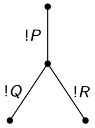
Example

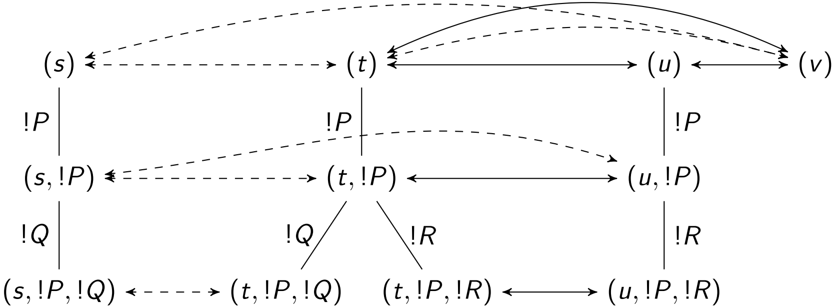




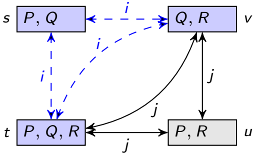
Example

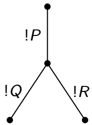

s
P, Q
i
Q, R
v
i
i
j
j
t
P, Q, R
j
P, R
u

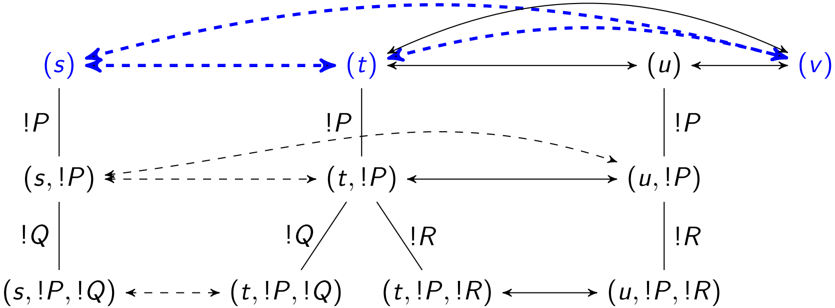

!P
!Q
!R


(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)

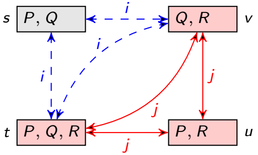
Example

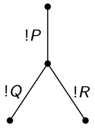

s
P, Q
i
Q, R
v
i
i
j
j
t
P, Q, R
j
P, R
u

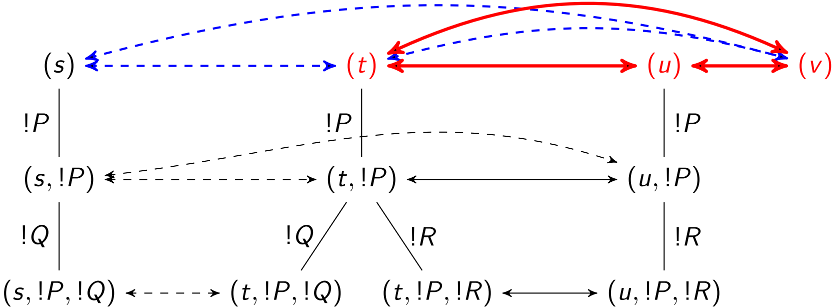

!P
!Q
!R


(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)

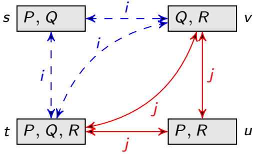
Example

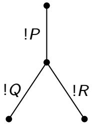

s
P, Q
i
Q, R
v
i
i
j
j
t
P, Q, R
P, R
u
j

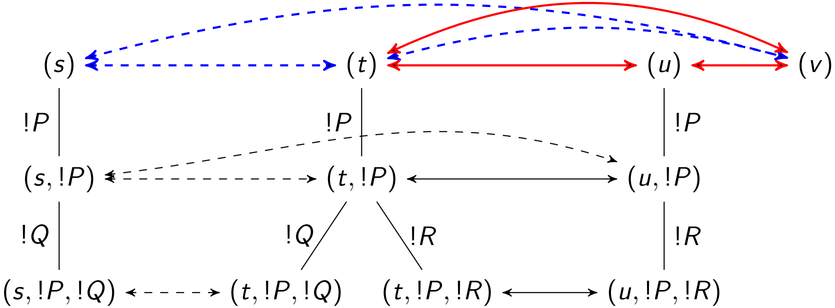

!P
!Q
!R


(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)

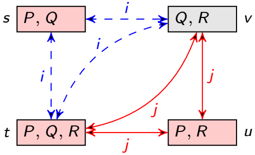
Example

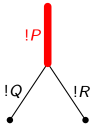

s
P, Q
Q, R
v
i
i
i
j
j
t
P, Q, R
P, R
u
j

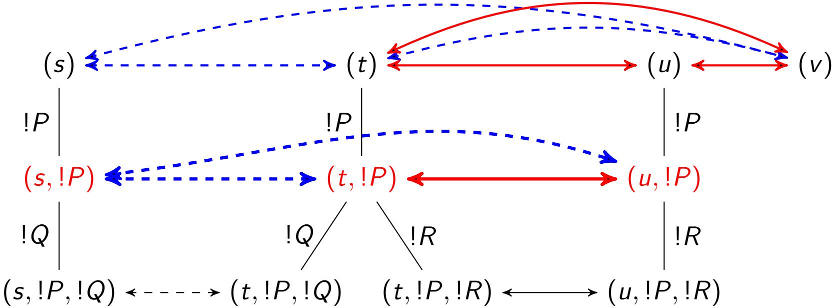

!P
!Q
!R


(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)

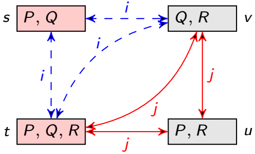
Example

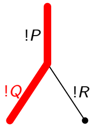

s
P, Q
Q, R
v
i
i
i
j
j
t
P, Q, R
P, R
u
j

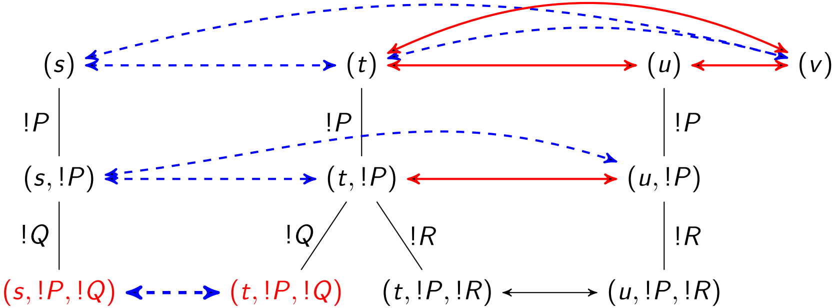

!P
!Q
!R


(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)

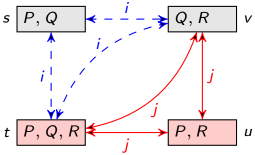
Example

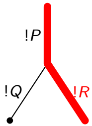

s
P, Q
i
Q, R
v
i
i
j
j
t
P, Q, R
P, R
u
j

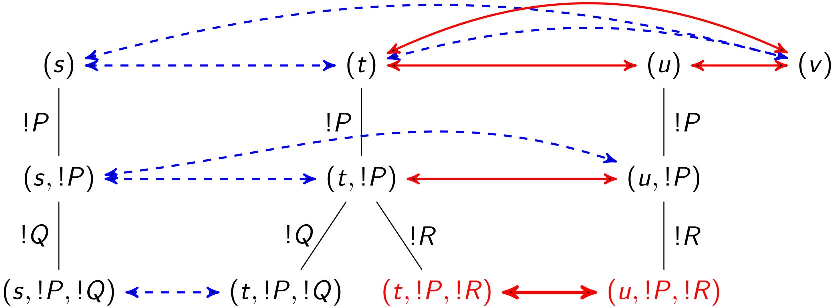

!P
!Q
!R


(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)

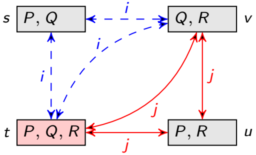
Example

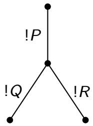

s
P, Q
i
Q, R
v
i
i
j
t
P, Q, R
j
P, R
u
j

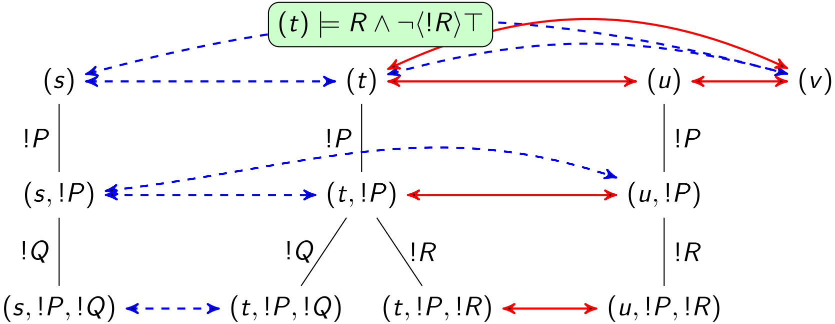

!P
!Q
!R


(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


(t) ⊨ R ∧ ¬⟨!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)

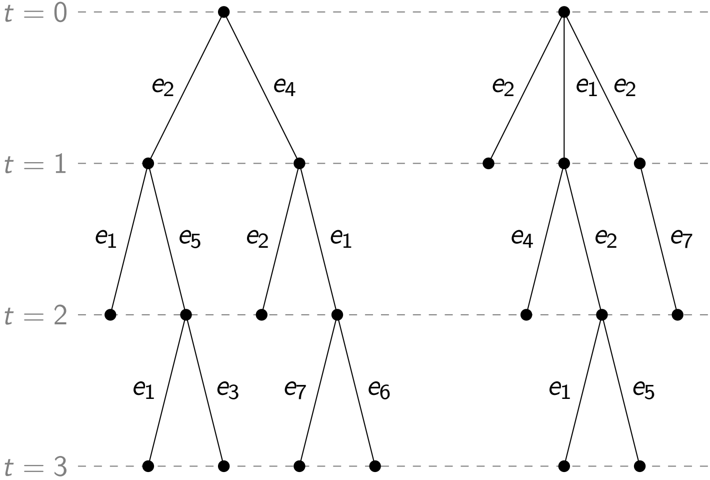
Representation Result

Given a set of DEL protocols $\mathbf{X}$, let $\mathbb{F}(\mathbf{X})$ be the class of ETL frames generated by protocols from $\mathbf{X}$.

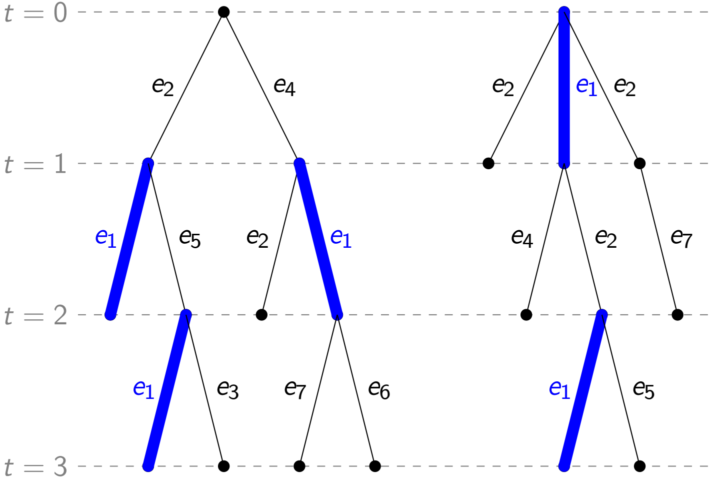
Theorem (Main Representation Theorem)

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). 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.

Bisimulation Invariance + Finiteness Condition



Bisimulation Invariance + Finiteness Condition



Recall that if $\mathbf{X}$ is a set of DEL protocols, we define $\mathbb{F}(\mathbf{X}) = \{\mathbb{F}(\mathcal{M}, \mathsf{P}) \mid \mathcal{M} \text{ an epistemic model and } \mathsf{P} \in \mathbf{X}\}$. This construction suggests the following natural questions:

- Which DEL protocols generate interesting ETL models?
- Which modal languages are most suitable to describe these models?
- Can we axiomatize interesting classes DEL-generated ETL models?

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

Announcement + Protocol Information

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

Announcement + Protocol Information

1. $A \rightarrow \langle A \rangle \top$ vs. $\langle A \rangle \top \rightarrow A$
2. $\langle A \rangle K_i P \leftrightarrow A \wedge K_i \langle A \rangle P$

Announcement + Protocol Information

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

Announcement + Protocol Information

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

Announcement + Protocol Information

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

Theorems Sound and complete axiomatizations of various generated ETL models.

Conclusions

Merging logics of rational agency

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

Logics of Knowledge and Preference

$K(\varphi \succeq \psi)$: "Ann knows that φ is at least as good as ψ"

$K\varphi \succeq K\psi$: "knowing φ is at least as good as knowing ψ

Logics of Knowledge and Preference

$K(\varphi \succeq \psi)$: "Ann knows that φ is at least as good as ψ"

$K\varphi \succeq K\psi$: "knowing φ is at least as good as knowing ψ

$\mathcal{M} = \langle W, \sim, \succeq, V \rangle$

Logics of Knowledge and Preference

$K(\varphi \succeq \psi)$: "Ann knows that φ is at least as good as ψ"

$K\varphi \succeq K\psi$: "knowing φ is at least as good as knowing ψ

$\mathcal{M} = \langle W, \sim, \succeq, V\rangle$

J. van Eijck. *Yet more modal logics of preference change and belief revision*. manuscript, 2009.

F. Liu. *Changing for the Better: Preference Dynamics and Agent Diversity*. PhD thesis, ILLC, 2008.

$$A(\psi \rightarrow \langle \succeq \rangle \varphi) \quad \textit{vs.} \quad K(\psi \rightarrow \langle \succeq \rangle \varphi)$$

$$A(\psi \rightarrow \langle \succeq \rangle \varphi) \quad \textit{vs.} \quad K(\psi \rightarrow \langle \succeq \rangle \varphi)$$

Should preferences be restricted to information sets?

$$A(\psi \rightarrow \langle \succeq \rangle \varphi) \;\; \textit{vs.} \;\; K(\psi \rightarrow \langle \succeq \rangle \varphi)$$

Should preferences be restricted to information sets?

$\mathcal{M}, w \models \langle \succeq \cap \sim \rangle \varphi$ iff there is a v with $w \sim v$ and $w \preceq v$ such that $\mathcal{M}, v \models \varphi$

$K(\psi \rightarrow \langle \succeq \cap \sim \rangle \varphi)$

Defining Beliefs from Preferences

- Starting with the work of Savage (based on Ramsey and de Finetti), there is a tradition in game theory and decision theory to *define* beliefs and utilities in terms of the agent's preferences

Defining Beliefs from Preferences

- Starting with the work of Savage (based on Ramsey and de Finetti), there is a tradition in game theory and decision theory to *define* beliefs and utilities in terms of the agent's preferences

- Typically the results come in the form of a representation theorem:

  If the agents preferences satisfy such-and-such properties, then there is a set of conditional probability functions and a (state independent) utility function such that the agent can be assumed to act as an expected utility maximizer.

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

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

The Framework

Let Ω be a set of states.

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

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

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

Belief Operators

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

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

B is normal if

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

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

Defining Beliefs from Preferences

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

Defining Beliefs from Preferences

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

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

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

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

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

▸ Conclusions

Merging logics of rational agency

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

Some Literature

Stemming from Bratman's planning theory of intention a number of *BDI logics*:

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

Some Literature

Stemming from Bratman's planning theory of intention a number of *BDI logics*:

- Cohen and Levesque; Rao and Georgeff; Meyer, 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.

Bratman's Planning Theory of Intention

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

Bratman's Planning Theory of Intention

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

A plan is a *conduct-controlling* mental attitude

Bratman's Planning Theory of Intention

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

A plan is a *conduct-controlling* mental attitude

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

Bratman's Planning Theory of Intention

An agent commits to a (partial) plan that is

Bratman's Planning Theory of Intention

An agent commits to a (partial) plan that is

1. means-end coherent,

Bratman's Planning Theory of Intention

An agent commits to a (partial) plan that is

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

Bratman's Planning Theory of Intention

An agent commits to a (partial) plan that is

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

Bratman's Planning Theory of Intention

An agent commits to a (partial) plan that is

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

Bratman's Planning Theory of Intention

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

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

A Methodological Issue

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

A Methodological Issue

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

Two Extremes:

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

A Methodological Issue

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

Two Extremes:

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

A Methodological Issue

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

Two Extremes:

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

2. Reasoning *about* multiagent systems.

A Methodological Issue

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

Two Extremes:

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

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

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

C & L Logic of Intention

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

C & L Logic of Intention

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

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

Methodological Issues

A third alternative:

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

Methodological Issues

A third alternative:

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

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

Planning vs. Database Management

1. How does an agent *generate* new intentions?
2. Given that the agent's intentions specify a *partial plan*, how and when is the plan "filled out"?
3. How does an agent choose a particular *action* (that is under its control) given its current intentions?
4. How should an agent *maintain* its current state of beliefs and intentions in the presence of new information or new intentions?
5. When should an agent *reconsider* its intentions?

Thomas Icard, EP and Yoav Shoham. *Intention and Belief Revision*. in preparation.

Our Framework

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

Elements of a Logic of Intention Revision

Elements of a Logic of Intention Revision

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

Elements of a Logic of Intention Revision

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

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

Elements of a Logic of Intention Revision

- Beliefs in a dynamic environment: certainty (irrevocable knowledge, hard information), belief (revisable, soft information), *safe* belief
- Three views of actions: PDL (state changing), Temporal (lay out time and actions are sequences of time points), STIT (choices, or actions, constrain the future).
- Two types of beliefs: those about the state of the world and those about the future *which are governed by the agent's plans*

Intention Revision

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

Intention Revision

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

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

Intention Revision

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

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

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

Intention Revision

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

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

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

Intention Revision

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

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

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

Intention Revision

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

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

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

Intention Revision

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

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

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

Our Framework

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

Our Framework

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

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

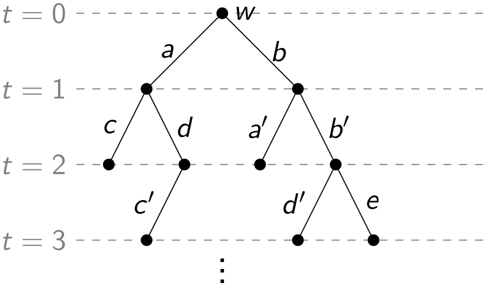
Our Framework

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

Skip Details

Choice Situations

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



Choice Situations: $\mathcal{L}_1$

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

Choice Situations: $\mathcal{L}_1$

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

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

Choice Situations: $\mathcal{L}_1$

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

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

Notation: If $\alpha = a_1a_2a_3\cdots a_n$, $\langle\alpha\rangle\varphi := \langle a_1\rangle\cdots\langle a_n\rangle\varphi$

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

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

Adding Beliefs

Standard picture where worlds are choice situations

Adding Beliefs

Standard picture where worlds are choice situations

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

Adding Beliefs

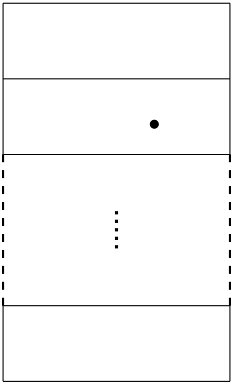
Standard picture where worlds are choice situations

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

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

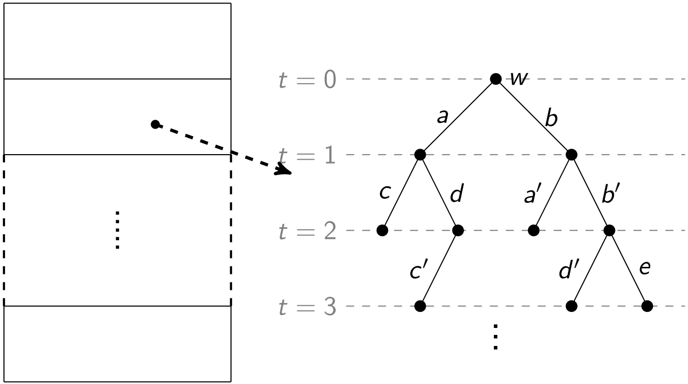
Belief Structures

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



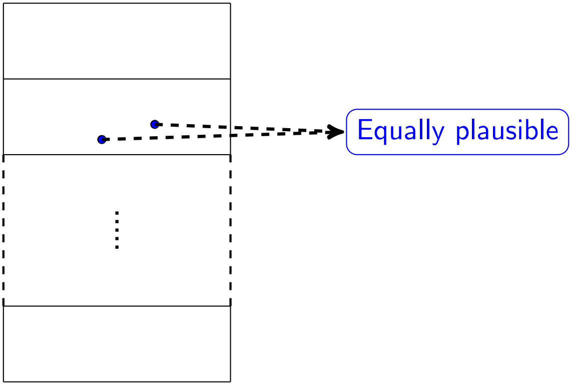
Belief Structures

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



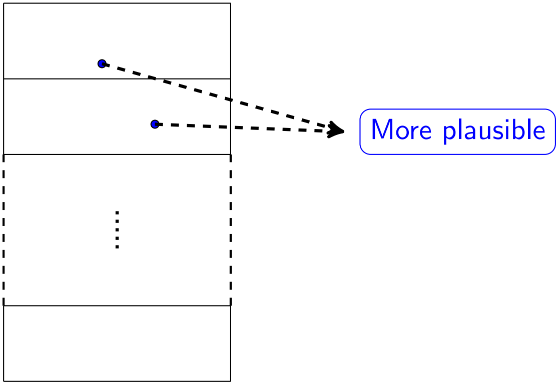
Belief Structures

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



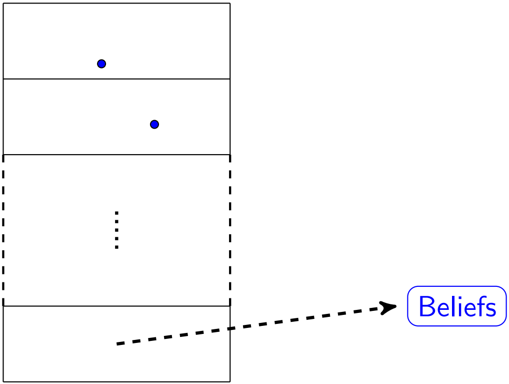
Belief Structures

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



Belief Structures

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



Belief Structures

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

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

(i) S a set of choice situations

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

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

Belief Structures

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

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

(i) S a set of choice situations

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

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

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

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

Belief Structures

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

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

(i) S a set of choice situations

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

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

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

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

Belief Structures

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

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

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

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

Completeness

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

Instructions

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

Instructions

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

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

Instructions

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

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

Instructions

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

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

Instructions

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

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

Instructions

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

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

Dynamics

There are three sources of dynamics:

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

Dynamics

There are three sources of dynamics:

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

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

Dynamics

There are three sources of dynamics:

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

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

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

Dynamics

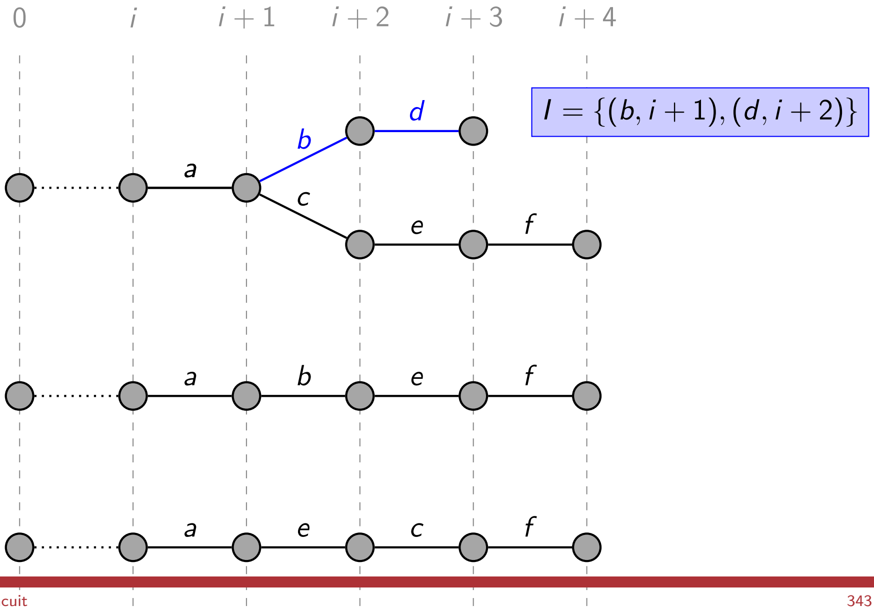
There are three sources of dynamics:

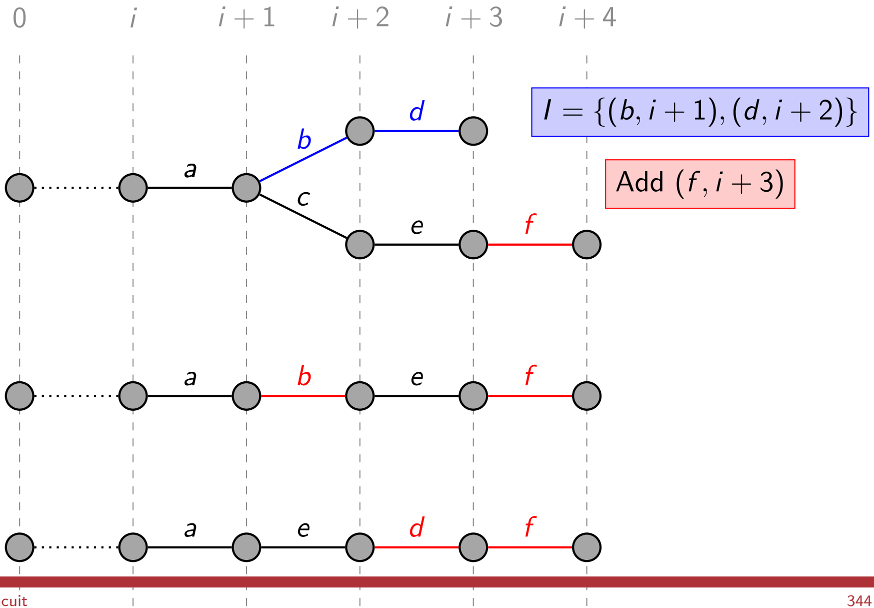
1. Nature can reveal (true) facts about the current choice situation (eg., facts that are true, choices that are available/not available in the future).

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

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

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


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


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

Selection Function

Say a set of *beliefs* $\mathcal{B}$ and a set of *instructions* I is **coherent** if the agent doesn't believe the instructions are impossible.
A **selection function** γ maps a set of beliefs $\mathcal{B}$ and instructions to a set of instructions: $\gamma(\mathcal{B}, I) = I'$

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

Selection Function

Say a set of *beliefs* $\mathcal{B}$ and a set of *instructions* I is **coherent** if the agent doesn't believe the instructions are impossible.
A **selection function** γ maps a set of beliefs $\mathcal{B}$ and instructions to a set of instructions: $\gamma(\mathcal{B}, I) = I'$

1. $\gamma(\mathcal{B}, I) \subseteq I$.
2. $\gamma(\mathcal{B}, I)$ is coherent with $\mathcal{B}$.
3. additional principles.....

Where we are going

AGM-style principles and representation theorem; Modal-style completeness (with dynamic operators get considerably more technical: *reduction axioms* are not available).

Where we are going

AGM-style principles and representation theorem; Modal-style completeness (with dynamic operators get considerably more technical: *reduction axioms* are not available).

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

1. The notion of Belief-Plan consistency must be updated
2. Define intentions *semantically*: the agent "intends a, t just in case it is a *necessary component* of the current plan".
3. Many agents
4. .....

Conclusions

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

Conclusions

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

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

Conclusions

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

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

- Normative vs. Descriptive

Conclusions

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

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

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

Conclusions

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

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

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

Conclusions

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

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

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

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

Conclusions

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

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

Logics of Rational Agency

- What's going on in the area:
  www.loriweb.org

- Special Issue of Synthese: Knowledge, Rationality and Interaction. *Logic and Intelligent Interaction*, Volume 169, Number 2 / July, 2009
  (eds. T. Agotnes, J. van Benthem and EP)

- New subarea of Stanford Encyclopedia of Philosophy on logic and rational agency
  (eds. J. van Benthem, EP, and O. Roy)

Calls for....

- **Papers**: LOFT 2010. University of Toulouse, July 21 - 23. Deadline: March 15, 2010.

- **Ph.D. position**: TiLPS, Tilburg University, "A formal analysis of social procedures". Deadline: **October 15** (to start in February).

Thank You!