# Principles of
# Knowledge Representation and Reasoning

## 3. Modal Logics

### 3.1 Boxes and Diamonds

*Bernhard Nebel*

- Motivation
- Syntax
- Semantics: Possible Worlds
- Different Systems of Modal Logics
- Proof Methods
- Outlook

# Motivation for Studying Modal Logics

- Some KR formalisms can be understood as (fragments of) a **propositional modal logic**
- Complexity and decidability results can be re-used
- Algorithms/proof methods can be re-used

$\leadsto$ Motivation for introducing modal logic as an independent topic

$\rightarrow$ Will be used for the qualitative spatial representation formalism **RCC8**

$\rightarrow$ Another application will be **description logics**

# Motivation for Modal Logics

Often, we want to state something where we have an **“embedded proposition”**:

- John believes that theory is nonsense
- I know that $2^{10} = 1024$

Reasoning with embedded propositions:

- John believes that theory is nonsense
- John believes that if theory is nonsense then theory is dispensable

$\rightarrow$ This implies (assuming *belief* is closed under *modus ponens*):

  $\leadsto$ John believes that theory is dispensable

$\leadsto$ How to formalize this?

# Syntax

Propositional logic $+$ two new unary operators: $\Box, \Diamond$ ***(Box & Diamond)***:

$$\begin{array}{rcll} \varphi & \longrightarrow & \ldots & \textit{classical propositional formula} \\ & | & \Box\varphi' & \textit{Box} \\ & | & \Diamond\varphi' & \textit{Diamond} \end{array}$$

$\Box$ and $\Diamond$ have the same operator precedence as $\neg$. Some possible readings of $\Box\varphi$:

- Necessarily $\varphi$ (alethic)
- Always $\varphi$ (temporal)
- Make $\varphi$ true! (deontic)
- Agent $A$ believes $\varphi$ (doxastic)
- Agent $A$ knows $\varphi$ (epistemic)

$\leadsto$ different formalizations for different intended readings

# Truth Functional Semantics?

- Could it be possible to define the meaning of $\Box\varphi$ **truth functionally**, i.e. by referring to the truth value of $\varphi$ only?

- Trial for the *necessity* interpretation:
  - If $\varphi$ is false, then $\Box\varphi$ should be false.
  - If $\varphi$ is true, then ...

  $\leadsto$ ... $\Box\varphi$ should be true $\leadsto$ $\Box$ is the identity function

  $\leadsto$ ... $\Box\varphi$ should be false $\leadsto$ $\Box\varphi$ is identical to falsity

- **Note**: There are only 4 different unary Boolean functions.

# Semantics: The Idea

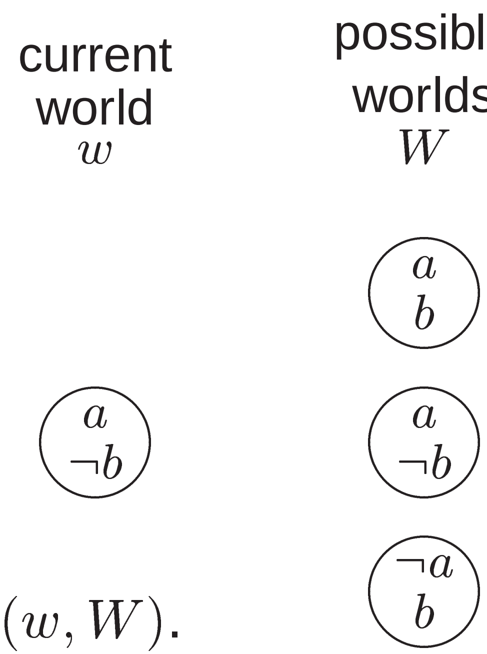
In classical propositional logic, formulae are interpreted over single interpretations and are evaluated to *true* or *false*.

In modal logics one considers always **sets** of such interpretations: **possible worlds** (physically possible, conceivable, . . . )

**Main idea**:

- Consider a world (interpretation) $w$ and a set of worlds $W$, which are possible with respect to $w$
- A classical formula (with no modal operators) $\varphi$ is true relative to $(w, W)$ iff $\varphi$ is true in $w$
- $\Box\varphi$ is true relative to $(w, W)$ iff $\varphi$ is true in **all worlds** in $W$
- $\Diamond\varphi$ is true relative to $(w, W)$ iff $\varphi$ is true in **one world** in $W$

# Semantics: An Example




**Examples**:

- $a \wedge \neg b$ is true relative to $(w, W)$.
- $\Box a$ is not true relative to $(w, W)$.
- $\Box(a \vee b)$ is true relative to $(w, W)$.

**Question**: How shall we evaluate modal formulae in $w \in W$?

$\leadsto$ for each world, we specify the possible worlds

$\leadsto$ **frames**

# Frames, Interpretations, and Worlds

A **frame** is a pair $\mathcal{F} = \langle W, R \rangle$, where $W$ is a non-empty set (of *worlds*) and $R \subseteq W \times W$ (the *accessibility relation*).

For $(w, v) \in R$ we write also $wRv$.We say that $v$ is an **$R$-successor** of $w$ and that $v$ is **reachable** (or $R$-reachable) from $w$.

A **($\Sigma$)-interpretation** (or model) **based on the frame** $\mathcal{F} = \langle W, R \rangle$ is a triple $\mathcal{I} = \langle W, R, \pi \rangle$, where $\pi$ is a function from worlds to truth assignments:

$$\pi\colon W \to (\Sigma \to \{T, F\})$$

# Semantics: Truth in one World

A formula $\varphi$ is **true in world $w$ of an interpretation $\mathcal{I} = \langle W, R, \pi \rangle$** under the following conditions:

$$
\begin{array}{rcl}
\mathcal{I}, w \models a & \textit{iff} & \pi(w)(a) = T \\
\mathcal{I}, w \models \top & & \\
\mathcal{I}, w \not\models \bot & & \\
\mathcal{I}, w \models \neg\varphi & \textit{iff} & \mathcal{I}, w \not\models \varphi \\
\mathcal{I}, w \models \varphi \wedge \psi & \textit{iff} & \mathcal{I}, w \models \varphi \text{ and } \mathcal{I}, w \models \psi \\
\mathcal{I}, w \models \varphi \vee \psi & \textit{iff} & \mathcal{I}, w \models \varphi \text{ or } \mathcal{I}, w \models \psi \\
\mathcal{I}, w \models \varphi \rightarrow \psi & \textit{iff} & \text{if } \mathcal{I}, w \models \varphi, \text{ then } \mathcal{I}, w \models \psi \\
\mathcal{I}, w \models \varphi \leftrightarrow \psi & \textit{iff} & \mathcal{I}, w \models \varphi, \text{ if and only if } \mathcal{I}, w \models \psi \\
\mathcal{I}, w \models \Box\varphi & \textit{iff} & \mathcal{I}, u \models \varphi \text{ for all } u \text{ s.t. } wRu \\
\mathcal{I}, w \models \Diamond\varphi & \textit{iff} & \mathcal{I}, u \models \varphi \text{ for one } u \text{ s.t. } wRu
\end{array}
$$

# Satisfiability and Validity

A formula $\varphi$ is called **satisfiable in an interpretation** $\mathcal{I}$ (or in a frame $\mathcal{F}$, or in a class of frames $\mathcal{C}$) if there exists a world in $\mathcal{I}$ (or an interpretation $\mathcal{I}$ based on $\mathcal{F}$, or an interpretation $\mathcal{I}$ based on a frame contained in the class $\mathcal{C}$, respectively) such that $\mathcal{I}, w \models \varphi$.

A formula $\varphi$ is called **true in an interpretation** $\mathcal{I}$ (symbolically $\mathcal{I} \models \varphi$) if $\varphi$ is true in all worlds of $\mathcal{I}$.

A formula $\varphi$ is called **valid in a frame** $\mathcal{F}$ or $\mathcal{F}$-valid (symbolically $\mathcal{F} \models \varphi$) if $\varphi$ is true in all interpretations based on $\mathcal{F}$.

A formula $\varphi$ is called **valid in a class of frames** $\mathcal{C}$ or $\mathcal{C}$-valid (symbolically $\mathcal{C} \models \varphi$) if $\mathcal{F} \models \varphi$ for all $\mathcal{F} \in \mathcal{C}$.

**K** is the class of all frames – named after Kripke, who invented this semantics.

# Validity: Some Examples

- $\varphi \vee \neg\varphi$;
- $\Box(\varphi \vee \neg\varphi)$;
- $\Box\varphi$, if $\varphi$ is a classical tautology;
- $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$ (called axiom schema $K$).

**Theorem**. $K$ is **K**-valid.

**Proof**. Let $\mathcal{I}$ be an interpretation and let $w$ be a world in $\mathcal{I}$.

Assumption: $\mathcal{I}, w \models \Box(\varphi \to \psi)$, i.e., in all worlds $u$ with $wRu$ , we have that if $\varphi$ is true, then $\psi$ must be true(otherwise $K$ is true in any case).

If $\Box\varphi$ is false in $w$, then $(\Box\varphi \to \Box\psi)$ is true.

If $\Box\varphi$ is true in $w$, then $\varphi$ is true in all worlds $u$. Because of our assumption, $\Box\psi$ is true in $w$, i.e., $(\Box\varphi \to \Box\psi)$ is true in $w$.

Since $\mathcal{I}$ and $w$ were arbitrary, the argument goes through for any $\mathcal{I}, w$, i.e., $K$ is **K**-valid.