Modal Logics and Agents

- In propositional or predicate logic, formulas are either true, or false, in any model.
- In natural language we often distinguish between many modes of truth:
  - *necessarily true*
  - *known to be true*
  - *believed to be true*
  - *true in the future*
- For example, the sentence: “G. Bush is the president of the U.S.A.”, although currently true, will not be true at some point in the future.
- The sentence: “There are nine planets in the solar system is true, and may be true forever in the future, is not necessarily true, in the sense that it could have been a different number.”
- The sentence: “The cubic root of 27 is 3” is true, necessarily true, and true in the future. However, it may not be known to be true by some people, or not believed to be true.

Modes of Truth

- CTL could distinguish between truth at different points in the future, as well as different futures. Temporal logic is thus a special case of modal logic.
- Modal logic is also very useful in modelling artificial intelligence problems, like the interaction of agents in various environments. Each agent has different knowledge about its environment, and we use modalities to model the knowledge of the agents.
- Modal logic adds unary connectives to express one, or more of these modes of truth. The simplest modal logics deal with only one concept, such as knowledge, necessity, or time.
- We take a logic engineering approach: given a particular mode of truth, how can we develop a logic capable of expressing that modality.

We add the extra connectives $\Box$ and $\Diamond$ to the set of connectives given in propositional logic.

Definition: The *formulas* of basic modal logic ϕ are defined by the following BNF:

$$\Phi ::= \bot \mid \top \mid p \mid \neg\Phi \mid (\Phi \wedge \Phi) \mid (\Phi \vee \Phi) \mid (\Phi \rightarrow \Phi) \mid (\Phi \leftrightarrow \Phi) \mid \Box\Phi \mid \Diamond\Phi$$

For these formulas, we define *syntax trees* in the usual way.

Convention: We assume that the unary connectives $\neg$, $\Box$, and $\Diamond$ bind most closely, followed by $\wedge$, and $\vee$, followed by $\rightarrow$ and $\leftrightarrow$.

Definition: A *model* $\mathcal{M}$ of basic modal logic is specified by three things:

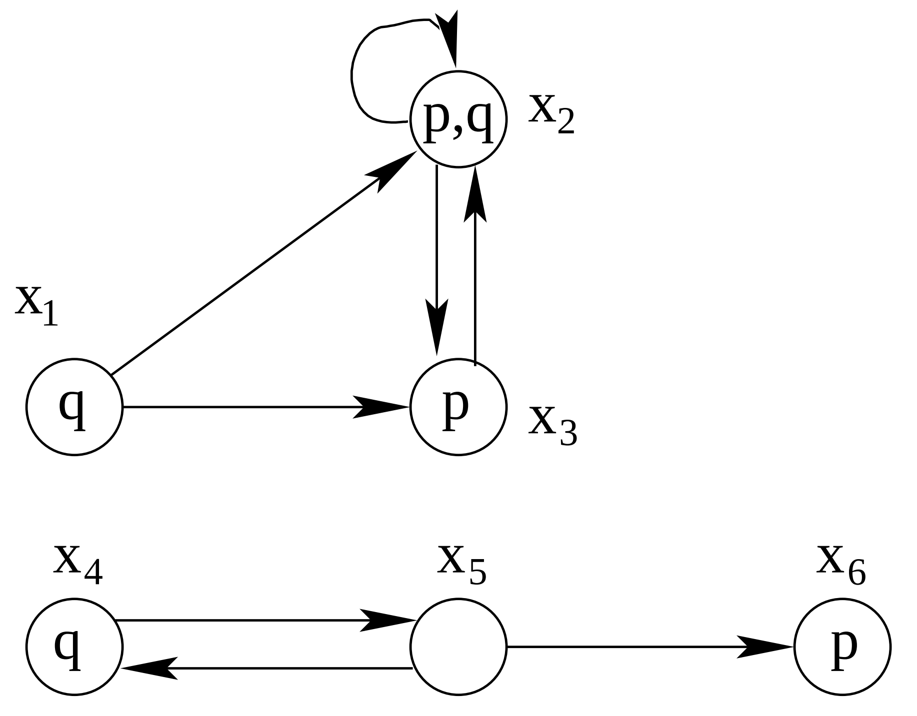
- A set W, whose elements are called *worlds*;
- A relation $R \subseteq W \times W$, called the accessibility relation;
- A *labelling function* $L : W \to \mathcal{P}(\mathtt{Atoms})$.

Such models are called *Kripke models*. Intuitively, $w \in W$ stands for a possible world and wRw' means that w' is a world related to w. The actual nature of the relationship depends on what we really want to model.

If the wet of worlds W is finite, we can use an easy graphical notation to represent a model. Suppose $W = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, the relation $R = \{(x_1,x_2),(x_1,x_3),(x_2,x_2),(x_2,x_3),(x_3,x_2),(x_4,x_5),(x_5,x_4),(x_5,x_6)\}$, and the labelling function given by the table:

x	$L(x)$
x_1	$\{q\}$
x_2	$\{p,q\}$
x_3	$\{p\}$
x_4	$\{q\}$
x_5	$\emptyset$
x_6	$\{p\}$

The Kripke model can be represented as:



Definition: Let $\mathcal{M} = (W, R, L)$ be a model of basic modal logic. Suppose $x \in W$ and ϕ is a formula. We define that ϕ is *true* in the world x, denoted $x \Vdash \phi$, inductively:

$$
\begin{array}{rl}
x \Vdash \top & \\
x \nVdash \bot & \\
x \Vdash p & \text{iff } p \in L(x) \\
x \nVdash \neg\phi & \text{iff } x \nVdash \phi \\
x \Vdash \phi \wedge \psi & \text{iff } x \Vdash \phi \text{ and } x \Vdash \psi \\
x \Vdash \phi \vee \psi & \text{iff } x \Vdash \phi\text{, or } x \Vdash \psi \\
x \Vdash \phi \rightarrow \psi & \text{iff } x \Vdash \phi\text{, whenever we have } x \Vdash \psi \\
x \Vdash \phi \leftrightarrow \psi & \text{iff } x \Vdash \phi \text{ iff } x \Vdash \psi \\
x \Vdash \Box\psi & \text{iff , for each } y \in W \text{ with } R(x,y)\text{, we have } y \Vdash \psi \\
x \Vdash \Diamond\psi & \text{iff there is a } y \in W \text{ such that } R(x,y) \text{ and } y \Vdash \psi
\end{array}
$$

We also write $\mathcal{M}, x \Vdash \phi$ to denote the fact that $x \Vdash \phi$ in the model $\mathcal{M}$.

Model of Basic Modal Logic (cont'd)

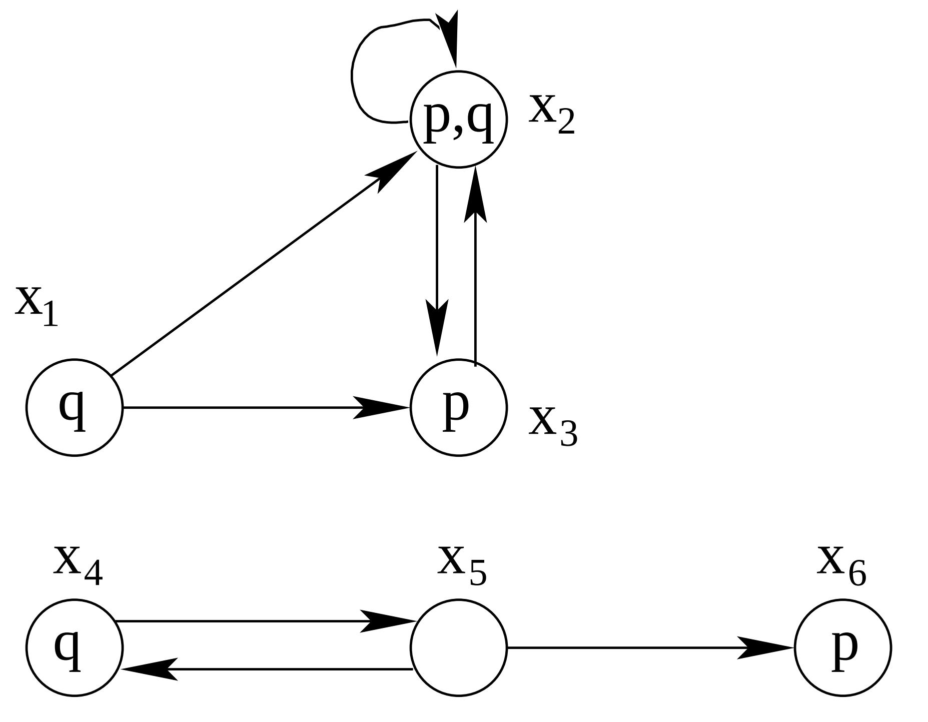
In the previous definition:

- The first two clauses simply express the fact that $\top$ is always true, while $\bot$ is always false.
- $L(x)$ is the set of all atomic formulas that are true at x.
- The boolean connectives $\neg$, $\wedge$, $\vee$, $\rightarrow$, and $\leftrightarrow$ is the same as in propositional logic.
- For $\Box\phi$ to be true at x, we require that ϕ be true in all worlds related to x (resembles AX of CTL).
- For $\Diamond\phi$ to be true at x, we require that ϕ to be true in at least one related world (resembles EX of CTL).

Definition: A model $\mathcal{M} = (W, R, L)$ of basic modal logic is said to satisfy a formula if every state in the modle satisfies it. Thus, we write $\mathcal{M} \models \phi$ iff, for each $x \in W$, $x \Vdash \phi$.

Examples

- $x_1 \Vdash q$, since $q \in L(x_1)$.
- $x_1 \Vdash \Diamond q$, since $x_1 R x_2$ and $x_2 \Vdash q$.
- $x_1 \nVdash \Box q$, since $x_3 \nVdash q$.
- $x_5 \nVdash \Box p$, and $x_5 \nVdash \Box q$.
- $x_5 \Vdash \Box(p \vee q)$



Comment

Worlds like x_6 in the previous slide, which have no worlds related to them, have the property

$$x_6 \not\Vdash \Diamond\phi$$

for all formulas ϕ (even if ϕ is $\top$). On the other hand, we have

$$x_6 \Vdash \Box\phi$$

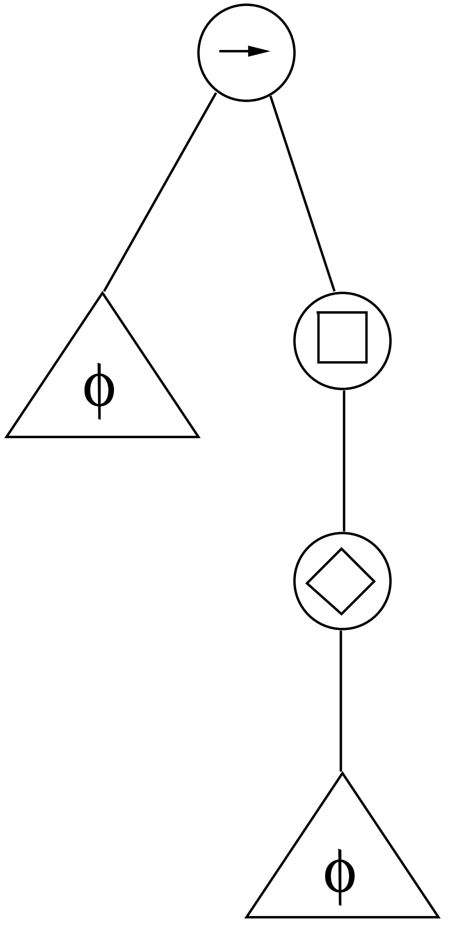
for all formulas ϕ (even when ϕ is $\bot$).

Formulas and Formula Schemes

Example: the formulas $p \longrightarrow \Box\Diamond p$, $q \longrightarrow \Box\Diamond q$, and $(p \wedge \Diamond q) \longrightarrow \Box\Diamond(p \wedge \Diamond q)$ are instances of the scheme

$$\phi \longrightarrow \Box\Diamond\phi.$$

The syntax tree of this scheme is:



Equivalences Between Modal Formulas

Definition: We say that a set of formulas Γ of basic modal logic *semantically entails* a formula ϕ of basic modal logic if, in any world x of any model $\mathcal{M} = (W, R, L)$, we have $x \Vdash \phi$ whenever $x \Vdash \psi$, for all $\psi \in \Gamma$. In that case, we write $\Gamma \models \phi$. We say that ϕ and ψ are *semantically equivalent* if $\phi \models \psi$ and $\psi \models \phi$ hold. We denote this by $\phi \equiv \psi$.

Note: Any equivalence in propositional logic is an equivalence in modal logic.

Example

Remeber that $p \longrightarrow \neg q$ and $\neg(p \wedge q)$ are equivalent in propositional logic. Let us perform the substitution:

$$p \mapsto \Box p \wedge (q \longrightarrow p)$$
$$q \mapsto r \longrightarrow \Diamond(q \vee p)$$

The result is the pair of modal logic formulas:

$$\Box p \wedge (q \longrightarrow p) \longrightarrow \neg(q \mapsto r \longrightarrow \Diamond(q \vee p))$$
$$\neg((\Box p \wedge (q \longrightarrow p)) \wedge (r \longrightarrow \Diamond(q \vee p)))$$

are semantically equivalent.

De Morgan Rules and Distributivity

De Morgan rules:

$$\neg\Box\phi \equiv \Diamond\neg\phi \text{ (resembles } \neg\forall x\phi \equiv \exists x\neg\phi)$$
$$\neg\Diamond\phi \equiv \Box\neg\phi \text{ (resembles } \neg\exists x\phi \equiv \forall\neg\phi)$$

$\Box$ distributes over $\wedge$ and $\Diamond$ distributes over $\vee$.

$$\Box(\phi \wedge \psi) \equiv \Box\phi \wedge \Box\psi \text{ and } \Diamond(\phi \vee \psi) \equiv \Diamond\phi \vee \Diamond\psi$$

Valid Formulas

Definition: A formula of basic modal logic is said to be *valid* if it is true in every world of every model, i.e. iff $\models \phi$.

Every propositional tautology is valid in modal logic.

Valid Formulas (cont'd)

The following formulas are valid:

$$\neg\Box\phi \leftrightarrow \Diamond\neg\phi$$
$$\Box(\phi \wedge \psi) \leftrightarrow \Box\phi \wedge \Box\psi$$
$$\Diamond(\phi \vee \psi) \leftrightarrow \Diamond\phi \vee \Diamond\psi$$
$$\Box(\phi \rightarrow \psi) \wedge \Box\phi \rightarrow \Box\phi$$
$$\Box(\phi \rightarrow \psi) \rightarrow (\Box\phi \rightarrow \Box\phi)$$

Proof of validity for the first formula: Let $\mathcal{M} = (W, R, L)$ be a model, and $x \in W$. We want to show $x \Vdash \neg\Box\phi \leftrightarrow \Diamond\neg\phi$, i.e. that $x \Vdash \neg\Box\phi$ iff $x \Vdash \Box\neg\phi$. The following chain of reasoning follows:

$x \Vdash \neg\Box\phi$

iff it isn't the case that $x \Vdash \Box\phi$

iff it isn't the case that, for all y such that $R(x,y)$, $y \Vdash \phi$

iff there is some y such that $R(x,y)$ and not $y \Vdash \phi$

iff there is some y such that $R(x,y)$ and $y \Vdash \neg\phi$

iff $x \Vdash \Diamond\neg\phi$

- The basic framework is quite general and can be refined in various ways to provide the properties that are appropriate for the intended applications.
- Logic engineering is the subject of engineering logics to fit new applications.
- We will consider how to engineer the basic framework for modal logic to fir the following readings of $\Box\phi$.
  - 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.

Logic Engineering (cont'd)

Since $\Diamond$ can be expressed as $\neg\Box\neg$, we can find the corresponding readings for $\Diamond$.

For example, the reading of $\Diamond\phi$ corresponding to "It is necessarily true that ϕ can be inferred in the following steps:

- "It is *not* necessarily true that ϕ" = "It is possible that $\neg\phi$."
- "It is *not* necessarily true that $\neg\phi$" = "It is possible that *not not* ϕ"= "It is possible that ϕ".

The reading of $\Diamond\phi$ corresponding to "Agent Q knows ϕ" can be derived as follows:

agent Q does *not* know *not* ϕ
= as far as Q's knowledge is concerned, ϕ could be the case
= ϕ is consistent with what agent Q knows
= for all agent Q knows, ϕ.

Readings of Modal Connectives

$\Box\phi$	$\Diamond\phi$
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 execution of program P, ϕ holds	After some execution of P, ϕ holds

The Stock of Valid Formulas

- For a logic that captures the concept of *necessity*, we must have $\Box p \rightarrow p$.
- This can be read as: “for anything which is necessarily true, is also simply true.”
- Or, “whatever agent Q knows is true”. (everything that is known must also be true; on the other hand, things that are believed are not necessarily true).

The Stock of Valid Formulas (2)

The table shows six interesting readings for $\Box$ and eight formula schemes. It is rather debatable whether to put a cross or a tick in some of the boxes; this depends on the concept of truth we are trying to formalize.

$\Box\phi$	$\Box\phi \to \phi$	$\Box\phi \to \Box\Box\phi$	$\Diamond\phi \to \Box\Diamond\phi$	$\Diamond\top$	$\Box\phi \to \Diamond\phi$	$\Box\phi \lor \Box\neg\phi$	$\Box(\phi \to \psi) \land \Box\phi \to \Box\psi$	$\Diamond\phi \land \Diamond\psi \to \Diamond(\phi \land \psi)$
It is necessarily true that ϕ	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\times$	$\surd$	$\times$
It will always be true 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 any execution of program P, ϕ holds	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\surd$	$\times$

Necessity

Consider the formulas $\Box\phi \to \Box\Box\phi$.

- Reading: "What is necessary, is also *necessarily* necessary."
- If we are dealing with physical necessity, this amounts to whether the laws of the universe are a physical necessity.
- That is, do the laws of the universe entail that they should be the laws of the universe? The answer seems to be "no".
- What if we meant logical necessity? The answer must now be "yes", since the laws of logic are assertions whose truth cannot be denied.

Always in the Future

The future may or may not include the present.

- If the future does not include the present, then $\Box\phi \rightarrow \phi$ is not valid.
- The validity of the formula $\Diamond\top$ can be interpreted as "time has no end". Whether this formula should be valid or not also depends of how we want to interpret time.

Ought

- In this case, $\Box$ is interpreted as “ought”, and $\Diamond$ is interpreted as “permitted”.
- The formula $\Box\phi \rightarrow \Box\Box\phi$ is not valid.
- Example: “It ought to be the case that we wear a seat-belt” does not compel us to believe that “It ought to be the case that we *ought* to wear a belt”.
- However, anything which ought to be so, should be permitted to be so, and therefore $\Diamond\phi \rightarrow \Diamond\phi$.

Belief

- To decide if $\Diamond\top$ should be valid, let us express it as $\neg\Box\bot$.
- $\neg\Box\bot$ is read as "Agent Q does not believe any contradictions".
- We need to be precise about whether we model human beings, which may not be capable of perfect reasoning, or not.
- Typically, it is preferred to model idealized agents, capable of perfect reasoning, and therefore able to spot any contradiction.
- The formula $\Diamond\phi \to \Box\Diamond\phi$ reads as "If agent Q doesn't believe something, then he believes he doesn't believe it." —it is reasonable to assume it is valid.
- Validity of the formula $\Box\phi \vee \Box\neg\phi$ means that agent Q has an opinion on everything —we assume this is not reasonable, and therefore this formula is not valid.

Knowledge

While agent Q may have false beliefs, he may only *know* only that which is true.

- $\Box\phi \rightarrow \Box\Box\phi$ —"if one knows something, he/she knows he/she knows it" —should be valid.
- $\neg\Box\phi \rightarrow \Box\neg\Box\phi$ —"if one doesn't know something, he/she knows he/she doesn't know it." —should be valid only if we model idealized agents. (it is not true of ordinary humans).

Execution of Programs

- $\Box\phi \to \Box\Box\phi$ says that running the program twice is the same as running it once —false if the program deducts money from a bank account!
- $\Diamond\top$ is read as "there is an execution of the program which terminates." —again false.

Properties of the Accessibility Relation

- We have been engineering logics at the level of deciding what formulas should be valid for the various readings of $\Box$.
- We can also engineer at the level of Kripke models.
- For each of our seven readings of $\Box$, there is a corresponding reading of the accessibility relation R.
- In some of these readings, it will be useful to stipulate that R is reflexive, or transitive, or has other properties.

The meaning of R, for readings of $\Box$

$\Box\phi$	$R(x,y)$
It is necessarily true that ϕ	y is possible according to the information at x
It will always be true that ϕ	y is in the future of x
It ought to be that ϕ	y is acceptable 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 program P, ϕ holds	y is a possible resulting state after execution of P at x

Types of Binary Relations

- *reflexive*: if, for every $x \in W$, we have $R(x,x)$.
- *symmetric*: if, for every $x, y \in W$, we have $R(x,y)$ implies $R(y,x)$.
- *serial*: if, for every x there is a y such that $R(x,y)$.
- *transitive*: if, for every $x, y, z \in W$, we have $R(x,y)$ and $R(y,z)$ imply $R(x,z)$.
- *Euclidian*: if, for every $x, y, z \in W$ with $R(x,y)$ and $R(x,z)$, we have $R(y,z)$.
- *functional*: if, for each x there is a unique y such that $R(x,y)$.
- *linear*: if, for every $x, y, z \in W$ we have that $R(x,y)$ and $R(x,z)$ together imply that $R(y,z)$, or y equals z, or $R(z,y)$.
- *total*: if for every $x, y \in W$ we have $R(x,y)$ or $R(y,x)$.
- *equivalence relation*: reflexive, symmetric and tansitive.

Example

If $\Box\phi$ means "Agent Q knows ϕ", then $R(x,y)$ means y could be the actual world according to Q's knowledge at x.

- Should R be reflexive? This would say: x could be the actual world according to Q's knowledge at x. In other words, Q cannot know that things are different from how they really are —i.e. Q *cannot have false knowledge*. This seems a desirable property to have.
- Should R be transitive? It would say: if y is possible according to Q's knowledge at x and z is possible according to Q's knowledge at y, then z is possible according to Q's knowledge at x. This again should be true.