Logics of Rational Agency

Lecture 2

Eric Pacuit

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

July 28, 2009

✓ Introduction, Motivation and Background

Lecture 2: Basic Ingredients for a Logic of Rational Agency

Lecture 3: Logics of Rational Agency and Social Interaction, Part I

Lecture 4: Logics of Rational Agency and Social Interaction, Part II

Lecture 5: Conclusions and General Issues

Logics of Rational Agency

Basic Ingredients

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

Basic Ingredients

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

Basic Ingredients

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

Basic Ingredients

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

- Single agent vs. many agents.

- What the the primitive operators?
  - Informational attitudes
  - Motivational attitudes
  - Normative attitudes

- Static vs. dynamic

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

Single-Agent Epistemic Logic

Typically, we write KP instead of $\Box P$ when the intended interpretation is "P is *known*"

Single-Agent Epistemic Logic

Typically, we write KP instead of $\Box P$ 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 instead of $\Box P$ 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 instead of $\Box P$ 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 instead of $\Box P$ when the intended interpretation is "P is *known*"

$K(P \rightarrow Q)$: "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 instead of $\Box P$ when the intended interpretation is "P is *known*"

$K(P \rightarrow Q)$: "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.

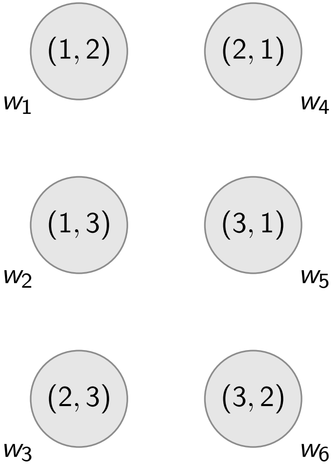
What are the relevant states?

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?

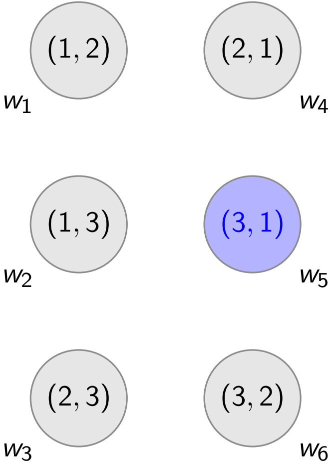


Example

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

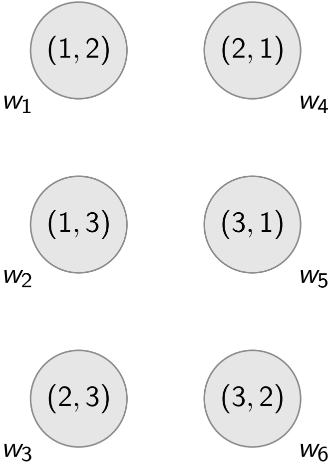


Example

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?

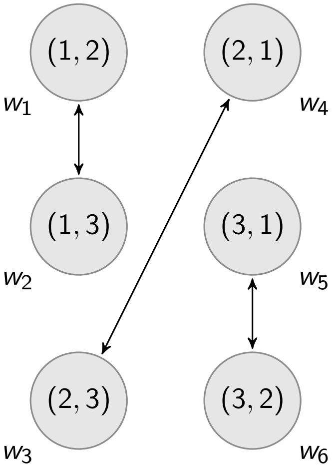


Example

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?

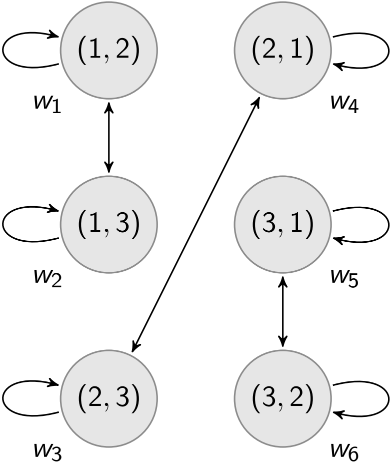


Example

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

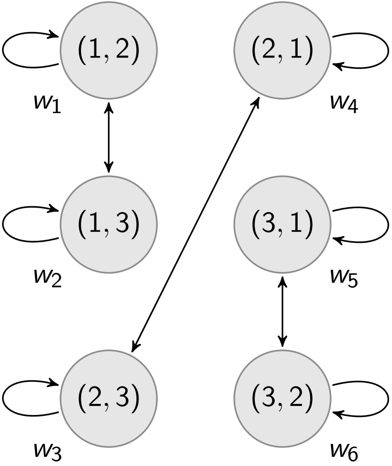
Suppose there are three cards: 1, 2 and 3.

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

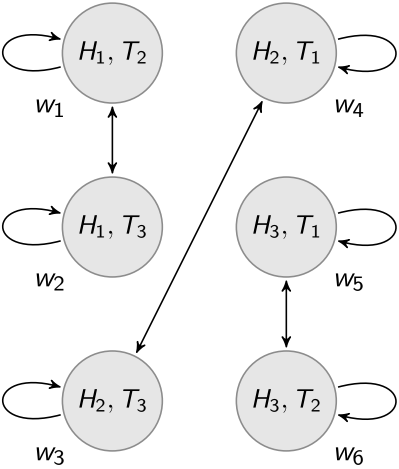
Suppose there are three cards: 1, 2 and 3.

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"

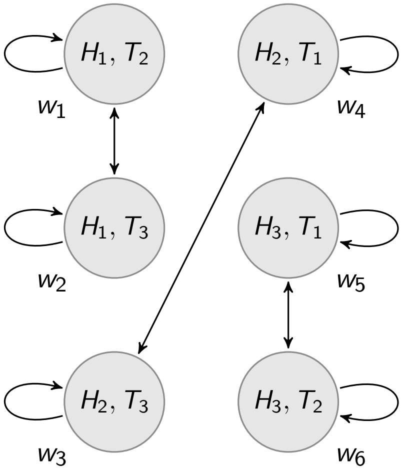
Eg., $V(H_1) = \{w_1, w_2\}$



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.

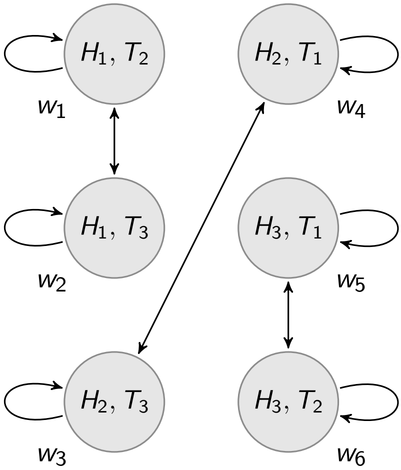


Example

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.

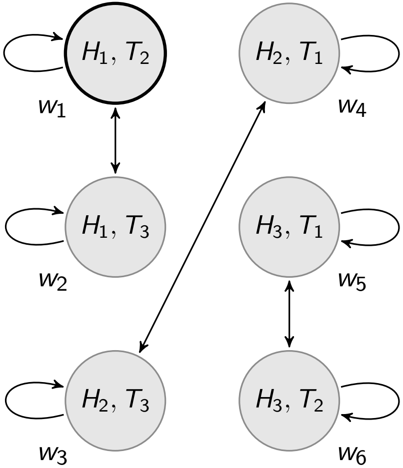


Example

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.

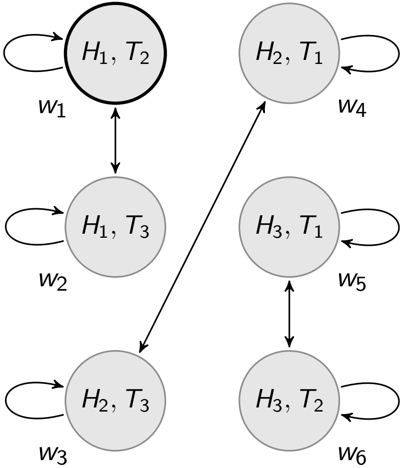


Example

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.

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

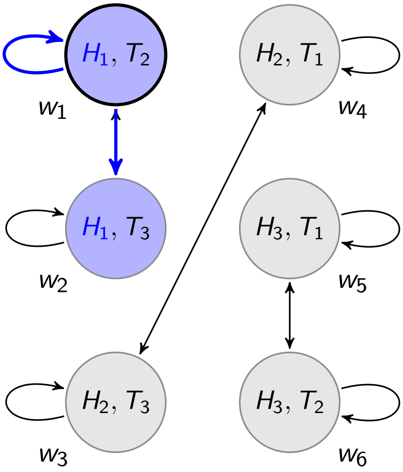


Example

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.

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



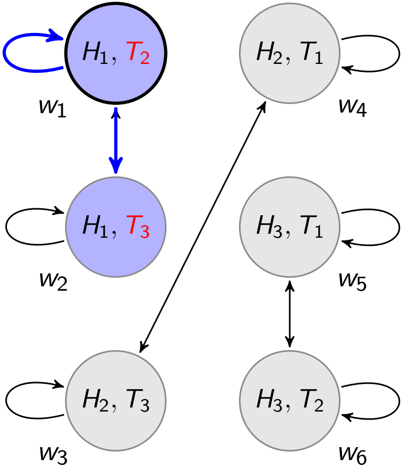
Example

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.

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

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

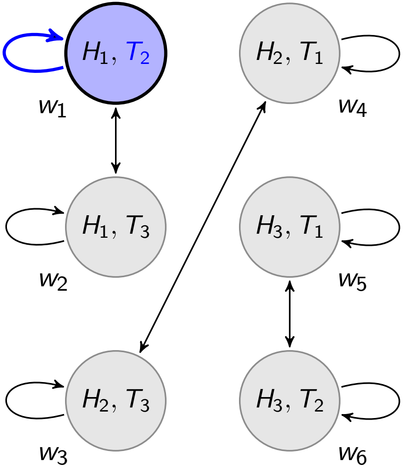


Example

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.

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

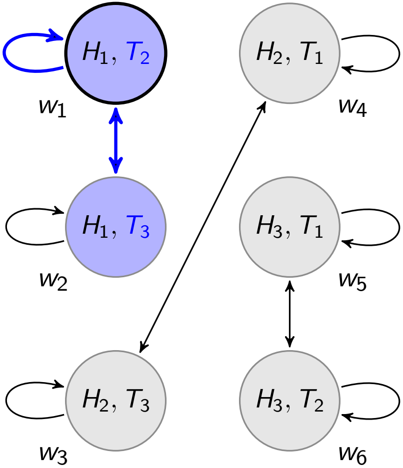


Example

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.

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



Some Questions

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

What idealizations have we made?

Some Notation

A Kripke Frame is a tuple $\langle W, R \rangle$ where $R \subseteq W \times W$.

φ is valid in a Kripke model $\mathcal{M}$ if $\mathcal{M}, w \models \varphi$ for all states w (we write $\mathcal{M} \models \varphi$).

φ is valid on a Kripke frame $\mathcal{F}$ if $\mathcal{M} \models \varphi$ for all models $\mathcal{M}$ based on $\mathcal{F}$.

Logical Omniscience

Fact: φ is valid then $K\varphi$ is valid

Logical Omniscience

Fact: $K\varphi \wedge K\psi \rightarrow K(\varphi \wedge \psi)$ is valid on all Kripke frames

Logical Omniscience

Fact: If $\varphi \to \psi$ is valid then $K\varphi \to K\psi$ is valid

Logical Omniscience

Fact: $K(\varphi \to \psi) \to (K\varphi \to K\psi)$ is valid on all Kripke frames.

Logical Omniscience

Fact: $\varphi \leftrightarrow \psi$ is valid then $K\varphi \leftrightarrow K\psi$ is valid

Correspondence

Definition

A model formula φ corresponds to a property P (of a relation in a Kripke frame) provided

$$\mathcal{F} \models \varphi \text{ iff } \mathcal{F} \text{ has } P$$

Correspondence

Definition

A model formula φ corresponds to a property P (of a relation in a Kripke frame) provided

$$\mathcal{F} \models \varphi \text{ iff } \mathcal{F} \text{ has } P$$

Modal Formula	Corresponding Property

Correspondence

Definition

A model formula φ corresponds to a property P (of a relation in a Kripke frame) provided

$$\mathcal{F} \models \varphi \text{ iff } \mathcal{F} \text{ has } P$$

Modal Formula	Corresponding Property
$K\varphi \rightarrow \varphi$	Reflexive

Correspondence

Definition

A model formula φ corresponds to a property P (of a relation in a Kripke frame) provided

$$\mathcal{F} \models \varphi \text{ iff } \mathcal{F} \text{ has } P$$

Modal Formula	Corresponding Property
$K\varphi \rightarrow \varphi$	Reflexive
$K\varphi \rightarrow KK\varphi$	Transitive

Correspondence

Definition

A model formula φ corresponds to a property P (of a relation in a Kripke frame) provided

$$\mathcal{F} \models \varphi \text{ iff } \mathcal{F} \text{ has } P$$

Modal Formula	Corresponding Property
$K\varphi \rightarrow \varphi$	Reflexive
$K\varphi \rightarrow KK\varphi$	Transitive
$\neg K\varphi \rightarrow K\neg K\varphi$	Euclidean

Correspondence

Definition

A model formula φ corresponds to a property P (of a relation in a Kripke frame) provided

$$\mathcal{F} \models \varphi \text{ iff } \mathcal{F} \text{ has } P$$

Modal Formula	Corresponding Property
$K\varphi \rightarrow \varphi$	Reflexive
$K\varphi \rightarrow KK\varphi$	Transitive
$\neg K\varphi \rightarrow K\neg K\varphi$	Euclidean
$\varphi \rightarrow KL\varphi$	Symmetric

Correspondence

Definition

A model formula φ corresponds to a property P (of a relation in a Kripke frame) provided

$$\mathcal{F} \models \varphi \text{ iff } \mathcal{F} \text{ has } P$$

Modal Formula	Corresponding Property
$K\varphi \rightarrow \varphi$	Reflexive
$K\varphi \rightarrow KK\varphi$	Transitive
$\neg K\varphi \rightarrow K\neg K\varphi$	Euclidean
$\varphi \rightarrow KL\varphi$	Symmetric
$\neg K\bot$	Serial

Modal Formula	Property	Philosophical Assumption

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

The Logic **S5**

The logic **S5** contains the following axioms and rules:

Pc	Axiomatization of Propositional Calculus
K	$K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$
T	$K\varphi \rightarrow \varphi$
4	$K\varphi \rightarrow KK\varphi$
5	$\neg K\varphi \rightarrow K\neg K\varphi$
MP	$\dfrac{\varphi \qquad \varphi \rightarrow \psi}{\psi}$
Nec	$\dfrac{\varphi}{K\psi}$

The Logic **S5**

The logic **S5** contains the following axioms and rules:

Pc	Axiomatization of Propositional Calculus
K	$K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$
T	$K\varphi \rightarrow \varphi$
4	$K\varphi \rightarrow KK\varphi$
5	$\neg K\varphi \rightarrow K\neg K\varphi$
MP	$\dfrac{\varphi \qquad \varphi \rightarrow \psi}{\psi}$
Nec	$\dfrac{\varphi}{K\psi}$

Theorem

S5 *is sound and strongly complete with respect to the class of Kripke frames with equivalence relations.*

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 $p \in \mathrm{At}$)
- $\mathcal{M}, w \models \neg\varphi$ if $\mathcal{M}, w \not\models \varphi$
- $\mathcal{M}, w \models \varphi \wedge \psi$ if $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models K\varphi$ 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

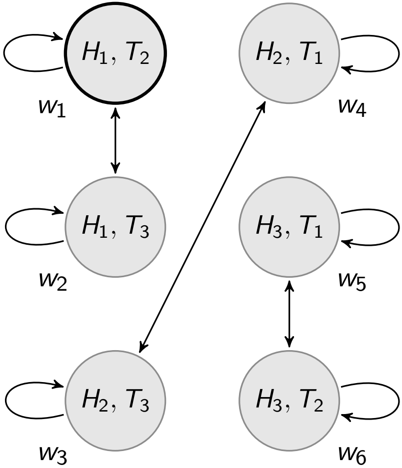
- $K_A K_B \varphi$: "Ann knows that Bob knows φ"
- $K_A(K_B\varphi \vee K_B\neg\varphi)$: "Ann knows that Bob knows whether φ
- $\neg K_B K_A K_B(\varphi)$: "Bob does not know that Ann knows that Bob knows that φ"

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.

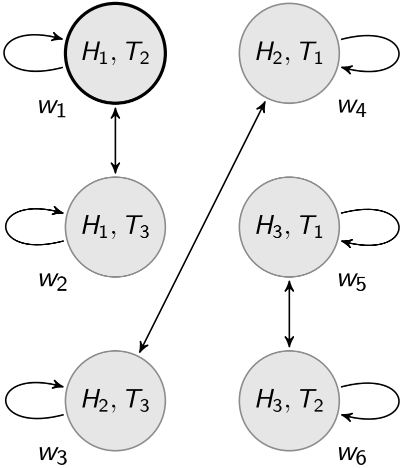


Example

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.

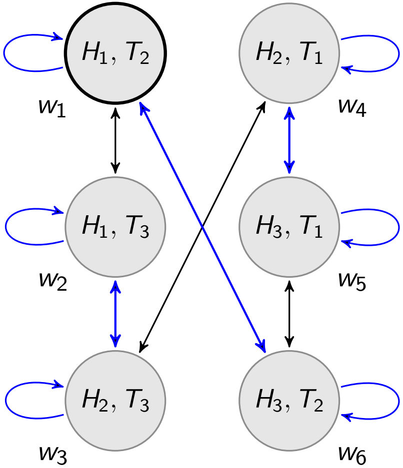


Example

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.

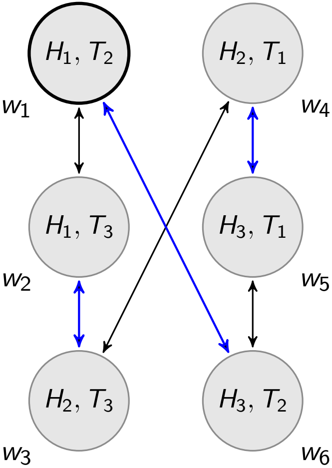


Example

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.



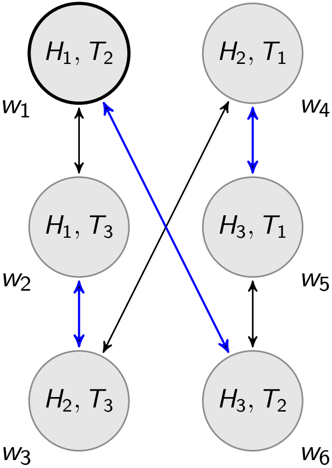
Example

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.

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



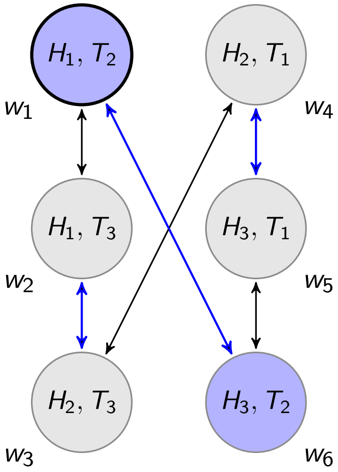
Example

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.

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



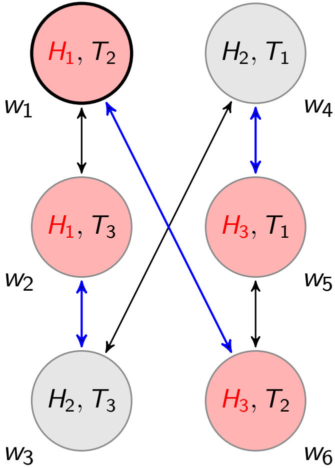
Example

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

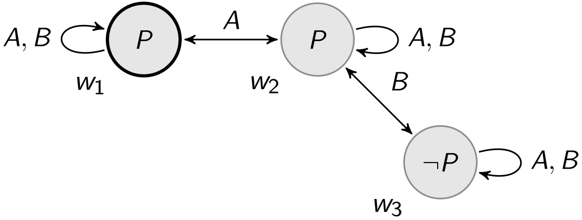
CP: "It is common knowledge that *P*" — "Everyone knows that everyone knows that everyone knows that $\cdots$ *P*".

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*?

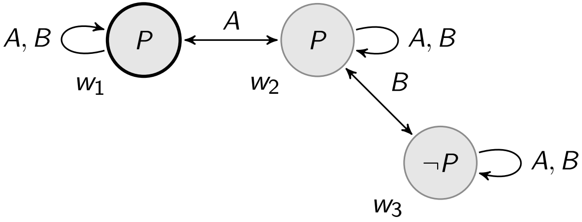


$$w_1 \models EP \wedge \neg CP$$

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*?

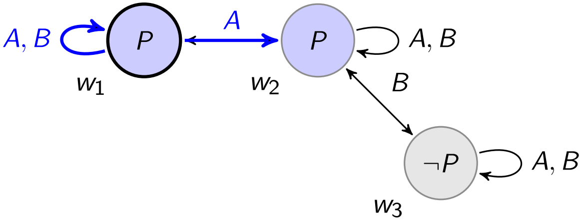


$$w_1 \models EP \wedge \neg CP$$

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*?

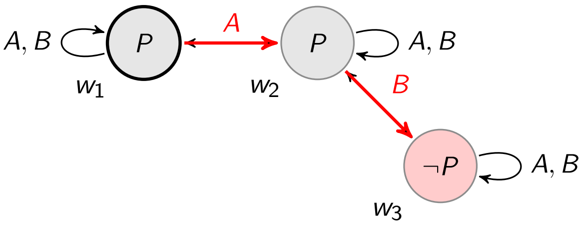


$$w_1 \models EP \wedge \neg CP$$

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*?



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

Spreading Gossip

Suppose that there are three friends, Ann, Bob and Charles, and Ann learns a interesting piece of news (P). If each of the friends are at home, how many calls are needed to create common knowledge that P?

Common Knowledge and Coordination

Suppose there are two friends Ann and Bob are on a bus separated by a crowd.

Common Knowledge and Coordination

Suppose there are two friends Ann and Bob are on a bus separated by a crowd. Before the bus comes to the next stop a mutual friend from outside the bus yells "get off at the next stop to get a drink?".

Common Knowledge and Coordination

Suppose there are two friends Ann and Bob are on a bus separated by a crowd. Before the bus comes to the next stop a mutual friend from outside the bus yells "get off at the next stop to get a drink?".

Say Ann is standing near the front door and Bob near the back door.

Common Knowledge and Coordination

Suppose there are two friends Ann and Bob are on a bus separated by a crowd. Before the bus comes to the next stop a mutual friend from outside the bus yells "get off at the next stop to get a drink?".

Say Ann is standing near the front door and Bob near the back door. When the bus comes to a stop, will they get off?

Common Knowledge and Coordination

Suppose there are two friends Ann and Bob are on a bus separated by a crowd. Before the bus comes to the next stop a mutual friend from outside the bus yells "get off at the next stop to get a drink?".

Say Ann is standing near the front door and Bob near the back door. When the bus comes to a stop, will they get off?

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.

$$CP \rightarrow ECP$$

$CP \rightarrow ECP$

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)*

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

Another 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.

Another 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).

Another 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?

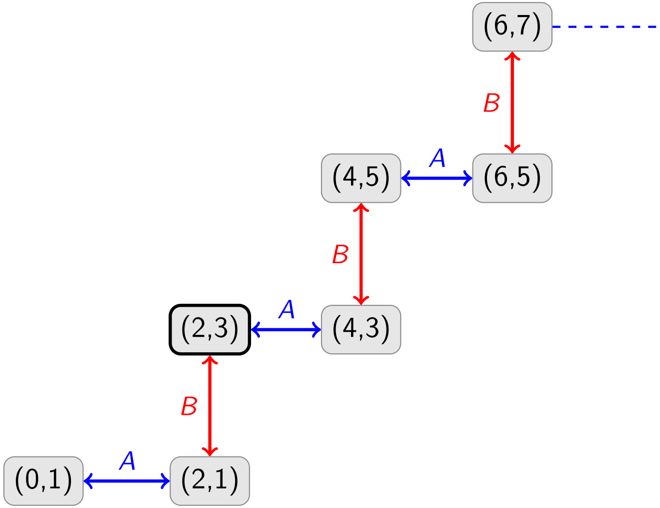
Another Example

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)

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

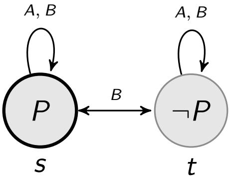
Ann would like Bob to attend her talk; however, she only wants Bob to attend if he is interested in the subject of her talk, not because he is just being polite.

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

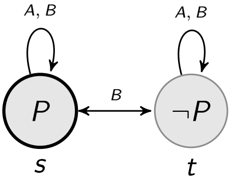
1. Ann knows about the talk.
2. Bob knows about the talk.
3. Ann knows that Bob knows about the talk.
4. Bob *does not* know that Ann knows that he knows about the talk.
5. *And nothing else.*

Example



P means "The talk is at 2PM".

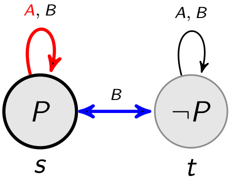
Example



P means "The talk is at 2PM".

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

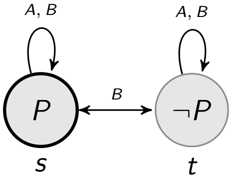
Example

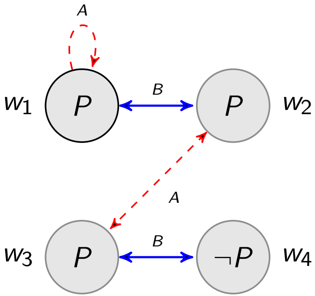


P means "The talk is at 2PM".

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

Example



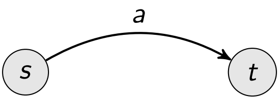


Actions

Actions as *transitions between states, or situations*:

Actions

Actions as *transitions between states, or situations*:



Propositional Dynamic Logic

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

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

$\mathcal{M}, w \models [\alpha]\varphi$ iff for each v, if $wR_\alpha v$ 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. $[\alpha](\varphi \to \psi) \to ([\alpha]\varphi \to [\alpha]\psi)$
3. $[\alpha \cup \beta]\varphi \leftrightarrow [\alpha]\varphi \wedge [\beta]\varphi$
4. $[\alpha;\beta]\varphi \leftrightarrow [\alpha][\beta]\varphi$
5. $[\psi?]\varphi \leftrightarrow (\psi \to \varphi)$
6. $\varphi \wedge [\alpha][\alpha^*]\varphi \leftrightarrow [\alpha^*]\varphi$ (Fixed-Point Axiom)
7. $\varphi \wedge [\alpha^*](\varphi \to [\alpha]\varphi) \to [\alpha^*]\varphi$ (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 ($\mathrm{EXPTIME}$-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:

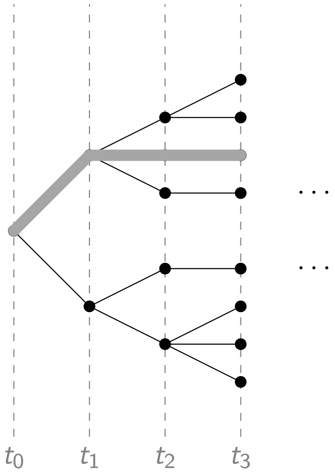
3. p is optimal (in some sense: shortest, maximally efficient, most convenient, etc.) in the set of computations satisfying conditions (1) and (2).

The axioms:

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

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

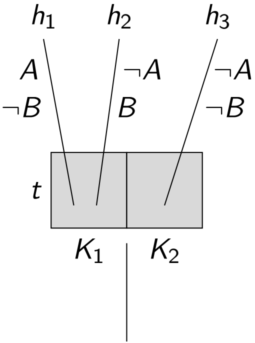
Example Continuing with this example, suppose that Ann has the ability to hit the dart board, but has no other control over the placement of the dart. Now, when she throws the dart, as a matter of fact, it will either hit the top half of the board or the bottom half of the board. Since, Ann has the ability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

End of lecture 2.