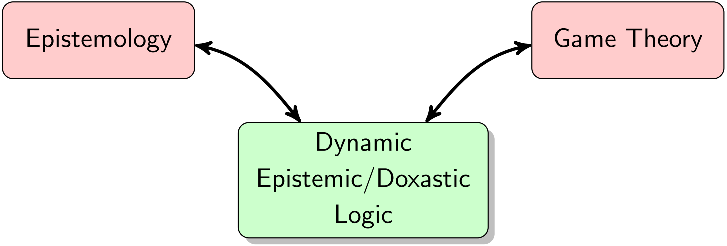
Modal Logic
Epistemic Logic

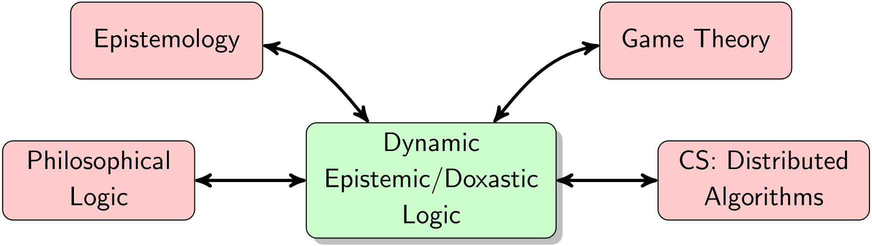
Eric Pacuit

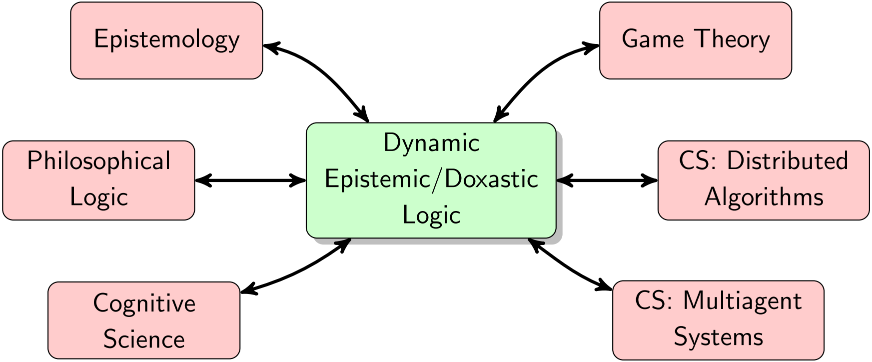
University of Maryland, College Park
ai.stanford.edu/∼epacuit

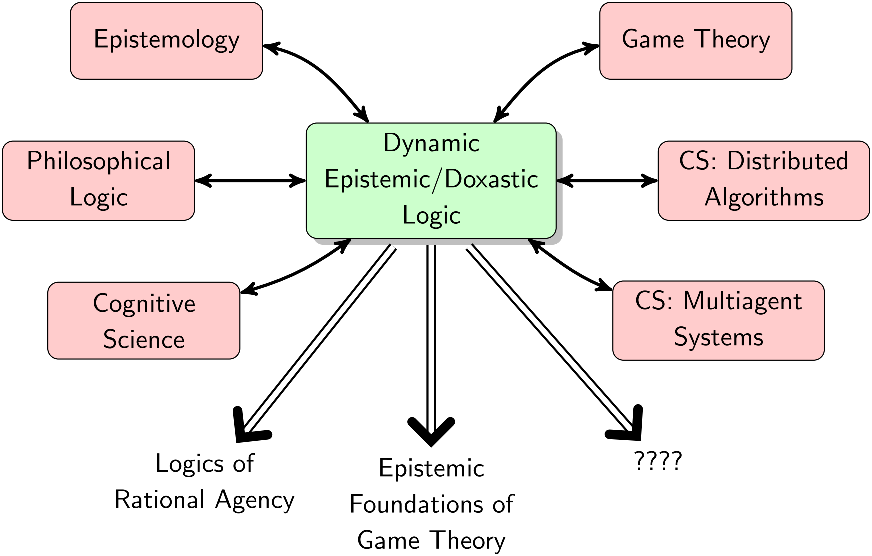
April 19, 2012

Dynamic
Epistemic/Doxastic
Logic

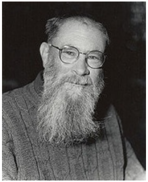

Epistemology
Game Theory
Dynamic Epistemic/Doxastic Logic


Epistemology
Game Theory
Philosophical Logic
Dynamic Epistemic/Doxastic Logic
CS: Distributed Algorithms


Epistemology
Game Theory
Philosophical Logic
Dynamic Epistemic/Doxastic Logic
CS: Distributed Algorithms
Cognitive Science
CS: Multiagent Systems

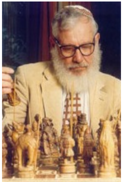

Epistemology
Game Theory
Philosophical Logic
Dynamic Epistemic/Doxastic Logic
CS: Distributed Algorithms
Cognitive Science
CS: Multiagent Systems
Logics of Rational Agency
Epistemic Foundations of Game Theory
????

Foundations of Epistemic Logic



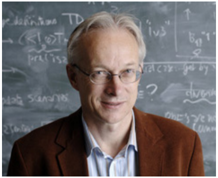
David Lewis

Jakko Hintikka



Robert Aumann

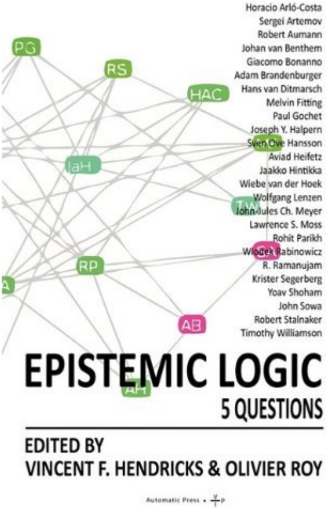
Larry Moss



Johan van Benthem

Alexandru Baltag

Foundations of Epistemic Logic



Single-Agent Epistemic Logic

Let KP informally mean “the agent knows that P (is true)”.

Single-Agent Epistemic Logic

Let KP informally mean "the agent knows that P (is true)".

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

Single-Agent Epistemic Logic

Let KP informally mean "the agent knows that P (is true)".

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

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

Single-Agent Epistemic Logic

Let KP informally mean "the agent knows that P (is true)".

$K(P \to Q)$: "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

Let KP informally mean "the agent knows that P (is true)".

$K(P \to Q)$: "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

Let KP informally mean "the agent knows that P (is true)".

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

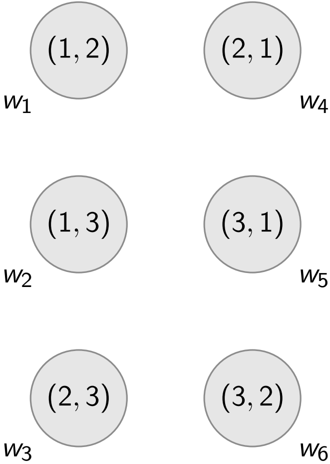
What are the relevant states?

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?

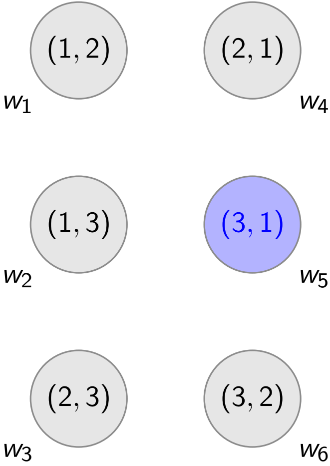


Example

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

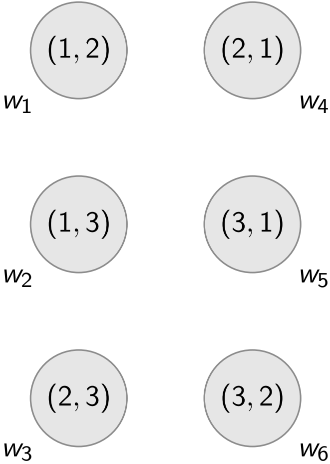


Example

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?

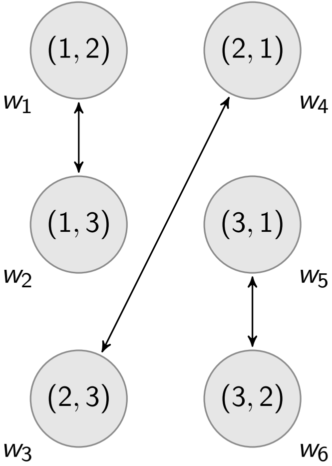


Example

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?

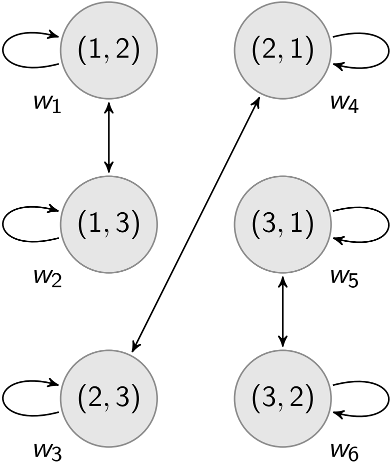


Example

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

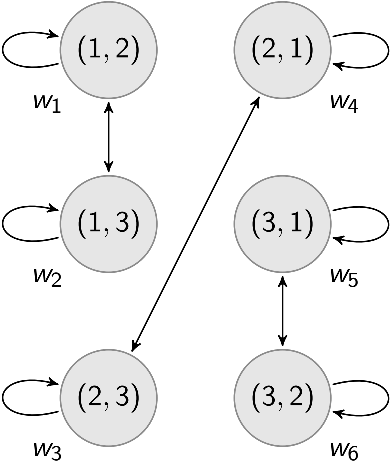
Suppose there are three cards: 1, 2 and 3.

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

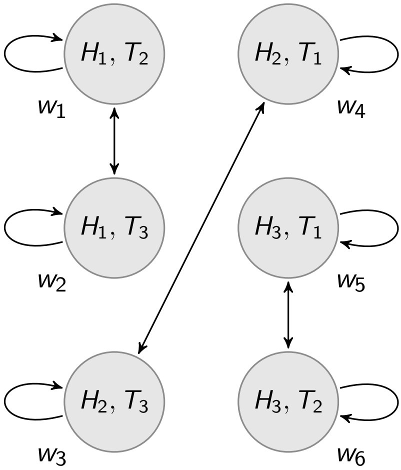
Suppose there are three cards: 1, 2 and 3.

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"

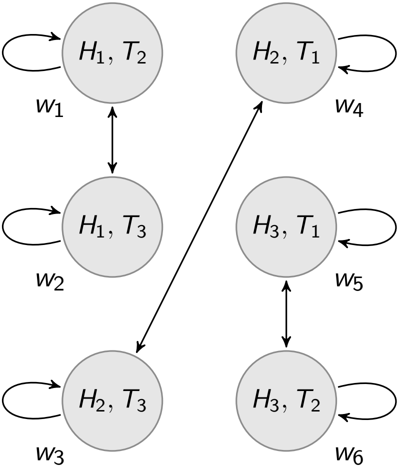
Eg., $V(H_1) = \{w_1, w_2\}$



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.

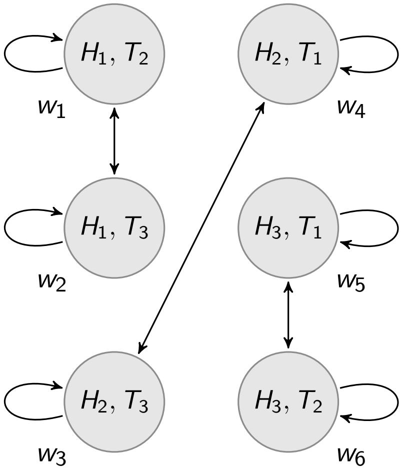


Example

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.

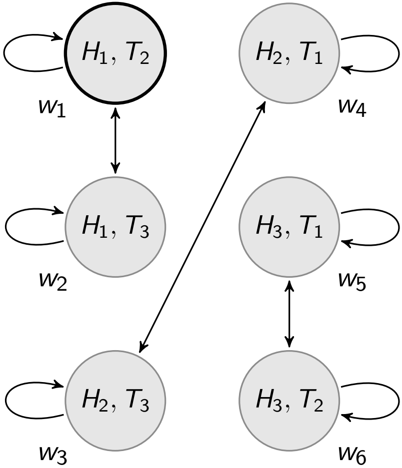


Example

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.

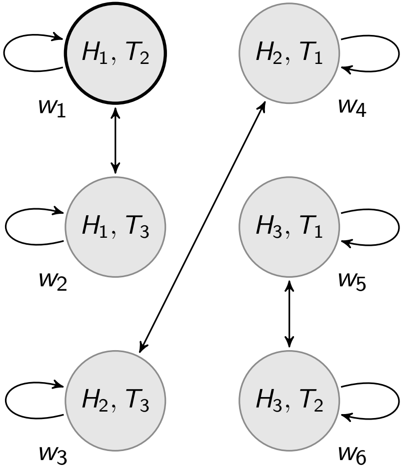


Example

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.

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

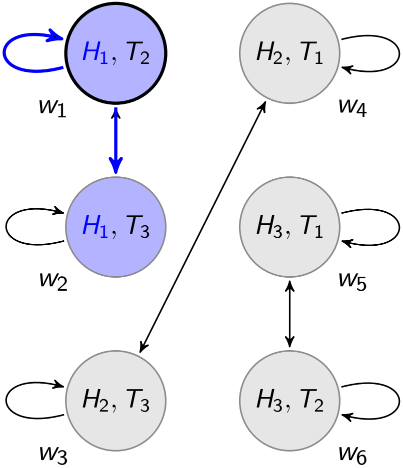


Example

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.

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



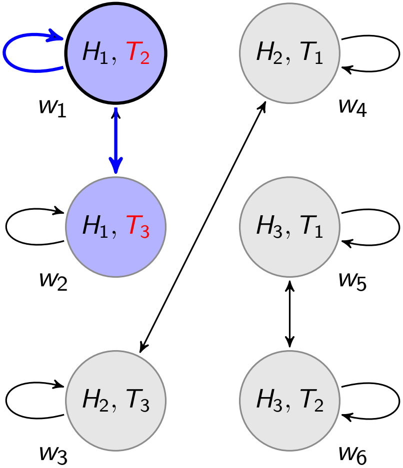
Example

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.

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

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

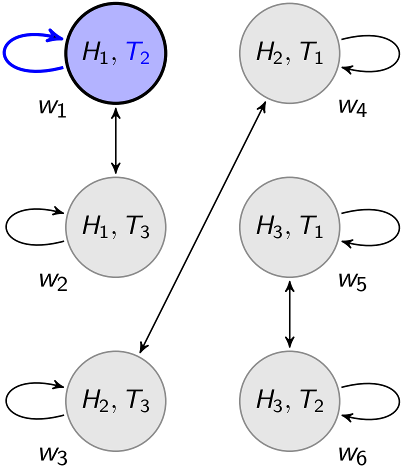


Example

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.

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

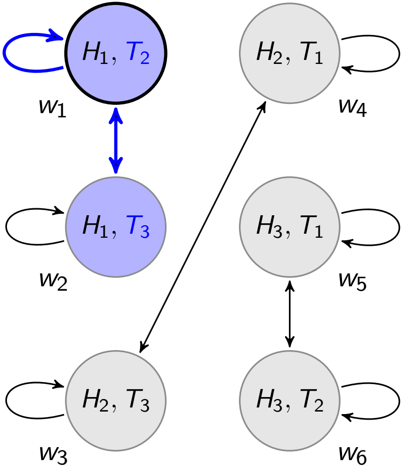


Example

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.

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



Multiagent Epistemic Logic

Many of the examples we are interested in involve more than one agent!

Multiagent Epistemic Logic

Many of the examples we are interested in involve more than one agent!

$K_A P$ means "Ann knows P"

$K_B P$ means "Bob knows P"

Multiagent Epistemic Logic

Many of the examples we are interested in involve more than one agent!

$K_A P$ means "Ann knows P"

$K_B P$ means "Bob knows P"

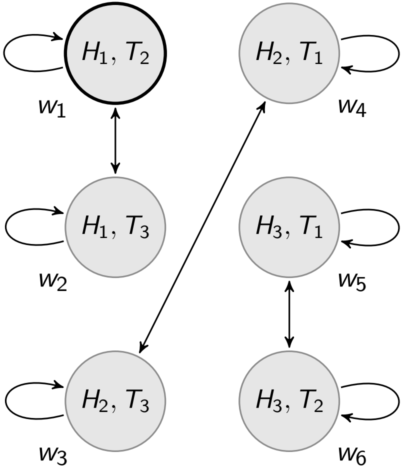
- $K_A K_B \varphi$: "Ann knows that Bob knows φ"
- $K_A(K_B\varphi \vee K_B\neg\varphi)$: "Ann knows that Bob knows whether φ
- $\neg K_B K_A K_B(\varphi)$: "Bob does not know that Ann knows that Bob knows that φ"

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.

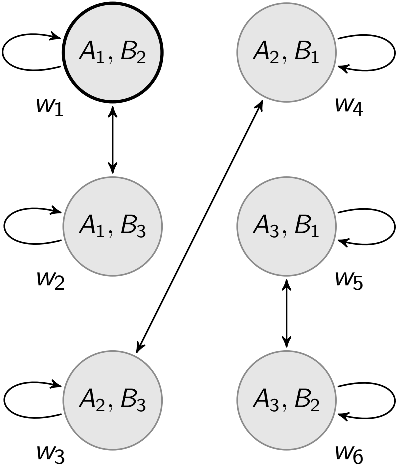


Example

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.

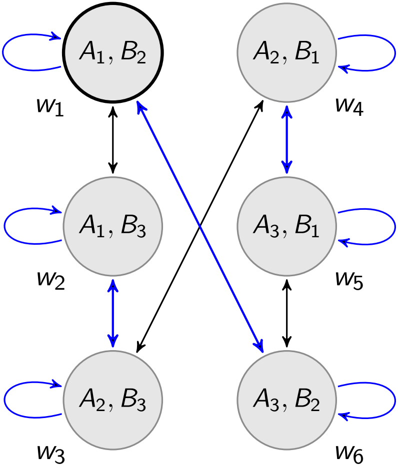


Example

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.

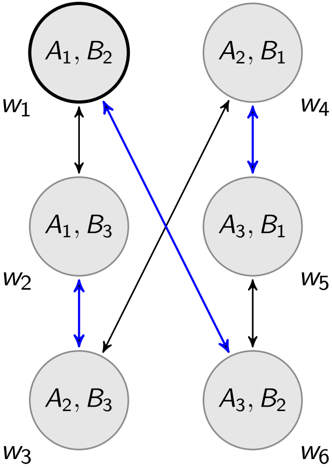


Example

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.



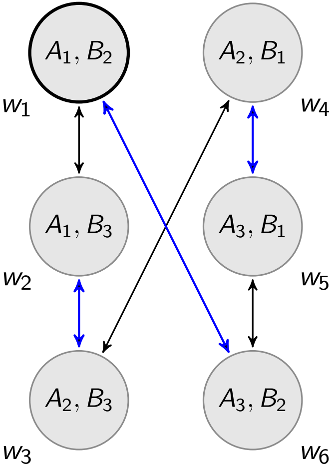
Example

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.

$\mathcal{M}, w_1 \models K_B(K_A A_1 \vee K_A \neg A_1)$



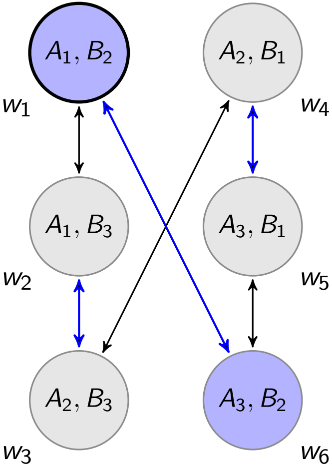
Example

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.

$\mathcal{M}, w_1 \models K_B(K_A A_1 \vee K_A \neg A_1)$



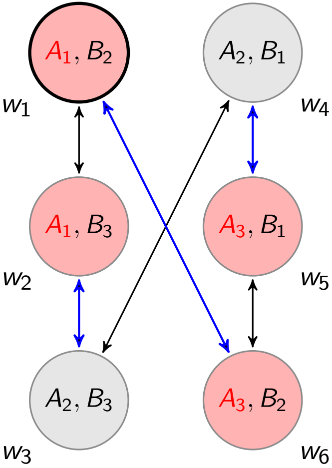
Example

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.

$\mathcal{M}, w_1 \models K_B(K_A A_1 \vee K_A \neg A_1)$



Single-Agent Epistemic Logic: The Language

φ is a formula of Epistemic Logic ($\mathcal{L}$) if it is of the form

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

Single-Agent Epistemic Logic: The Language

φ is a formula of Epistemic Logic ($\mathcal{L}$) if it is of the form

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

- $p \in \mathsf{At}$ is an atomic fact.
  - "It is raining"
  - "The talk is at 2PM"
  - "The card on the table is a 7 of Hearts"

Single-Agent Epistemic Logic: The Language

φ is a formula of Epistemic Logic ($\mathcal{L}$) if it is of the form

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

- $p \in \mathsf{At}$ is an atomic fact.
- The usual propositional language ($\mathcal{L}_0$)

Single-Agent Epistemic Logic: The Language

φ is a formula of Epistemic Logic ($\mathcal{L}$) if it is of the form

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

- $p \in \mathsf{At}$ is an atomic fact.
- The usual propositional language ($\mathcal{L}_0$)
- $K\varphi$ is intended to mean "The agent knows that φ is true".

Single-Agent Epistemic Logic: The Language

φ is a formula of Epistemic Logic ($\mathcal{L}$) if it is of the form

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

- $p \in \mathsf{At}$ is an atomic fact.
- The usual propositional language ($\mathcal{L}_0$)
- $K\varphi$ is intended to mean "The agent knows that φ is true".
- The usual definitions for $\rightarrow, \vee, \leftrightarrow$ apply
- Define $L\varphi$ as $\neg K \neg \varphi$

Single-Agent Epistemic Logic: The Language

φ is a formula of Epistemic Logic ($\mathcal{L}$) if it is of the form

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

$K(p \rightarrow q)$: "Ann knows that p implies q"

$Kp \vee \neg Kp$:

$Kp \vee K\neg p$:

$L\varphi$:

$KL\varphi$:

Single-Agent Epistemic Logic: The Language

φ is a formula of Epistemic Logic ($\mathcal{L}$) if it is of the form

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

$K(p \to q)$: "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"

$L\varphi$:

$KL\varphi$:

Single-Agent Epistemic Logic: The Language

φ is a formula of Epistemic Logic ($\mathcal{L}$) if it is of the form

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

$K(p \to q)$: "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"

$L\varphi$: "φ is an epistemic possibility"

$KL\varphi$: "Ann knows that she thinks φ is possible"

Single-Agent Epistemic Logic: Kripke Models

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

- $W \neq \emptyset$ is the set of all relevant *scenarios* (states of affairs, possible worlds)
- $R \subseteq W \times W$ is the **epistemic accessibility relation**: wRv provided "state v is epistemically accessible for the agent from state w"
- $V : \mathsf{At} \to \wp(W)$ is a valuation function assigning atomic sentences to states

$\mathcal{M}, w \models \varphi$ means "φ is a correct description of some aspect of the scenario w"

$\mathcal{M}, w \models \varphi$ means "φ is a correct description of some aspect of the scenario w"

"An agent is *informed* that φ is true if φ is true throughout the agent's current range of possibilities".

$\mathcal{M}, w \models \varphi$ means "φ is a correct description of some aspect of the scenario w"

"An agent is *informed* that φ is true if φ is true throughout the agent's current range of possibilities".

$\mathcal{M}, w \models K_i\varphi$:

$\mathcal{M}, w \models \varphi$ means "φ is a correct description of some aspect of the scenario w"

"An agent is *informed* that φ is true if φ is true throughout the agent's current range of possibilities".

$\mathcal{M}, w \models K_i\varphi$:

- wR_iv if "everything i knows in state w is true in v
- wR_iv if "agent i has the same experiences and memories in both w and v"
- wR_iv if "agent i has cannot *rule-out* v (given her evidence and observations)"
- wR_iv if "agent i is in the same *local state* in w and v"
- wR_iv if "agent i has the same *information* in w and v"

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

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
$K(\varphi \to \psi) \to (K\varphi \to K\psi)$	—	Logical Omniscience
$K\varphi \to \varphi$	Reflexive	Truth

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

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

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

Unawareness

Why would an agent not know some fact φ? (i.e., why would $\neg K_i\varphi$ be true?)

Unawareness

Why would an agent not know some fact φ? (i.e., why would $\neg K_i \varphi$ be true?)

- The agent may or may not believe φ, but has not ruled out all the $\neg\varphi$-worlds

Unawareness

Why would an agent not know some fact φ? (i.e., why would $\neg K_i\varphi$ be true?)

- The agent may or may not believe φ, but has not ruled out all the $\neg\varphi$-worlds
- The agent may believe φ and ruled-out the $\neg\varphi$-worlds, but this was based on "bad" evidence, or was not justified, or the agent was "epistemically lucky" (eg., Gettier cases),...

Unawareness

Why would an agent not know some fact φ? (i.e., why would $\neg K_i\varphi$ be true?)

- The agent may or may not believe φ, but has not ruled out all the $\neg\varphi$-worlds
- The agent may believe φ and ruled-out the $\neg\varphi$-worlds, but this was based on "bad" evidence, or was not justified, or the agent was "epistemically lucky" (eg., Gettier cases),...
- The agent has not yet entertained possibilities relevant to the truth of φ (the agent is unaware of φ).

Can we model unawareness in state-space models?

Can we model unawareness in state-space models?

E. Dekel, B. Lipman and A. Rustichini. *Standard State-Space Models Preclude Unawareness*. Econometrica, 55:1, pp. 159 - 173 (1998).

Sherlock Holmes

While Watson never reports it, Sherlock Holmes once noted an even more curious incident, that of the dog that barked and the cat that howled in the night.

Sherlock Holmes

While Watson never reports it, Sherlock Holmes once noted an even more curious incident, that of the dog that barked and the cat that howled in the night. When Watson objected that the dog did not bark and the cat did not Howl, Holmes replied "that is the curious incident to which I refer."

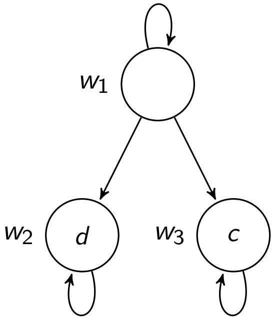
Sherlock Holmes

While Watson never reports it, Sherlock Holmes once noted an even more curious incident, that of the dog that barked and the cat that howled in the night. When Watson objected that the dog did not bark and the cat did not Howl, Holmes replied "that is the curious incident to which I refer." Holmes knew that this meant that no one, neither man nor dog, had intruded on the premises the previous night.

Sherlock Holmes

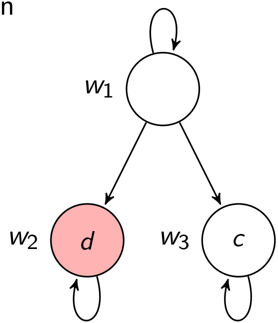
While Watson never reports it, Sherlock Holmes once noted an even more curious incident, that of the dog that barked and the cat that howled in the night. When Watson objected that the dog did not bark and the cat did not Howl, Holmes replied "that is the curious incident to which I refer." Holmes knew that this meant that no one, neither man nor dog, had intruded on the premises the previous night. For had a man intruded, the dog would have barked. Had a dog intruded, the cat would have howled. Hence the lack of either of these two signals means that there could not have been a human or canine intruder.

Modeling Watson's Unawareness



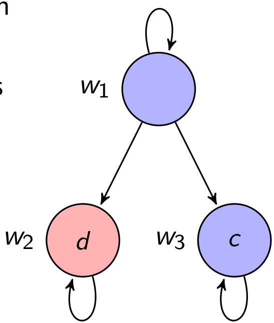
Modeling Watson's Unawareness

- Let $E = \{w_2\}$ be the event that there was an intruder.



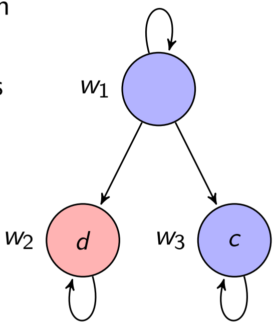
Modeling Watson's Unawareness

- Let $E = \{w_2\}$ be the event that there was an intruder.
- $K(E) = \{w_2\}$ (at w_2, Watson knows there is a human intruder) and $-K(E) = \{w_1, w_3\}$



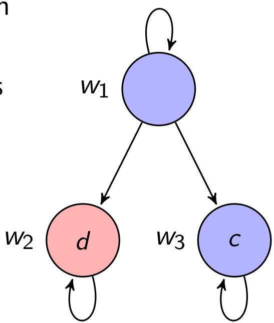
Modeling Watson's Unawareness

- Let $E = \{w_2\}$ be the event that there was an intruder.
- $K(E) = \{w_2\}$ (at w_2, Watson knows there is a human intruder) and $-K(E) = \{w_1, w_3\}$
- $K(-K(E)) = \{w_3\}$ (at w_3, Watson knows that she does not know E), and $-K(-K(E)) = \{w_1, w_2\}$



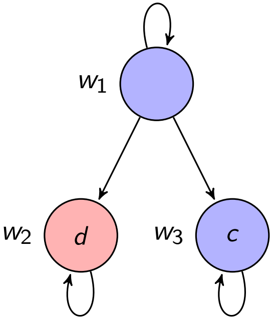
Modeling Watson's Unawareness

- Let $E = \{w_2\}$ be the event that there was an intruder.
- $K(E) = \{w_2\}$ (at w_2, Watson knows there is a human intruder) and $-K(E) = \{w_1, w_3\}$
- $K(-K(E)) = \{w_3\}$ (at w_3, Watson knows that she does not know E), and $-K(-K(E)) = \{w_1, w_2\}$
- $-K(E) \cap -K(-K(E)) = \{w_1\}$ and, in fact, $\bigcap_{i=1}^{\infty}(-K)^i(E) = \{w_1\}$



Modeling Watson's Unawareness

- $E = \{w_2\}$
- $K(E) = \{w_2\}$, $-K(E) = \{w_1, w_3\}$
- $K(-K(E)) = \{w_3\}$,
  $-K(-K(E)) = \{w_1, w_2\}$
- $-K(E) \cap -K(-K(E)) = \{w_1\}$,
  $\bigcap_{i=1}^{\infty}(-K)^i(E) = \{w_1\}$

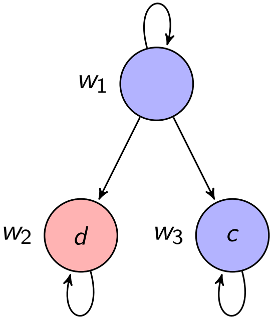


Let $U(F) = \bigcap_{i=1}^{\infty}(-K)^i(F)$. Then,

- $U(\emptyset) = U(W) = U(\{w_1\}) = U(\{w_2, w_3\}) = \emptyset$
- $U(E) = U(\{w_3\}) = U(\{w_1, w_3\}) = U(\{w_1, w_2\} = \{w_1\}$

Modeling Watson's Unawareness

- $E = \{w_2\}$
- $K(E) = \{w_2\}$, $-K(E) = \{w_1, w_3\}$
- $K(-K(E)) = \{w_3\}$, $-K(-K(E)) = \{w_1, w_2\}$
- $-K(E) \cap -K(-K(E)) = \{w_1\}$, $\bigcap_{i=1}^{\infty}(-K)^i(E) = \{w_1\}$



Let $U(F) = \bigcap_{i=1}^{\infty}(-K)^i(F)$. Then,

- $U(\emptyset) = U(W) = U(\{w_1\}) = U(\{w_2, w_3\}) = \emptyset$
- $U(E) = U(\{w_3\}) = U(\{w_1, w_3\}) = U(\{w_1, w_2\} = \{w_1\}$

Then, $U(E) = \{w_1\}$ and $U(U(E)) = U(\{w_1\}) = \emptyset$

Properties of Unawareness

1. $U\varphi \rightarrow (\neg K\varphi \wedge \neg K\neg K\varphi)$

Properties of Unawareness

1. $U\varphi \to (\neg K\varphi \wedge \neg K\neg K\varphi)$
2. $\neg KU\varphi$

Properties of Unawareness

1. $U\varphi \to (\neg K\varphi \wedge \neg K \neg K\varphi)$
2. $\neg KU\varphi$
3. $U\varphi \to UU\varphi$

Properties of Unawareness

1. $U\varphi \to (\neg K\varphi \wedge \neg K \neg K\varphi)$
2. $\neg KU\varphi$
3. $U\varphi \to UU\varphi$

Theorem. In any logic where U satisfies the above axiom schemes, we have

1. If K satisfies Necessitation (from φ infer $K\varphi$), then for all formulas φ, $\neg U\varphi$ is derivable (the agent is aware of everything); and
2. If K satisfies Monotonicity (from $\varphi \to \psi$ infer $K\varphi \to K\psi$), then for all φ and ψ, $U\varphi \to \neg K\psi$ is derivable (if the agent is unaware of something then the agent does not know anything).

B. Schipper. *Online Bibliography on Models of Unawareness*. `http://www.econ.ucdavis.edu/faculty/schipper/unaw.htm`.

J. Halpern. *Alternative semantics for unawareness*. Games and Economic Behavior, 37, 321-339, 2001.

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 $i \in \mathcal{A}$

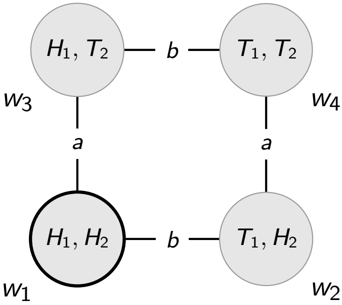
Kripke Models: $\mathcal{M} = \langle W, \{R_i\}_{i\in\mathcal{A}}, V\rangle$ and $w \in W$

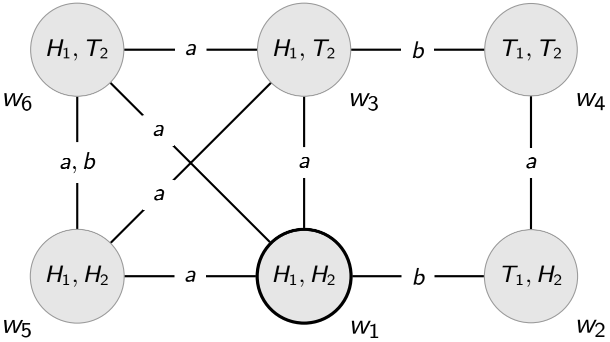
Truth: $\mathcal{M}, w \models \varphi$ is defined as follows:

- $\mathcal{M}, w \models p$ 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

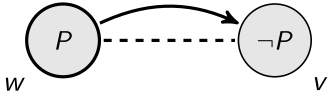
- $K_a K_b \varphi$: "Ann knows that Bob knows φ"
- $K_a(K_b\varphi \vee K_b\neg\varphi)$: "Ann knows that Bob knows whether φ
- $\neg K_b K_a K_b(\varphi)$: "Bob does not know that Ann knows that Bob knows that φ"


H_1, T_2
T_1, T_2
H_1, H_2
T_1, H_2
w_3
w_4
w_1
w_2
a
a
b
b

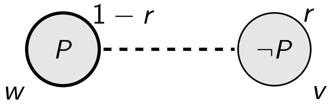


H_1, T_2
w_6
H_1, T_2
w_3
T_1, T_2
w_4
a
b
a
a, b
a
a
a
H_1, H_2
w_5
H_1, H_2
w_1
T_1, H_2
w_2
a
b

Ann does not know that P



Ann does not know that P, but she believes that $\neg P$



Ann does not know that P, but she believes that $\neg P$ is true to degree r.

Combining Logics of Knowledge and Belief

$\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathcal{A}}, \{R_i\}_{i \in \mathcal{A}}, V \rangle$ where

- $W \neq \emptyset$ is a set of states;
- each $\sim_i$ is an equivalence relation on W;
- each R_i is a serial, transitive, Euclidean relation on W; and
- V is a valuation function.

Combining Logics of Knowledge and Belief

$\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{R_i\}_{i\in\mathcal{A}}, V\rangle$ where

- $W \neq \emptyset$ is a set of states;
- each $\sim_i$ is an equivalence relation on W;
- each R_i is a serial, transitive, Euclidean relation on W; and
- V is a valuation function.

What is the relationship between knowledge (K_i) and believe (B_i)?

- Each K_i is **S5**

Combining Logics of Knowledge and Belief

$\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{R_i\}_{i\in\mathcal{A}}, V\rangle$ where

- $W \neq \emptyset$ is a set of states;
- each $\sim_i$ is an equivalence relation on W;
- each R_i is a serial, transitive, Euclidean relation on W; and
- V is a valuation function.

What is the relationship between knowledge (K_i) and believe (B_i)?

- Each K_i is **S5**
- Each B_i is **KD45**

Combining Logics of Knowledge and Belief

$\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathcal{A}}, \{R_i\}_{i \in \mathcal{A}}, V \rangle$ where

- $W \neq \emptyset$ is a set of states;
- each $\sim_i$ is an equivalence relation on W;
- each R_i is a serial, transitive, Euclidean relation on W; and
- V is a valuation function.

What is the relationship between knowledge (K_i) and believe (B_i)?

- Each K_i is **S5**
- Each B_i is **KD45**
- $K_i\varphi \rightarrow B_i\varphi$? "knowledge implies belief"

Combining Logics of Knowledge and Belief

$\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathcal{A}}, \{R_i\}_{i \in \mathcal{A}}, V \rangle$ where

- $W \neq \emptyset$ is a set of states;
- each $\sim_i$ is an equivalence relation on W;
- each R_i is a serial, transitive, Euclidean relation on W; and
- V is a valuation function.

What is the relationship between knowledge (K_i) and believe (B_i)?

- Each K_i is **S5**
- Each B_i is **KD45**
- $K_i\varphi \to B_i\varphi$? "knowledge implies belief"
- $B_i\varphi \to B_iK_i\varphi$? "positive certainty"

Combining Logics of Knowledge and Belief

$\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{R_i\}_{i\in\mathcal{A}}, V\rangle$ where

- $W \neq \emptyset$ is a set of states;
- each $\sim_i$ is an equivalence relation on W;
- each R_i is a serial, transitive, Euclidean relation on W; and
- V is a valuation function.

What is the relationship between knowledge (K_i) and believe (B_i)?

- Each K_i is **S5**
- Each B_i is **KD45**
- $K_i\varphi \to B_i\varphi$? "knowledge implies belief"
- $B_i\varphi \to B_iK_i\varphi$? "positive certainty"
- $B_i\varphi \to K_iB_i\varphi$?

An Issue

- Suppose that p is something you are certain of (you *believe* it with probability one), but is false: $\neg p \wedge Bp$

An Issue

- Suppose that p is something you are certain of (you *believe* it with probability one), but is false: $\neg p \wedge Bp$
- Assuming 1. B satisfies **KD45**, 2. K satisfies **S5**, 3. knowledge implies believe and 4. positive certainty leads to a contradiction.

An Issue

- Suppose that p is something you are certain of (you *believe* it with probability one), but is false: $\neg p \wedge Bp$
- Assuming 1. B satisfies **KD45**, 2. K satisfies **S5**, 3. knowledge implies believe and 4. positive certainty leads to a contradiction.
- $Bp \to BKp$

An Issue

- Suppose that p is something you are certain of (you *believe* it with probability one), but is false: $\neg p \wedge Bp$
- Assuming 1. B satisfies **KD45**, 2. K satisfies **S5**, 3. knowledge implies believe and 4. positive certainty leads to a contradiction.
- $Bp \rightarrow BKp$
- $\neg p \rightarrow \neg Kp$

An Issue

- Suppose that p is something you are certain of (you *believe* it with probability one), but is false: $\neg p \wedge Bp$
- Assuming 1. B satisfies **KD45**, 2. K satisfies **S5**, 3. knowledge implies believe and 4. positive certainty leads to a contradiction.
- $Bp \to BKp$
- $\neg p \to \neg Kp \to K\neg Kp$

An Issue

- Suppose that p is something you are certain of (you *believe* it with probability one), but is false: $\neg p \wedge Bp$
- Assuming 1. B satisfies **KD45**, 2. K satisfies **S5**, 3. knowledge implies believe and 4. positive certainty leads to a contradiction.
- $Bp \rightarrow BKp$
- $\neg p \rightarrow \neg Kp \rightarrow K\neg Kp \rightarrow B\neg Kp$

An Issue

- Suppose that p is something you are certain of (you *believe* it with probability one), but is false: $\neg p \wedge Bp$
- Assuming 1. B satisfies **KD45**, 2. K satisfies **S5**, 3. knowledge implies believe and 4. positive certainty leads to a contradiction.
- $Bp \rightarrow BKp$
- $\neg p \rightarrow \neg Kp \rightarrow K\neg Kp \rightarrow B\neg Kp$
- So, $BKp \wedge B\neg Kp$ also holds, but this contradictions $B\varphi \rightarrow \neg B\neg\varphi$.

J. Halpern. *Should Knowledge Entail Belief?*. Journal of Philosophical Logic, 25:5, 1996, pp. 483-494.

Adding Beliefs

Epistemic Models: $\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathcal{A}}, V \rangle$

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 $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models K_i\varphi$ if for each $v \in W$, if $w \sim_i v$, then $\mathcal{M}, v \models \varphi$

Adding Beliefs

Epistemic-Doxastic Models: $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, V\rangle$

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 $w \sim_i v$, then $\mathcal{M}, v \models \varphi$

Adding Beliefs

Epistemic-Doxastic Models: $\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathcal{A}}, \{\preceq_i\}_{i \in \mathcal{A}}, V\rangle$

Plausibility Relation: $\preceq_i \subseteq W \times W$. $w \preceq_i v$ means

"v is at least as plausibile as w."

Adding Beliefs

Epistemic-Doxastic Models: $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, V\rangle$

Plausibility Relation: $\preceq_i \subseteq W \times W$. $w \preceq_i v$ means

"v is at least as plausibile as w."

Properties of $\preceq_i$: reflexive, transitive, and *well-founded*.

Adding Beliefs

Epistemic-Doxastic Models: $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, V\rangle$

Plausibility Relation: $\preceq_i \subseteq W \times W$. $w \preceq_i v$ means

"v is at least as plausibile as w."

Properties of $\preceq_i$: reflexive, transitive, and *well-founded*.

Most Plausible: For $X \subseteq W$, let

$$Min_{\preceq_i}(X) = \{v \in W \mid v \preceq_i w \text{ for all } w \in X \}$$

Adding Beliefs

Epistemic-Doxastic Models: $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, V\rangle$

Plausibility Relation: $\preceq_i \subseteq W \times W$. $w \preceq_i v$ means

"v is at least as plausibile as w."

Properties of $\preceq_i$: reflexive, transitive, and *well-founded*.

Most Plausible: For $X \subseteq W$, let

$$Min_{\preceq_i}(X) = \{v \in W \mid v \preceq_i w \text{ for all } w \in X\ \}$$

Assumptions:

1. *plausibility implies possibility*: if $w \preceq_i v$ then $w \sim_i v$.
2. *locally-connected*: if $w \sim_i v$ then either $w \preceq_i v$ or $v \preceq_i w$.

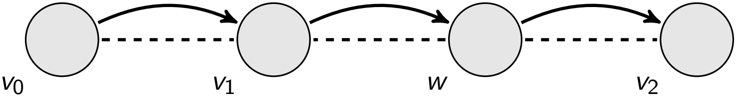
Adding Beliefs

Epistemic-Doxastic Models: $\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathcal{A}}, \{\preceq_i\}_{i \in \mathcal{A}}, V \rangle$

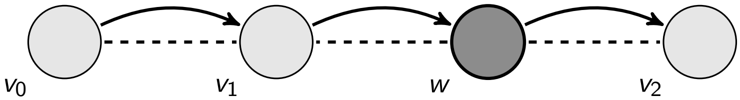
Truth: $\mathcal{M}, w \models \varphi$ is defined as follows:

- $\mathcal{M}, w \models p$ 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 $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models K_i\varphi$ if for each $v \in W$, if $w \sim_i v$, then $\mathcal{M}, v \models \varphi$
- $\mathcal{M}, w \models B_i\varphi$ if for each $v \in Min_{\preceq_i}([w]_i)$, $\mathcal{M}, v \models \varphi$
  $[w]_i = \{v \mid w \sim_i v\}$ is the agent's **information cell**.

Grades of Doxastic Strength

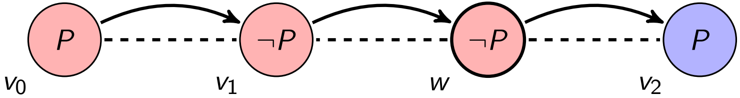


Grades of Doxastic Strength



Suppose that w is the current state.

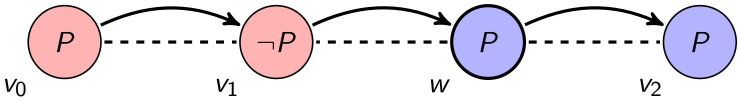
Grades of Doxastic Strength



Suppose that w is the current state.

- **Belief** (BP)

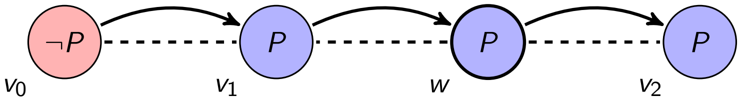
Grades of Doxastic Strength



Suppose that w is the current state.

- **Belief** (BP)
- **Robust Belief** ($\Box P$)

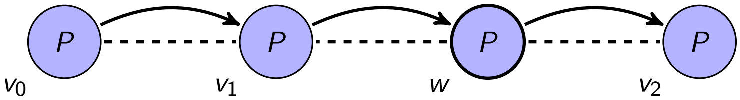
Grades of Doxastic Strength



Suppose that w is the current state.

- **Belief** (BP)
- **Robust Belief** ($\Box P$)
- **Strong Belief** ($B^s P$)

Grades of Doxastic Strength

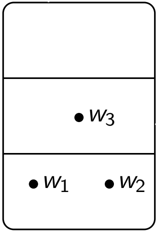


Suppose that w is the current state.

- **Belief** (BP)
- **Robust Belief** ($\Box P$)
- **Strong Belief** ($B^s P$)
- **Knowledge** (KP)

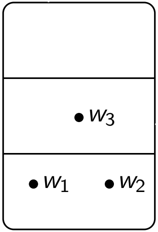
Conditional Beliefs

- $w_1 \sim w_2 \sim w_3$



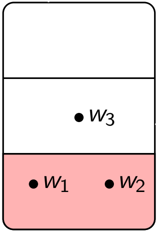
Conditional Beliefs

- $w_1 \sim w_2 \sim w_3$
- $w_1 \preceq w_2$ and $w_2 \preceq w_1$ (w_1 and w_2 are equi-plausbile)
- $w_1 \prec w_3$ ($w_1 \preceq w_3$ and $w_3 \not\preceq w_1$)
- $w_2 \prec w_3$ ($w_2 \preceq w_3$ and $w_3 \not\preceq w_2$)

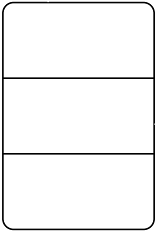


Conditional Beliefs

- $w_1 \sim w_2 \sim w_3$
- $w_1 \preceq w_2$ and $w_2 \preceq w_1$ (w_1 and w_2 are equi-plausbile)
- $w_1 \prec w_3$ ($w_1 \preceq w_3$ and $w_3 \not\preceq w_1$)
- $w_2 \prec w_3$ ($w_2 \preceq w_3$ and $w_3 \not\preceq w_2$)
- $\{w_1, w_2\} \subseteq Min_{\preceq}([w_i])$

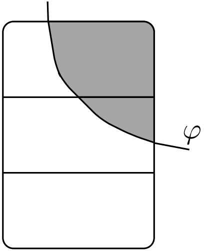


Conditional Beliefs



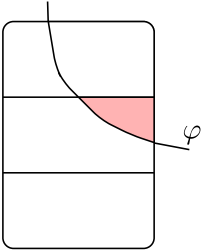
$B_i^{\varphi}\psi$: Agent i believes ψ, *given that φ is true*.

Conditional Beliefs



$B_i^{\varphi}\psi$: Agent i believes ψ, *given that φ is true*.

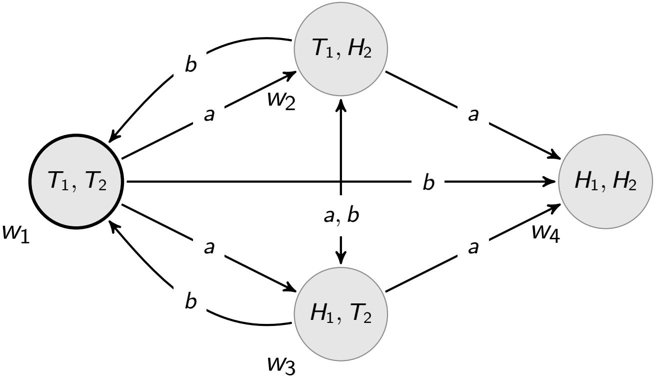
Conditional Beliefs



$B_i^{\varphi}\psi$: Agent i believes ψ, *given that φ is true*.

$\mathcal{M}, w \models B_i^{\varphi}\psi$ if for each $v \in Min_{\preceq_i}([w]_i \cap [\![\varphi]\!])$, $\mathcal{M}, v \models \varphi$
where $[\![\varphi]\!] = \{w \mid \mathcal{M}, w \models \varphi\}$

Example



Success:	$B_i^{\varphi}\varphi$
Knowledge entails belief	$K_i\varphi \to B_i^{\psi}\varphi$
Full introspection:	$B_i^{\varphi}\psi \to K_i B_i^{\varphi}\psi$ and $\neg B_i^{\varphi}\psi \to K_i \neg B_i^{\varphi}\psi$
Cautious Monotonicity:	$(B_i^{\varphi}\alpha \land B_i^{\varphi}\beta) \to B_i^{\varphi\land\beta}\alpha$
Rational Monotonicity:	$(B_i^{\varphi}\alpha \land \neg B_i^{\varphi}\neg\beta) \to B_i^{\varphi\land\beta}\alpha$

Rational Monotonicity, I

Rational Monotonicity: $(B_i^{\varphi}\alpha \wedge \neg B_i^{\varphi}\neg\beta) \rightarrow B_i^{\varphi\wedge\beta}\alpha$

R. Stalnaker. *Nonmonotonic consequence relations*. Fundamenta Informaticae, 21: 721, 1994.

Rational Monotonicity, I

Rational Monotonicity: $(B_i^{\varphi}\alpha \wedge \neg B_i^{\varphi}\neg\beta) \rightarrow B_i^{\varphi\wedge\beta}\alpha$

R. Stalnaker. *Nonmonotonic consequence relations*. Fundamenta Informaticae, 21: 721, 1994.

Consider the three composers: Verdi, Bizet, and Satie, and suppose that we initially accept (correctly but defeasibly) that Verdi is Italian $I(v)$, while Bizet and Satie are French $(F(b) \wedge F(s))$.

Rational Monotonicity, II

Suppose now that we are told by a reliable (but not infallible!) source of information that that Verdi and Bizet are compatriots ($C(v, b)$). This leads us no longer to endorse either the proposition that Verdi is Italian (because he could be French), or that Bizet is French (because he could be Italian); but we would still draw the defeasible consequence that Satie is French, since nothing that we have learned conflicts with it.

$$B^{C(v,b)}F(s)$$

Rational Monotonicity, III

Now consider the proposition $C(v, s)$ that Verdi and Satie are compatriots. Before learning that $C(v, b)$ we would be inclined to reject the proposition $C(v, s)$ because we accept $I(v)$ and $F(s)$, but after learning that Verdi and Bizet are compatriots, we can no longer endorse $I(v)$, and therefore no longer reject $C(v, s)$.

$$\neg B^{C(v,b)} \neg C(v, s)$$

Rational Monotonicity, IV

However, if we added $C(v, s)$ to our stock of beliefs, we would lose the inference to $F(s)$: in the context of $C(v, b)$, the proposition $C(v, s)$ is equivalent to the statement that all three composers have the same nationality. This leads us to suspend our belief in the proposition $F(s)$.

$$\neg B^{C(v,b) \wedge C(v,s)} F(s)$$

Rational Monotonicity, IV

However, if we added $C(v, s)$ to our stock of beliefs, we would lose the inference to $F(s)$: in the context of $C(v, b)$, the proposition $C(v, s)$ is equivalent to the statement that all three composers have the same nationality. This leads us to suspend our belief in the proposition $F(s)$.

$$\neg B^{C(v,b) \wedge C(v,s)} F(s)$$

$B^{C(v,b)} F(s)$ and $\neg B^{C(v,b)} \neg C(v, s)$ but $\neg B^{C(v,b) \wedge C(v,s)} F(s)$

Next: Common Knowledge