Slides thanks to

- Martin Henz Aquinas Hobor
- CS 3234: Logic and Formal Systems

Notions of Truth

- Often, it is not enough to distinguish between "true" and "false".
- We need to consider *modalities* if truth, such as:
  - necessity
  - time
  - knowledge by an agent
- Modal logic constructs a framework using which modalities can be formalized and reasoning methods can be established.

Syntax of Basic Modal Logic

$$
\begin{aligned}
\phi \quad ::= \quad & \top \mid \bot \mid p \mid (\neg\phi) \mid (\phi \wedge \phi) \\
& \mid (\phi \vee \phi) \mid (\phi \rightarrow \phi) \\
& \mid (\phi \leftrightarrow \phi) \\
& \mid (\Box\phi) \mid (\Diamond\phi)
\end{aligned}
$$

Kripke Models

Definition

A model $\mathcal{M}$ of propositional modal logic over a set of propositional atoms A is specified by three things:

1. A W of *worlds*;

Kripke Models

Definition

A model $\mathcal{M}$ of propositional modal logic over a set of propositional atoms A is specified by three things:

1. A W of *worlds*;
2. a relation R on W, meaning $R \subseteq W \times W$, called the *accessibility relation*;

Kripke Models

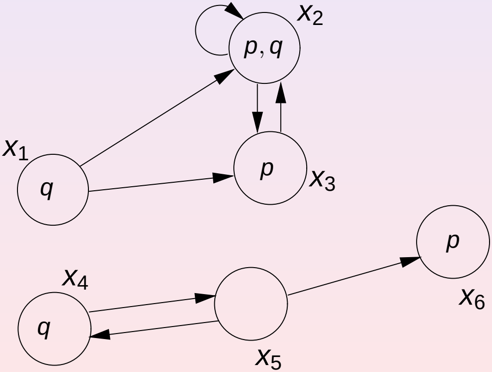
Definition

A model $\mathcal{M}$ of propositional modal logic over a set of propositional atoms A is specified by three things:

1. A W of *worlds*;
2. a relation R on W, meaning $R \subseteq W \times W$, called the *accessibility relation*;
3. a function $L : W \to A \to \{T, F\}$, called *labeling function*.

Example

$$W = \{x_1, x_2, x_3, x_4, x_5, x_6\}$$
$$R = \{(x_1, x_2), (x_1, x_3), (x_2, x_2), (x_2, x_3), (x_3, x_2), (x_4, x_5), (x_5, x_4), (x_5, x_6)\}$$
$$L = \{(x_1, \{q\}), (x_2, \{p, q\}), (x_3, \{p\}), (x_4, \{q\}), (x_5, \{\}), (x_6, \{p\})\}$$



When is a formula true in a possible world?

Definition

Let $\mathcal{M} = (W, R, L)$, $x \in W$, and ϕ a formula in basic modal logic. We define $x \Vdash \phi$ via structural induction:

- $x \Vdash \top$
- $x \nVdash \bot$
- $x \Vdash p$ iff $p \in L(x)(p) = T$
- $x \Vdash \neg\phi$ iff $x \nVdash \phi$
- $x \Vdash \phi \wedge \psi$ iff $x \Vdash \phi$ and $x \Vdash \psi$
- $x \Vdash \phi \vee \psi$ iff $x \Vdash \phi$ or $x \Vdash \psi$
- ...

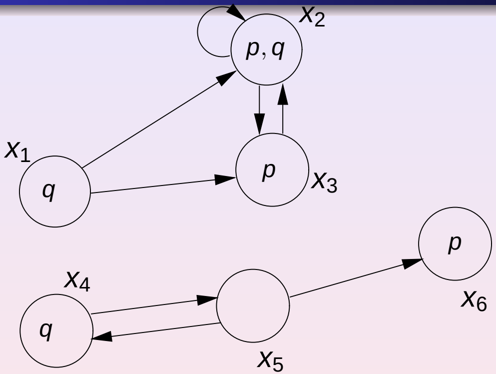
When is a formula true in a possible world?

Definition (continued)

Let $\mathcal{M} = (W, R, L)$, $x \in W$, and ϕ a formula in basic modal logic. We define $x \Vdash \phi$ via structural induction:

- ...
- $x \Vdash \phi \rightarrow \psi$ iff $x \Vdash \psi$, whenever $x \Vdash \phi$
- $x \Vdash \phi \leftrightarrow \psi$ iff ($x \Vdash \phi$ iff $x \Vdash \psi$)
- $x \Vdash \Box\phi$ iff for each $y \in W$ with $R(x, y)$, we have $y \Vdash \phi$
- $x \Vdash \Diamond\phi$ iff there is a $y \in W$ such that $R(x, y)$ and $y \Vdash \phi$.

Example



- $x_1 \Vdash q$
- $x_1 \Vdash \Diamond q$, $x_1 \nVdash \Box q$
- $x_5 \nVdash \Box p$, $x_5 \nVdash \Box q$, $x_5 \nVdash \Box p \vee \Box q$, $x_5 \Vdash \Box(p \vee q)$
- $x_6 \Vdash \Box \phi$ holds for all ϕ, but $x_6 \nVdash \Diamond \phi$

A Range of Modalities

In a particular context $\Box\phi$ could mean:

- It is necessarily true that ϕ
- It will always be true that ϕ
- It ought to be that ϕ
- Agent Q believes that ϕ
- Agent Q knows that ϕ
- After any execution of program P, ϕ holds.

Since $\Diamond\phi \equiv \neg\Box\neg\phi$, we can infer the meaning of $\Diamond$ in each context.

A Range of Modalities

From the meaning of $\Box\phi$, we can conclude the meaning of $\Diamond\phi$, since $\Diamond\phi \equiv \neg\Box\neg\phi$:

$\Box\phi$	$\Diamond\phi$
It is necessarily true that ϕ	It is possibly true that ϕ
It will always be true that ϕ	Sometime in the future ϕ
It ought to be that ϕ	It is permitted to be that ϕ
Agent Q believes that ϕ	ϕ is consistent with Q's beliefs
Agent Q knows that ϕ	For all Q knows, ϕ
After any run of P, ϕ holds.	After some run of P, ϕ holds

Formula Schemes that hold wrt some Modalities

$\Box\phi$	$\Box\phi \to \phi$	$\Box\phi \to \Box\Box\phi$	$\Diamond\phi \to \Box\Diamond\phi$	$\Diamond\top$	$\Box\phi \to \Diamond\phi$	$\Box\phi \lor \Box\neg\phi$	$\Box(\phi \to \psi) \land \Box\phi \to \Box\psi$	$\Diamond\phi \land \Diamond\psi \to \Diamond(\phi \land \psi)$
It is necessary that ϕ	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\times$	$\surd$	$\times$
It will always be that ϕ	$\times$	$\surd$	$\times$	$\times$	$\times$	$\times$	$\surd$	$\times$
It ought to be that ϕ	$\times$	$\times$	$\times$	$\surd$	$\surd$	$\times$	$\surd$	$\times$
Agent Q believes that ϕ	$\times$	$\surd$	$\surd$	$\surd$	$\surd$	$\times$	$\surd$	$\times$
Agent Q knows that ϕ	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\times$	$\surd$	$\times$
After running P, ϕ	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\surd$	$\times$

Modalities lead to Interpretations of R

$\Box\phi$	$R(x,y)$
It is necessarily true that ϕ	y is possible world according to info at x
It will always be true that ϕ	y is a future world of x
It ought to be that ϕ	y is an acceptable world according to the information at x
Agent Q believes that ϕ	y could be the actual world according to Q's beliefs at x
Agent Q knows that ϕ	y could be the actual world according to Q's knowledge at x
After any execution of P, ϕ holds	y is a possible resulting state after execution of P at x

Possible Properties of R

- reflexive: for every $w \in W$, we have $R(x,x)$.
- symmetric: for every $x,y \in W$, we have $R(x,y)$ implies $R(y,x)$.
- serial: for every x there is a y such that $R(x,y)$.
- transitive: for every $x,y,z \in W$, we have $R(x,y)$ and $R(y,z)$ imply $R(x,z)$.
- Euclidean: for every $x,y,z \in W$ with $R(x,y)$ and $R(x,z)$, we have $R(y,z)$.
- functional: for each x there is a unique y such that $R(x,y)$.
- linear: for every $x,y,z \in W$ with $R(x,y)$ and $R(x,z)$, we have $R(y,z)$ or $y = z$ or $R(z,y)$.
- total: for every $x,y \in W$, we have $R(x,y)$ and $R(y,x)$.
- equivalence: reflexive, symmetric and transitive.

Reflexivity and Transitivity

Theorem

The following statements are equivalent:

- R is reflexive;
- $\mathcal{F}$ satisfies $\Box\phi \to \phi$;
- $\mathcal{F}$ satisfies $\Box p \to p$;

Theorem

The following statements are equivalent:

- R is transitive;
- $\mathcal{F}$ satisfies $\Box\phi \to \Box\Box\phi$;
- $\mathcal{F}$ satisfies $\Box p \to \Box\Box p$;

Formula Schemes and Properties of R

name	formula scheme	property of R
T	$\Box\phi \to \phi$	reflexive
B	$\phi \to \Box\Diamond\phi$	symmetric
D	$\Box\phi \to \Diamond\phi$	serial
4	$\Box\phi \to \Box\Box\phi$	transitive
5	$\Diamond\phi \to \Box\Diamond\phi$	Euclidean
	$\Box\phi \leftrightarrow \Diamond\phi$	functional
	$\Box(\phi \land \Box\phi \to \psi) \lor \Box(\psi \land \Box\psi \to \phi)$	linear

Which Formula Schemes to Choose?

Definition

Let $\mathcal{L}$ be a set of formula schemes and $\Gamma \cup \{\psi\}$ a set of formulas of basic modal logic.

Which Formula Schemes to Choose?

Definition

Let $\mathcal{L}$ be a set of formula schemes and $\Gamma \cup \{\psi\}$ a set of formulas of basic modal logic.

- A set of formula schemes is said to be *closed* iff it contains all substitution instances of its elements.

Which Formula Schemes to Choose?

Definition

Let $\mathcal{L}$ be a set of formula schemes and $\Gamma \cup \{\psi\}$ a set of formulas of basic modal logic.

- A set of formula schemes is said to be *closed* iff it contains all substitution instances of its elements.
- Let $\mathcal{L}_c$ be the smallest closed superset of $\mathcal{L}$.

Which Formula Schemes to Choose?

Definition

Let $\mathcal{L}$ be a set of formula schemes and $\Gamma \cup \{\psi\}$ a set of formulas of basic modal logic.

- A set of formula schemes is said to be *closed* iff it contains all substitution instances of its elements.
- Let $\mathcal{L}_c$ be the smallest closed superset of $\mathcal{L}$.
- Γ entails ψ in $\mathcal{L}$ iff $\Gamma \cup \mathcal{L}_c$ semantically entails ψ. We say $\Gamma \models_{\mathcal{L}} \psi$.

Examples of Modal Logics: K

K is the weakest modal logic, $\mathcal{L} = \emptyset$.

Examples of Modal Logics: KT45

$\mathcal{L} = \{T, 4, 5\}$

Examples of Modal Logics: KT45

$\mathcal{L} = \{T, 4, 5\}$

Used for reasoning about knowledge.

Examples of Modal Logics: KT45

$\mathcal{L} = \{T, 4, 5\}$

Used for reasoning about knowledge.

name	formula scheme	property of R
T	$\Box\phi \rightarrow \phi$	reflexive
4	$\Box\phi \rightarrow \Box\Box\phi$	transitive
5	$\Diamond\phi \rightarrow \Box\Diamond\phi$	Euclidean

Examples of Modal Logics: KT45

$\mathcal{L} = \{T, 4, 5\}$

Used for reasoning about knowledge.

name	formula scheme	property of R
T	$\Box\phi \rightarrow \phi$	reflexive
4	$\Box\phi \rightarrow \Box\Box\phi$	transitive
5	$\Diamond\phi \rightarrow \Box\Diamond\phi$	Euclidean

- T: Truth: agent Q only knows true things.
- 4: Positive introspection: If Q knows something, he knows that he knows it.
- 5: Negative introspection: If Q doesn't know something, he knows that he doesn't know it.

Explanation of Negative Introspection

name	formula scheme	property of R
...	...	...
5	$\Diamond\phi \to \Box\Diamond\phi$	Euclidean

Explanation of Negative Introspection

name	formula scheme	property of R
...	...	...
5	$\Diamond\phi \rightarrow \Box\Diamond\phi$	Euclidean

$$\Diamond\phi \quad \rightarrow \quad \Box\Diamond\phi$$

Explanation of Negative Introspection

name	formula scheme	property of R
...	...	...
5	$\Diamond\phi \rightarrow \Box\Diamond\phi$	Euclidean

$$\begin{array}{rcl} \Diamond\phi & \rightarrow & \Box\Diamond\phi \\ \Diamond\neg\psi & \rightarrow & \Box\Diamond\neg\psi \end{array}$$

Explanation of Negative Introspection

name	formula scheme	property of R
...	...	...
5	$\Diamond\phi \to \Box\Diamond\phi$	Euclidean

$$\begin{array}{rcl} \Diamond\phi & \to & \Box\Diamond\phi \\ \Diamond\neg\psi & \to & \Box\Diamond\neg\psi \\ \neg\Box\neg\neg\psi & \to & \Box\neg\Box\neg\neg\psi \end{array}$$

Explanation of Negative Introspection

name	formula scheme	property of R
...	...	...
5	$\Diamond\phi \rightarrow \Box\Diamond\phi$	Euclidean

$$\begin{array}{rcl} \Diamond\phi & \rightarrow & \Box\Diamond\phi \\ \Diamond\neg\psi & \rightarrow & \Box\Diamond\neg\psi \\ \neg\Box\neg\neg\psi & \rightarrow & \Box\neg\Box\neg\neg\psi \\ \neg\Box\psi & \rightarrow & \Box\neg\Box\psi \end{array}$$

Explanation of Negative Introspection

name	formula scheme	property of *R*
...	...	...
5	$\Diamond\phi \to \Box\Diamond\phi$	Euclidean

$$
\begin{array}{rcl}
\Diamond\phi & \to & \Box\Diamond\phi \\
\Diamond\neg\psi & \to & \Box\Diamond\neg\psi \\
\neg\Box\neg\neg\psi & \to & \Box\neg\Box\neg\neg\psi \\
\neg\Box\psi & \to & \Box\neg\Box\psi
\end{array}
$$

If *Q* doesn't know ψ, he knows that he doesn't know ψ.

Correspondence for KT45

Accessibility relations for KT45

KT45 hold if and only if R is reflexive (T), transitive (4) and Euclidean (5).

Correspondence for KT45

Accessibility relations for KT45

KT45 hold if and only if R is reflexive (T), transitive (4) and Euclidean (5).

Fact on such relations

A relation is reflexive, transitive and Euclidean iff it is reflexive, transitive and symmetric, i.e. iff it is an equivalence relation.

Collapsing Modalities

Theorem

Any sequence of modal operators and negations is KT45 is equivalent to one of the following: $-$, $\Box$, $\Diamond$, $\neg$, $\neg\Box$, and $\neg\Diamond$, where $-$ indicates the absence of any negation or modality.

Collapsing Modalities

Theorem

Any sequence of modal operators and negations is KT45 is equivalent to one of the following: $-$, $\Box$, $\Diamond$, $\neg$, $\neg\Box$, and $\neg\Diamond$, where $-$ indicates the absence of any negation or modality.

Examples

- $\Box\Box\phi \equiv \Box\phi$

Collapsing Modalities

Theorem

Any sequence of modal operators and negations is KT45 is equivalent to one of the following: $-$, $\Box$, $\Diamond$, $\neg$, $\neg\Box$, and $\neg\Diamond$, where $-$ indicates the absence of any negation or modality.

Examples

- $\Box\Box\phi \equiv \Box\phi$
- $\Diamond\Box\phi \equiv \Diamond\phi$

Collapsing Modalities

Theorem

Any sequence of modal operators and negations is KT45 is equivalent to one of the following: $-$, $\Box$, $\Diamond$, $\neg$, $\neg\Box$, and $\neg\Diamond$, where $-$ indicates the absence of any negation or modality.

Examples

- $\Box\Box\phi \equiv \Box\phi$
- $\Diamond\Box\phi \equiv \Diamond\phi$
- $\neg\Diamond\neg\phi \equiv \Box\phi$

Examples of Modal Logics: KT4

$\mathcal{L} = \{T, 4\}$

Used for partial evaluation in computer science.

Examples of Modal Logics: KT4

$\mathcal{L} = \{T, 4\}$

Used for partial evaluation in computer science.

name	formula scheme	property of R
T	$\Box\phi \rightarrow \phi$	reflexive
4	$\Box\phi \rightarrow \Box\Box\phi$	transitive

Examples of Modal Logics: KT4

$\mathcal{L} = \{T, 4\}$

Used for partial evaluation in computer science.

name	formula scheme	property of R
T	$\Box\phi \rightarrow \phi$	reflexive
4	$\Box\phi \rightarrow \Box\Box\phi$	transitive

- T: Truth: agent Q only knows true things.

Examples of Modal Logics: KT4

$\mathcal{L} = \{T, 4\}$

Used for partial evaluation in computer science.

name	formula scheme	property of R
T	$\Box\phi \to \phi$	reflexive
4	$\Box\phi \to \Box\Box\phi$	transitive

- T: Truth: agent Q only knows true things.
- 4: Positive introspection: If Q knows something, he knows that he knows it.

Correspondence for KT4

Accessibility relations for KT4

KT4 hold if and only if R is reflexive (T), and transitive (4).

Correspondence for KT4

Accessibility relations for KT4

KT4 hold if and only if R is reflexive (T), and transitive (4).

Definition

A reflexive and transitive relation is called a *preorder*.

Collapsing Modalities

Theorem

Any sequence of modal operators and negations is KT4 is equivalent to one of the following:
$-$, $\Box$, $\Diamond$, $\Box\Diamond$, $\Diamond\Box$, $\Box\Diamond\Box$, $\Diamond\Box\Diamond$, $\neg$, $\neg\Box$, $\neg\Diamond$, $\neg\Box\Diamond$, $\neg\Diamond\Box$, $\neg\Box\Diamond\Box$, and $\neg\Diamond\Box\Diamond$.

Connection to Intuitionistic Logic

Definition

A model of intuitionistic propositional logic is a model $\mathcal{M} = (W, R, L)$ of KT4 such that $R(x, y)$ always implies $L(x)(p) \to L(y)(p)$.

Satisfaction in Intuitionistic Logic

Definition

We change the definition of $x \Vdash \phi$ as follows:

- $x \Vdash \top$
- $x \nVdash \bot$
- $x \Vdash p$ iff $p \in L(x)$
- $x \Vdash \phi \wedge \psi$ iff $x \Vdash \phi$ and $x \Vdash \psi$
- $x \Vdash \phi \vee \psi$ iff $x \Vdash \phi$ or $x \Vdash \psi$

as usual,

Satisfaction in Intuitionistic Logic

Definition

We change the definition of $x \Vdash \phi$ as follows:

- $x \Vdash \top$
- $x \nVdash \bot$
- $x \Vdash p$ iff $p \in L(x)$
- $x \Vdash \phi \wedge \psi$ iff $x \Vdash \phi$ and $x \Vdash \psi$
- $x \Vdash \phi \vee \psi$ iff $x \Vdash \phi$ or $x \Vdash \psi$

as usual, but now:

- $x \Vdash \neg\phi$ **iff** for all y with $R(x, y)$, we have $y \nVdash \phi$

Satisfaction in Intuitionistic Logic

Definition

We change the definition of $x \Vdash \phi$ as follows:

- $x \Vdash \top$
- $x \nVdash \bot$
- $x \Vdash p$ iff $p \in L(x)$
- $x \Vdash \phi \wedge \psi$ iff $x \Vdash \phi$ and $x \Vdash \psi$
- $x \Vdash \phi \vee \psi$ iff $x \Vdash \phi$ or $x \Vdash \psi$

as usual, but now:

- $x \Vdash \neg\phi$ **iff** for all y with $R(x, y)$, we have $y \nVdash \phi$
- $x \Vdash \phi \rightarrow \psi$ **iff** for all y with $R(x, y)$, we have $y \Vdash \psi$ whenever we have $y \Vdash \phi$.

Example

Let $W = \{x, y\}$, $R = \{(x, x), (x, y), (y, y)\}$,
$L(x)(p) = F, L(y)(p) = T$.

Example

Let $W = \{x, y\}$, $R = \{(x, x), (x, y), (y, y)\}$, $L(x)(p) = F, L(y)(p) = T$. Does $x \Vdash p \vee \neg p$ hold?

Example

Let $W = \{x, y\}$, $R = \{(x, x), (x, y), (y, y)\}$, $L(x)(p) = F, L(y)(p) = T$. Does $x \Vdash p \vee \neg p$ hold?
Since

- $x \Vdash \phi \vee \psi$ iff $x \Vdash \phi$ or $x \Vdash \psi$

we would need: $x \Vdash \neg p$.

Example

Let $W = \{x, y\}$, $R = \{(x, x), (x, y), (y, y)\}$, $L(x)(p) = F, L(y)(p) = T$. Does $x \Vdash p \vee \neg p$ hold?
Since

- $x \Vdash \phi \vee \psi$ iff $x \Vdash \phi$ or $x \Vdash \psi$

we would need: $x \Vdash \neg p$.
Since

- $x \Vdash \neg\phi$ **iff** for all y with $R(x, y)$, we have $y \nVdash \phi$

we cannot establish $x \Vdash \neg p$.

Example

Let $W = \{x, y\}$, $R = \{(x, x), (x, y), (y, y)\}$,
$L(x)(p) = F, L(y)(p) = T$. Does $x \Vdash p \vee \neg p$ hold?
Since

- $x \Vdash \phi \vee \psi$ iff $x \Vdash \phi$ or $x \Vdash \psi$

we would need: $x \Vdash \neg p$.
Since

- $x \Vdash \neg\phi$ **iff** for all y with $R(x, y)$, we have $y \nVdash \phi$

we cannot establish $x \Vdash \neg p$.

Idea

Do not allow "assumptions", even if they exhaust all possibilities.

Dashed Boxes

Idea

In addition to proof boxes for assumptions, we introduce *blue boxes* that express knowledge about an *arbitrary accessible world*.

Dashed Boxes

Idea

In addition to proof boxes for assumptions, we introduce *blue boxes* that express knowledge about an *arbitrary accessible world*.

Rules about blue boxes

Dashed Boxes

Idea

In addition to proof boxes for assumptions, we introduce *blue boxes* that express knowledge about an *arbitrary accessible world*.

Rules about blue boxes

- Whenever $\Box\phi$ occurs in a proof, ϕ may be put into a subsequent blue box.

Dashed Boxes

Idea

In addition to proof boxes for assumptions, we introduce *blue boxes* that express knowledge about an *arbitrary accessible world*.

Rules about blue boxes

- Whenever $\Box\phi$ occurs in a proof, ϕ may be put into a subsequent blue box.
- Whenever ϕ occurs at the end of a blue box, $\Box\phi$ may be put after that blue box.

Rules for □

Introduction of □:

$$\frac{\boxed{\begin{array}{c} \vdots \\ \phi \end{array}}}{\Box\phi}\ [\Box\ i]$$

Rules for $\Box$

Elimination of $\Box$:

$$\frac{\Box\phi}{\boxed{\begin{array}{c}\vdots\\ \phi\\ \vdots\end{array}}}\;[\Box\ e]$$

Extra Rules for KT45

$$\frac{\Box\phi}{\phi}\,[T] \qquad \frac{\Box\phi}{\Box\Box\phi}\,[4] \qquad \frac{\neg\Box\phi}{\Box\neg\Box\phi}\,[5]$$

Example Proof

$$\vdash_K \Box p \wedge \Box q \rightarrow \Box(p \wedge q)$$

1	$\Box p \wedge \Box q$	assumption
2	$\Box p$	$\wedge e_1$ 1
3	$\Box q$	$\wedge e_2$ 1
4	p	$\Box e$ 2
5	q	$\Box e$ 3
6	$p \wedge q$	$\wedge i$ 4,5
7	$\Box(p \wedge q)$	$\Box i$ 4–6
8	$\Box p \wedge \Box q \rightarrow \Box(p \wedge q)$	$\rightarrow i$ 1–7

Wise Women Puzzle

- Three wise women, each wearing one hat, among three available red hats and *two* available white hats

Wise Women Puzzle

- Three wise women, each wearing one hat, among three available red hats and *two* available white hats
- Each wise woman is wise, can see other hats but not her own

Wise Women Puzzle

- Three wise women, each wearing one hat, among three available red hats and *two* available white hats
- Each wise woman is wise, can see other hats but not her own
- Queen asks first wise woman: Do you know the color of your hat.

Wise Women Puzzle

- Three wise women, each wearing one hat, among three available red hats and *two* available white hats
- Each wise woman is wise, can see other hats but not her own
- Queen asks first wise woman: Do you know the color of your hat.
  Answer: No

Wise Women Puzzle

- Three wise women, each wearing one hat, among three available red hats and *two* available white hats
- Each wise woman is wise, can see other hats but not her own
- Queen asks first wise woman: Do you know the color of your hat.
  Answer: No
- Queen asks second wise woman: Do you know the color of your hat.

Wise Women Puzzle

- Three wise women, each wearing one hat, among three available red hats and *two* available white hats
- Each wise woman is wise, can see other hats but not her own
- Queen asks first wise woman: Do you know the color of your hat.
  Answer: No
- Queen asks second wise woman: Do you know the color of your hat.
  Answer: No

Wise Women Puzzle

- Three wise women, each wearing one hat, among three available red hats and *two* available white hats
- Each wise woman is wise, can see other hats but not her own
- Queen asks first wise woman: Do you know the color of your hat.
  Answer: No
- Queen asks second wise woman: Do you know the color of your hat.
  Answer: No
- Queen asks third wise woman: Do you know the color of your hat?

Wise Women Puzzle

- Three wise women, each wearing one hat, among three available red hats and *two* available white hats
- Each wise woman is wise, can see other hats but not her own
- Queen asks first wise woman: Do you know the color of your hat.
  Answer: No
- Queen asks second wise woman: Do you know the color of your hat.
  Answer: No
- Queen asks third wise woman: Do you know the color of your hat?
- What is her answer?

Motivation

Reasoning about knowledge

We saw that KT45 can be used to reason about an agent's knowledge.

Motivation

Reasoning about knowledge

We saw that KT45 can be used to reason about an agent's knowledge.

Difficulty

We have three agents (queen does not count), not just one. We want them to be able to reason about *each others* knowledge.

Motivation

Reasoning about knowledge

We saw that KT45 can be used to reason about an agent's knowledge.

Difficulty

We have three agents (queen does not count), not just one. We want them to be able to reason about *each others* knowledge.

Idea

Introduce a $\Box$ operator for each agent, and a $\Box$ operator for a group of agents.

Modal Logic KT45^n

Agents

Assume a set $\mathcal{A} = \{1, 2, \ldots, n\}$ of agents.

Modal Logic KT45^n

Agents

Assume a set $\mathcal{A} = \{1, 2, \ldots, n\}$ of agents.

Modal connectives

Replace $\Box$ by:

- K_i for each agent i
- E_G for any subset G of $\mathcal{A}$

Modal Logic $KT45^n$

Agents

Assume a set $\mathcal{A} = \{1, 2, \ldots, n\}$ of agents.

Modal connectives

Replace $\Box$ by:

- K_i for each agent i
- E_G for any subset G of $\mathcal{A}$

Example

$K_1\, p \wedge K_1\, \neg\, K_2\, K_1\, p$ means:

> *Agent 1 knows p, and also that Agent 2 does not know that Agent 1 knows p.*

Common Knowledge

“Everyone knows that everyone knows”

In KT45^n, $E_G\ E_G\ \phi$ is stronger than $E_G\ \phi$.

Common Knowledge

"Everyone knows that everyone knows"

In KT45^n, $E_G\ E_G\ \phi$ is stronger than $E_G\ \phi$.

"Everyone knows everyone knows everyone knows"

In KT45^n, $E_G\ E_G\ E_G\ \phi$ is stronger than $E_G\ E_G\ \phi$.

Common Knowledge

"Everyone knows that everyone knows"

In $KT45^n$, $E_G\ E_G\ \phi$ is stronger than $E_G\ \phi$.

"Everyone knows everyone knows everyone knows"

In $KT45^n$, $E_G\ E_G\ E_G\ \phi$ is stronger than $E_G\ E_G\ \phi$.

Common knowledge

The infinite conjunction $E_G\phi \wedge E_G\ E_G\phi \wedge \ldots$ is called "common knowledge of ϕ", denoted, $C_G\ \phi$.

Distributed Knowledge

Combine knowledge

If intelligent agents communicate with each other and use the knowledge each have, they can discover new knowledge.

Distributed Knowledge

Combine knowledge

If intelligent agents communicate with each other and use the knowledge each have, they can discover new knowledge.

Distributed knowledge

The operator $D_G\phi$ is called "distributed knowledge of ϕ", denoted, $D_G\ \phi$.

Models of $KT45^n$

Definition

A model $\mathcal{M} = (W, (R_i)_{i \in \mathcal{A}}, L)$ of the multi-modal logic $KT45^n$ is specified by three things:

Models of KT45^n

Definition

A model $\mathcal{M} = (W, (R_i)_{i \in \mathcal{A}}, L)$ of the multi-modal logic KT45^n is specified by three things:

1. A set W, whose elements are called *worlds*;

Models of KT45^n

Definition

A model $\mathcal{M} = (W, (R_i)_{i \in \mathcal{A}}, L)$ of the multi-modal logic KT45^n is specified by three things:

1. A set W, whose elements are called *worlds*;
2. For each $i \in \mathcal{A}$ a relation R_i on W, meaning $R_i \subseteq W \times W$, called the accessibility relations;

Models of KT45^n

Definition

A model $\mathcal{M} = (W, (R_i)_{i \in \mathcal{A}}, L)$ of the multi-modal logic KT45^n is specified by three things:

1. A set W, whose elements are called *worlds*;
2. For each $i \in \mathcal{A}$ a relation R_i on W, meaning $R_i \subseteq W \times W$, called the accessibility relations;
3. A labeling function $L : W \to \mathcal{P}(\texttt{Atoms})$.

Semantics of KT45^n

Definition

Take a model $\mathcal{M} = (W, (R_i)_{i \in \mathcal{A}}, L)$ and a world $x \in W$. We define $x \Vdash \phi$ via structural induction:

Semantics of KT45^n

Definition

Take a model $\mathcal{M} = (W, (R_i)_{i \in \mathcal{A}}, L)$ and a world $x \in W$. We define $x \Vdash \phi$ via structural induction:

- $x \Vdash p$ iff $p \in L(x)$

Semantics of KT45^n

Definition

Take a model $\mathcal{M} = (W, (R_i)_{i \in \mathcal{A}}, L)$ and a world $x \in W$. We define $x \Vdash \phi$ via structural induction:

- $x \Vdash p$ iff $p \in L(x)$
- $x \Vdash \neg\phi$ iff $x \nVdash \phi$

Semantics of KT45^n

Definition

Take a model $\mathcal{M} = (W, (R_i)_{i \in \mathcal{A}}, L)$ and a world $x \in W$. We define $x \Vdash \phi$ via structural induction:

- $x \Vdash p$ iff $p \in L(x)$
- $x \Vdash \neg\phi$ iff $x \nVdash \phi$
- $x \Vdash \phi \wedge \psi$ iff $x \Vdash \phi$ and $x \Vdash \psi$

Semantics of $KT45^n$

Definition

Take a model $\mathcal{M} = (W, (R_i)_{i \in \mathcal{A}}, L)$ and a world $x \in W$. We define $x \Vdash \phi$ via structural induction:

- $x \Vdash p$ iff $p \in L(x)$
- $x \Vdash \neg\phi$ iff $x \nVdash \phi$
- $x \Vdash \phi \wedge \psi$ iff $x \Vdash \phi$ and $x \Vdash \psi$
- $x \Vdash \phi \vee \psi$ iff $x \Vdash \phi$ or $x \Vdash \psi$

Semantics of KT45^n

Definition

Take a model $\mathcal{M} = (W, (R_i)_{i \in \mathcal{A}}, L)$ and a world $x \in W$. We define $x \Vdash \phi$ via structural induction:

- $x \Vdash p$ iff $p \in L(x)$
- $x \Vdash \neg\phi$ iff $x \nVdash \phi$
- $x \Vdash \phi \land \psi$ iff $x \Vdash \phi$ and $x \Vdash \psi$
- $x \Vdash \phi \lor \psi$ iff $x \Vdash \phi$ or $x \Vdash \psi$
- $x \Vdash \phi \rightarrow \psi$ iff $x \Vdash \psi$, whenever $x \Vdash \phi$
- ...

Semantics of KT45^n (continued)

Definition

Take a model $\mathcal{M} = (W, (R_i)_{i \in \mathcal{A}}, L)$ and a world $x \in W$. We define $x \Vdash \phi$ via structural induction:

- ...

Semantics of KT45^n (continued)

Definition

Take a model $\mathcal{M} = (W, (R_i)_{i \in \mathcal{A}}, L)$ and a world $x \in W$. We define $x \Vdash \phi$ via structural induction:

- ...
- $x \Vdash K_i \phi$ iff for each $y \in W$ with $R_i(x, y)$, we have $y \Vdash \phi$

Semantics of KT45^n (continued)

Definition

Take a model $\mathcal{M} = (W, (R_i)_{i \in \mathcal{A}}, L)$ and a world $x \in W$. We define $x \Vdash \phi$ via structural induction:

- ...
- $x \Vdash K_i\phi$ iff for each $y \in W$ with $R_i(x, y)$, we have $y \Vdash \phi$
- $x \Vdash E_G\phi$ iff for each $i \in G$, $x \Vdash K_i\phi$.

Semantics of KT45^n (continued)

Definition

Take a model $\mathcal{M} = (W, (R_i)_{i \in \mathcal{A}}, L)$ and a world $x \in W$. We define $x \Vdash \phi$ via structural induction:

- ...
- $x \Vdash K_i \phi$ iff for each $y \in W$ with $R_i(x, y)$, we have $y \Vdash \phi$
- $x \Vdash E_G \phi$ iff for each $i \in G$, $x \Vdash K_i \phi$.
- $x \Vdash C_G \phi$ iff for each $k \geq 1$, we have $x \Vdash {E_G}^k \phi$.

Semantics of KT45^n (continued)

Definition

Take a model $\mathcal{M} = (W, (R_i)_{i \in \mathcal{A}}, L)$ and a world $x \in W$. We define $x \Vdash \phi$ via structural induction:

- ...
- $x \Vdash K_i\phi$ iff for each $y \in W$ with $R_i(x, y)$, we have $y \Vdash \phi$
- $x \Vdash E_G\phi$ iff for each $i \in G$, $x \Vdash K_i\phi$.
- $x \Vdash C_G\phi$ iff for each $k \geq 1$, we have $x \Vdash {E_G}^k\phi$.
- $x \Vdash D_G\phi$ iff for each $y \in W$, we have $y \Vdash \phi$, whenever $R_i(x, y)$ for all $i \in G$.

Formulation of Wise-Women Puzzle

Setup

- Wise woman i has red hat: p_i
- Wise woman i knows that wise woman j has a red hat: $K_i\ p_j$

Formulation of Wise-Women Puzzle

Initial situation

$$\begin{aligned}
\Gamma \quad = \{ \quad & C(p_1 \vee p_2 \vee p_3), \\
& C(p_1 \rightarrow K_2 p_1), C(\neg p_1 \rightarrow K_2 \neg p_1), \\
& C(p_1 \rightarrow K_3 p_1), C(\neg p_1 \rightarrow K_3 \neg p_1), \\
& C(p_2 \rightarrow K_1 p_2), C(\neg p_2 \rightarrow K_1 \neg p_2), \\
& C(p_2 \rightarrow K_3 p_2), C(\neg p_2 \rightarrow K_3 \neg p_2), \\
& C(p_3 \rightarrow K_1 p_3), C(\neg p_2 \rightarrow K_1 \neg p_3), \\
& C(p_3 \rightarrow K_2 p_3), C(\neg p_2 \rightarrow K_2 \neg p_3)\}
\end{aligned}$$

Announcements

First wise woman says "No"

$$C(\neg K_1 p_1 \wedge \neg K_1 \neg p_1)$$

Second wise woman says "No"

$$C(\neg K_2 p_2 \wedge \neg K_2 \neg p_2)$$

First Attempt

$$\Gamma, C(\neg K_1 p_1 \wedge \neg K_1 \neg p_1), C(\neg K_2 p_2 \wedge \neg K_2 \neg p_2) \vdash K_3 p_3$$

First Attempt

$$\Gamma, C(\neg K_1 p_1 \wedge \neg K_1 \neg p_1), C(\neg K_2 p_2 \wedge \neg K_2 \neg p_2) \vdash K_3 p_3$$

Problem

This does not take time into account. The second announcement can take the first announcement into account.

Solution

Prove separately:

Entailment 1 :

$$\Gamma, C(\neg K_1 p_1 \wedge \neg K_1 \neg p_1) \vdash C(p_2 \vee p_3)$$

Solution

Prove separately:

Entailment 1 :

$$\Gamma, C(\neg K_1 p_1 \land \neg K_1 \neg p_1) \vdash C(p_2 \lor p_3)$$

Entailment 2 :

$$\Gamma, C(p_2 \lor p_3), C(\neg K_2 p_2 \land \neg K_2 \neg p_2) \vdash K_3 p_3$$

Solution

Prove separately:

Entailment 1 :

$$\Gamma, C(\neg K_1 p_1 \wedge \neg K_1 \neg p_1) \vdash C(p_2 \vee p_3)$$

Entailment 2 :

$$\Gamma, C(p_2 \vee p_3), C(\neg K_2 p_2 \wedge \neg K_2 \neg p_2) \vdash K_3 p_3$$

Proof

Through natural deduction in KT45^n.