UvA

Multi-Agent Systems

Modal Logic

Mina Young Pedersen
m.y.pedersen@uva.nl

Fall 2025

What is Modal Logic?

- **Modal logic** is a formal tool of reasoning with roots in philosophy
- It extends the language of propositional logic with two new symbols: a diamond $\Diamond$, and a box $\Box$
- In this part of the course, we are going to explore how we can use modal logic as a tool to reason about agents in a multi-agent system
    - To do this, we will look at different types of modal logic
- But first, a recap on the basic modal logic

Language of Basic Modal Logic

- There are many ways to introduce the basic modal logic
- Here, we begin by introducing its language

Definition (Basic Modal Language)
Formulas of the **basic modal language** are defined as follows. We first choose a set of propositions called **atoms:**
At := $p, q, r, \ldots,$ $\top$ ("always true"), $\bot$ ("always false")

Next, we define inductively how to construct further expressions, using the format called the Backus-Naur Form (BNF):

ϕ ::= At | $\neg\phi$ | $(\phi \wedge \psi)$ | $(\phi \vee \psi)$ | $(\phi \rightarrow \psi)$ | $\Diamond\phi$ | $\Box\,\phi$

Language of Basic Modal Logic

Definition (Basic Modal Language)

Formulas of the **basic modal language** are defined as follows. We first choose a set of propositions called **atoms:**

At $:= p, q, r, \ldots, \top$ ("always true"), $\bot$ ("always false")

Next, we define inductively how to construct further expressions, using the format called the Backus-Naur Form (BNF):

$$\phi, \psi ::= \text{At} \mid \neg\phi \mid (\phi \wedge \psi) \mid (\phi \vee \psi) \mid (\phi \to \psi) \mid \Diamond\phi \mid \Box\,\phi$$

$p, q, r,$ etc. and $\top$ and $\bot$ are formulas

if ϕ is a formula, then $\neg\phi$ is a formula

if ϕ and ψ are formulas, then $(\phi \wedge \psi)$ is a formula

(…)

Language of Basic Modal Logic

$$\phi ::= \mathrm{At} \mid \neg\phi \mid (\phi \wedge \psi) \mid (\phi \vee \psi) \mid (\phi \to \psi) \mid \Diamond\phi \mid \Box\phi$$

Are these (well-formed) formulas of the basic modal language?

$(pq \wedge r)$ ✗

$((p \vee q) \wedge r))$ ✓

$(\Box s \to r)$ ✓

$\Box s \to r$ ✓

$\Box(s \to r)$ ✓

$r \to (\Diamond p \wedge q)$ ✓

$\Box\Diamond\Diamond$ ✗

$\Diamond\Diamond p \to q$ ✓

(we can remove parenthesis when there is no ambiguity)

Example: Disambiguation

- The un-bracketed flat symbol string $\neg \Box p \to q$ has three different modal readings

$$\neg(\Box p \to q)$$

$$\neg \Box (p \to q)$$

$$(\neg \Box p \to q)$$

Parenthesis are important!

But for a formula (ϕ), we can write it as ϕ

A Note on Notation

- All formulas of the basic modal language are built up of p, q, r, s etc., the symbols $\top$ and $\bot$, the connectives $\neg$, $\wedge$, $\vee$, $\rightarrow$ and the modal operators $\Diamond$ and $\Box$
    - According to the structure of what forms a well-formed formula (defined in the Backus-Naur Form)
- However, when we want to denote *arbitrary* formulas, we use Greek letters ϕ, ψ, χ etc.

Intuitive Readings of $\Diamond$ and $\Box$

- There are many possible readings for the modalities $\Diamond$ and $\Box$
  - For example:
    - $\Diamond\phi$: "it is possible that ϕ" and $\Box\phi$: "it is necessary that ϕ"
    - $\tilde{K}_a\phi$: "agent *a* considers it possible that ϕ" and $K_a\phi$: "agent *a* knows that ϕ"
    - $F\phi$: "ϕ will happen some time in the future" and $G\phi$: "ϕ will always be the case"

Intuitive Readings of $\Diamond$ and $\Box$

- The crucial aspect is that $\Diamond$ and $\Box$ have a particular relationship: they are **dualities** of each other
  - The two following formulas are intuitively valid

$$\Box\phi \leftrightarrow \neg\Diamond\neg\phi \qquad\qquad \Diamond\phi \leftrightarrow \neg\Box\neg\phi$$

- Similar to the relationship between the universal quantifier $\forall$ and existential quantifier $\exists$ in first-order logic

Models and Frames

- Our language is interpreted over simple graph-like structures

Definition (Relational Model and Relational Frame)

A relational model (or a possible worlds model) is a tuple $M = (W, R, V)$, where:

- **W is a non-empty domain (whose elements are generically called (possible) worlds or states);**
- **$R \subseteq (W \times W)$ is a binary relation on W (connecting some worlds to some others);**
- **$V : \mathrm{At} \to \mathscr{P}(W)$ is an atomic valuation (indicating the set of worlds that satisfy each one of the atomic propositions).**

A relational frame $F = (W, R)$ is a relational model without valuation.

We write Rwu when the pair (w, u) is in the relation R

Intuitive Interpretations

- There are many interpretations for possible worlds, ranging from metaphysical worlds to states of a computer, board positions in chess, deals in a card game…

- The accessibility relation R can be universal (every world is accessible to every other)

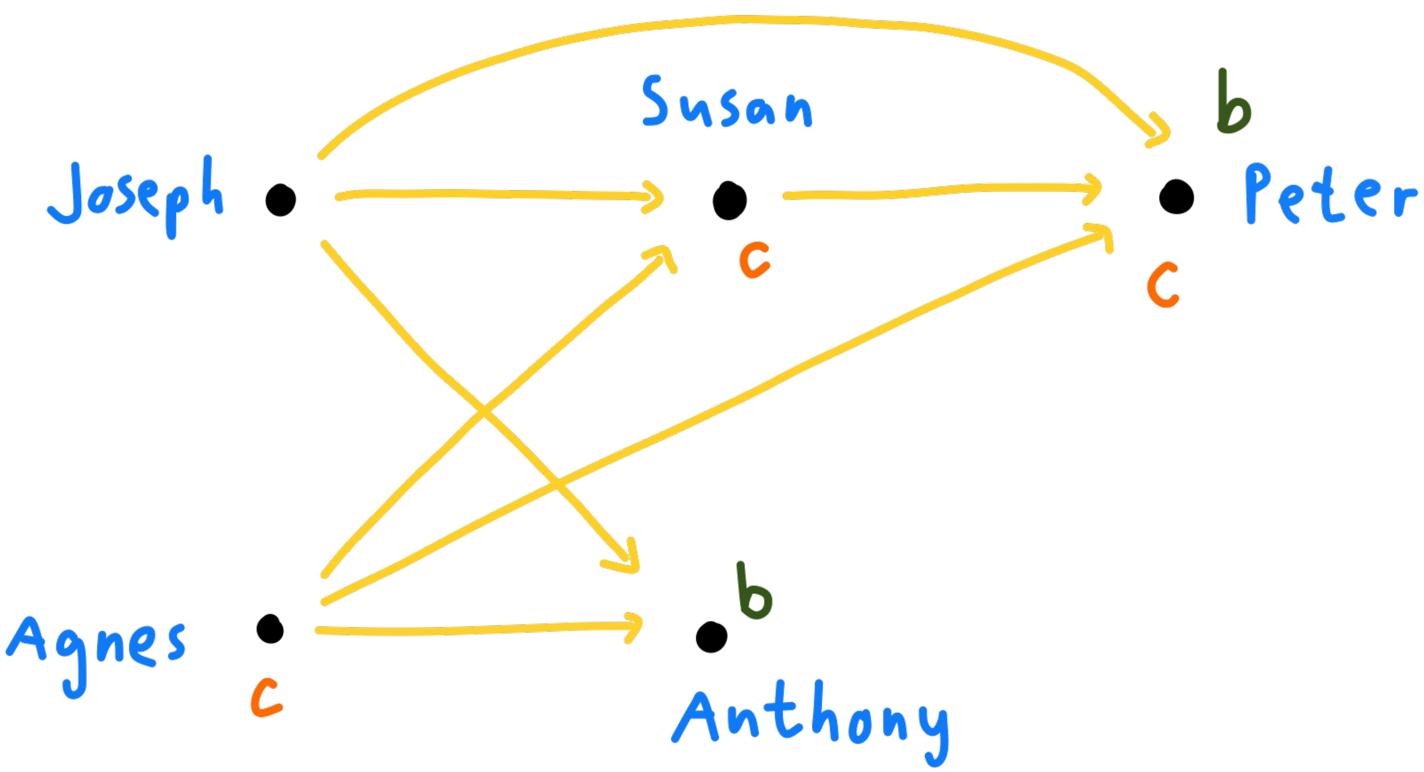
    - Or constrained to game states reachable by later play, epistemic states constrained by what agents can see, etc.

Example

b: “likes broccoli”

c: “likes carrot”

Rwu: “*u* is a descendant of *w*”



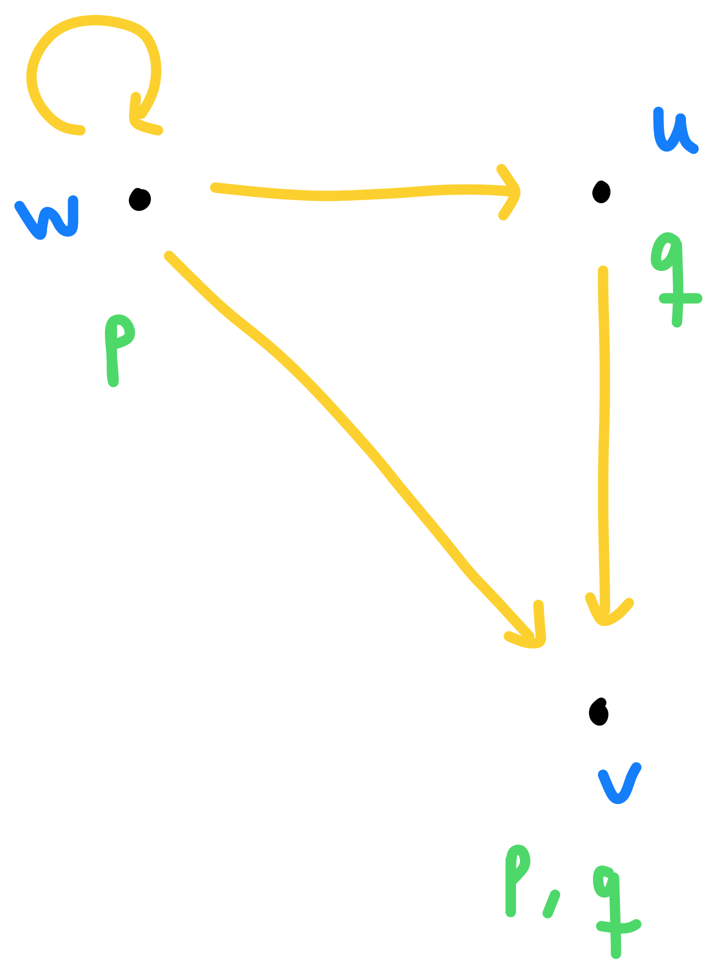
$W = \{$Joseph, Susan, Peter, Agnes, Anthony$\}$

RJosephSusan: “Susan is a descendant of Joseph”

RJosephPeter, RAgnesSusan, etc.

$V(b) = \{$Anthony, Peter$\}$ $V(c) = \{$Agnes, Susan, Peter$\}$

Another Example



$$W = \{w, v, u\}$$

$$R = \{(w, w), (w, u), (w, v), (u, v)\}$$

$$V(p) = \{w, v\}$$

$$V(q) = \{u, v\}$$

Semantics

- The semantic interpretation tells us exactly when formulas are true (and when they are false)

Definition (Semantic Interpretation/Truth)

Let $M = (W, R, V)$ be a relational model and $w \in W$ be a world.

Truth of a modal formula ϕ at world w in M, written $M, w \Vdash \phi$, is defined inductively as follows:

$M, w \Vdash p$ **iff** $w \in V(p)$

$M, w \Vdash \neg\phi$ **iff** $M, w \nVdash \phi$

$M, w \Vdash \phi \wedge \psi$ **iff** $M, w \Vdash \phi$ **and** $M, w \Vdash \psi$

$M, w \Vdash \phi \vee \psi$ **iff** $M, w \Vdash \phi$ **or** $M, w \Vdash \psi$

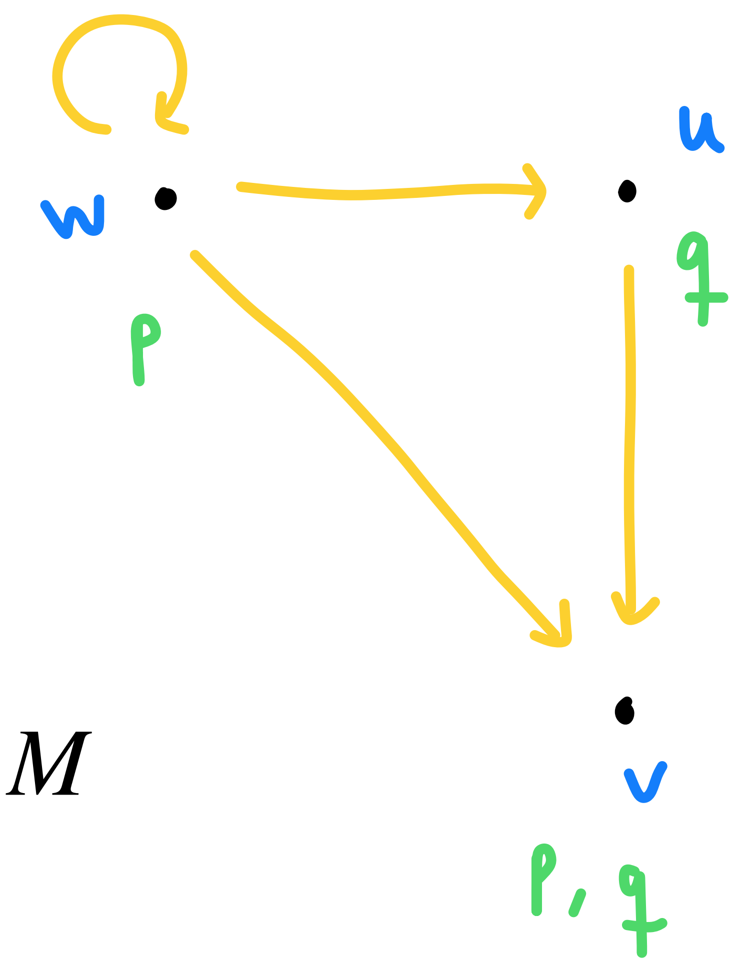
$M, w \Vdash \phi \rightarrow \psi$ **iff** (**if** $M, w \Vdash \phi$ **then** $M, w \Vdash \psi$)

$M, w \Vdash \Box\phi$ **iff for all** $v \in W$ **such that** $Rwv : M, v \Vdash \phi$

$M, w \Vdash \Diamond\phi$ **iff for some** $v \in W$ **such that** $Rwv : M, v \Vdash \phi$

We call (M, w) a **pointed model**

Return to the Example



M

$$W = \{w, v, u\}$$

$$R = \{(w,w), (w,u), (w,v), (u,v)\}$$

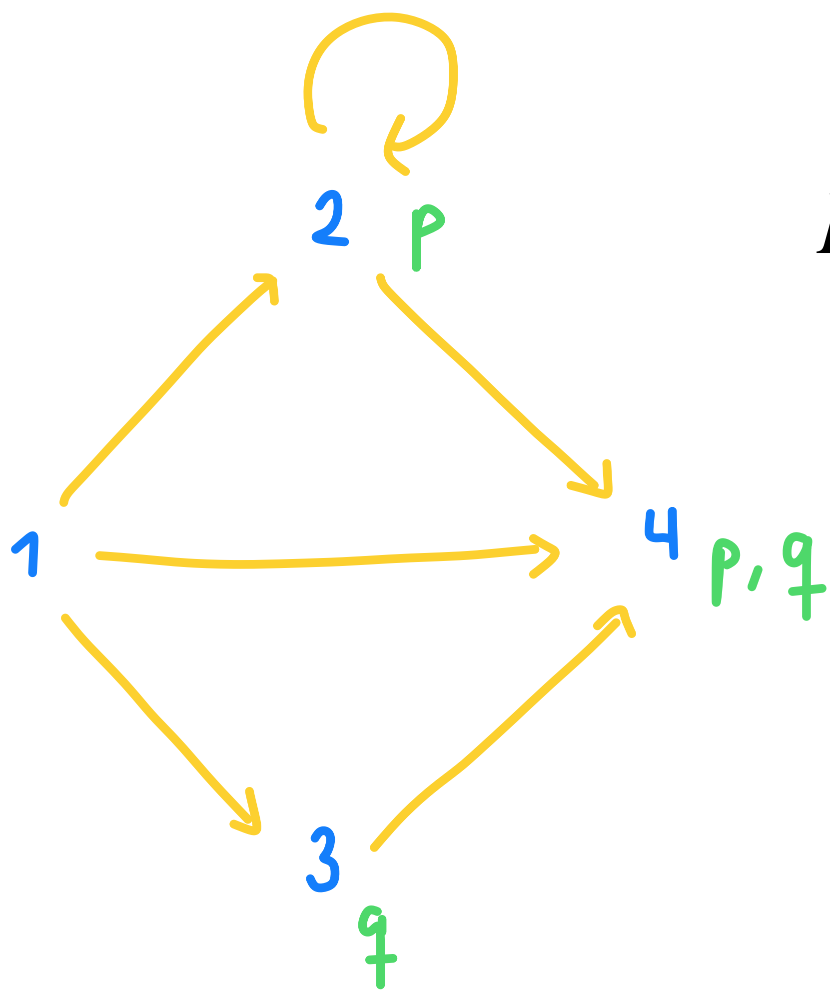
$$V(p) = \{w, v\}$$

$$V(q) = \{u, v\}$$

Which formulas are true at (M, w)?

$p \wedge q$ ✗	$\Diamond(p \wedge \neg q)$ ✓
$\Diamond(p \wedge q)$ ✓	$\Box\Diamond\top$ ✗
$\Diamond\Diamond(p \wedge q)$ ✓	$\Diamond\Box\bot$ ✓
$\Box(p \wedge q)$ ✗	

Another Example



$$W = \{1,2,3,4\}$$

$$R = \{(1,2), (2,2), (1,4), (2,4), (1,3), (3,4)\}$$

$$V(p) = \{2,4\}$$

$$V(q) = \{3,4\}$$

In what worlds is $\Diamond \Box p$ true?

1,2,3

In what worlds is $\Box (q \rightarrow p)$ true?

2,3,4

A Note on Redundancy

- From propositional logic, recall that $\vee$ can be defined in terms of $\wedge$ and $\neg$: $\phi \vee \psi := \neg(\neg\phi \wedge \neg\psi)$
  - And that $\wedge$ can be defined in terms of $\vee$ and $\neg$ in a similar way
- Also, recall that $\phi \to \psi := \neg(\phi \wedge \neg\psi)$
- And $\phi \leftrightarrow \psi := (\phi \to \psi) \wedge (\psi \to \phi)$
- Similarly, logical constants $\top$ (“always true”) and $\bot$ (“always false”) can be defined as $\top := p \vee \neg p$ and $\bot := p \wedge \neg p$
- Finally, we also have that $\Box\phi := \neg\Diamond\neg\phi$
- Therefore, when we present the language (and the semantics) of basic modal logic, we often just present it with: $\phi ::= \mathrm{At} \mid \neg\phi \mid (\phi \wedge \psi) \mid \Box\phi$

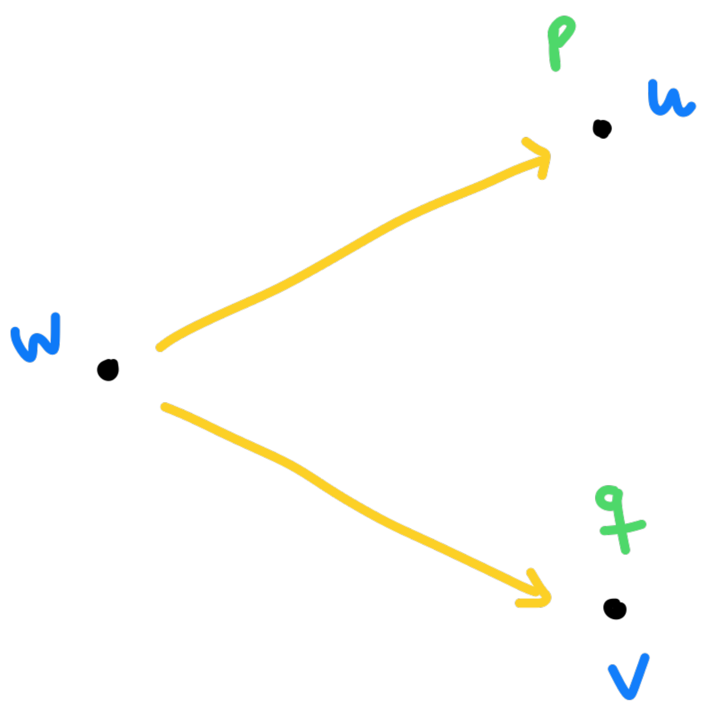
Finding Pointed Models for Formulas

- Sometimes, we might want to build a pointed model to show that a formula can be true
- And other times, similarly, we might want to build a pointed model to show that a formula is not always true (that it can be false)

Example: Finding Pointed Models for Formulas

Take the formula $\Diamond(p \wedge q) \leftrightarrow (\Diamond p \wedge \Diamond q)$

Can we find a pointed model where the formula is false?



$$M, w \nVdash \Diamond(p \wedge q) \leftrightarrow (\Diamond p \wedge \Diamond q)$$

$$M, u \Vdash \Diamond(p \wedge q) \leftrightarrow (\Diamond p \wedge \Diamond q)$$

Because $M, u \nVdash \Diamond(p \wedge q)$ and $M, u \nVdash \Diamond p \wedge \Diamond q$

Can we find a pointed model where the formula is true?

Because $M, w \Vdash \Diamond p \wedge \Diamond q$ and $M, w \nVdash \Diamond(p \wedge q)$

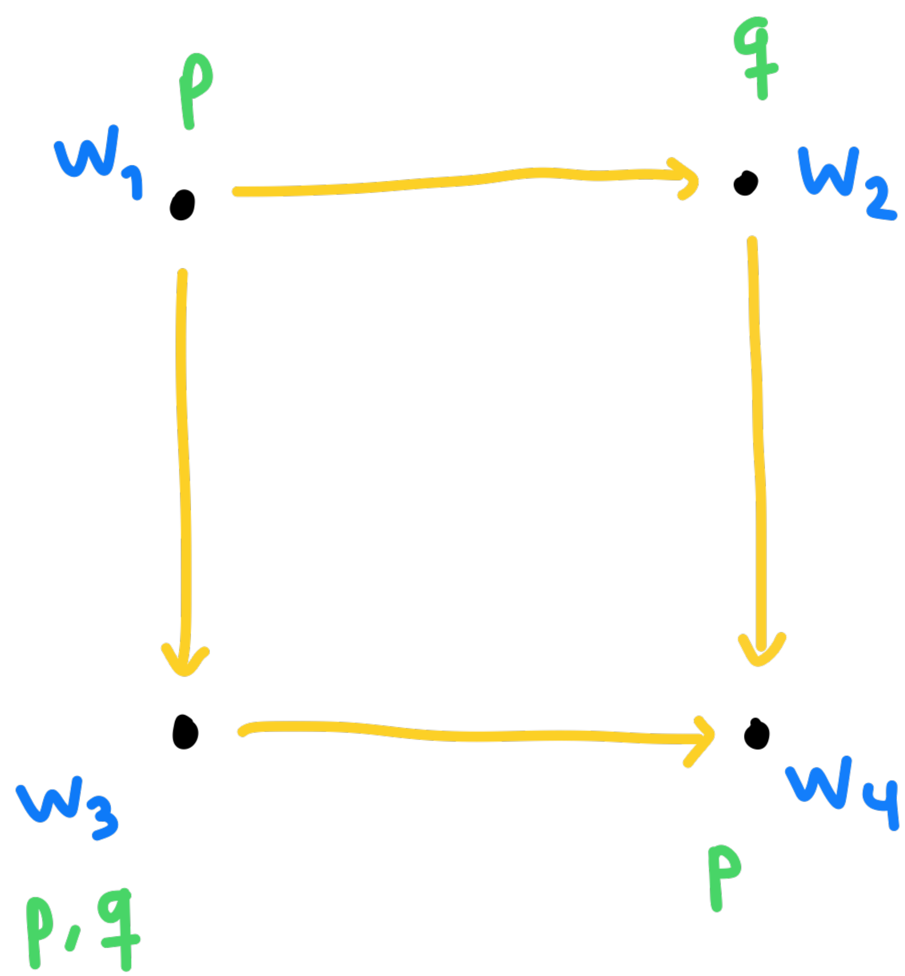
$\downarrow$

Because $M, w \nVdash (\Diamond p \wedge \Diamond q) \rightarrow \Diamond(p \wedge q)$

Finding Formulas for Worlds in a Model

- Sometimes we might want to find formulas that **characterize** worlds in particular models, i.e., a formula that is true at a given world, but false in all others

Example: Finding Formulas for Worlds in a Model



Can we find, for every world w_1, w_2, w_3, w_4 a formula that is true at the given world but false at all others?

$\neg p \wedge q$ characterizes w_2

Because w_2 is the only world where p is false and q is true

$p \wedge q$ characterizes w_3

Because w_3 is the only world where p and q are both true

$\Box \perp$ characterizes w_4

Because w_4 is the only world that does not have any successors

$p \wedge \neg q \wedge \neg \Box \perp$ characterizes w_1

Because w_1 is the only world where p is true and q is false that is not w_4

Validity

- When a formula is true in any world, in any model, we say that the formula is **valid**
- We say that a formula is **valid in a frame** when the formula is true at any state, with any valuation, on a given frame

Definition (Modal Validity)

A modal formula ϕ is **valid**, written as $\Vdash \phi$, if $M, w \Vdash \phi$ for all models and worlds.

A modal formula ϕ is **valid in a frame**, written as $F \Vdash \phi$, if it is true at every world w in every model (F, V) based on F.

Proving Validity

- We have already seen some valid formulas
  - Namely $\Box \phi \leftrightarrow \neg \Diamond \neg \phi$ and $\Diamond \phi \leftrightarrow \neg \Box \neg \phi$
- But how to we prove that a formula is valid?
  - We need to show that the formula is true in any model, at any state
  - To do that, we pick an *arbitrary* model M and an *arbitrary* state w in M, where we do not assume anything about them
  - If the formula is true in this arbitrary pointed model (M, w), it must be true in *any* pointed model

Example: Proving Validity

Prove that the formula $\Diamond(p \vee q) \to (\Diamond p \vee \Diamond q)$ is valid

Proof.

Fix an arbitrary model $M = (W, R, V)$, and an arbitrary state $w \in W$.

We want to prove that $M, w \Vdash \Diamond(p \vee q) \to (\Diamond p \vee \Diamond q)$.

Assume that $M, w \Vdash \Diamond(p \vee q)$.

Then, according to the semantics, there exists a world $v \in W$ such that Rwv and $M, v \Vdash p \vee q$.

So, again by the semantics, $M, v \Vdash p$ or $M, v \Vdash q$.

Since Rwv, $M, w \Vdash \Diamond p$ or $M, w \Vdash \Diamond q$.

Either way, $M, w \Vdash \Diamond p \vee \Diamond q$.

Since we assumed that $M, w \Vdash \Diamond(p \vee q)$, it follows that

$$M, w \Vdash \Diamond(p \vee q) \to (\Diamond p \vee \Diamond q).$$

Since we fixed arbitrary M and w, it follows that

$\Diamond(p \vee q) \to (\Diamond p \vee \Diamond q)$ is valid.

Disproving Validity

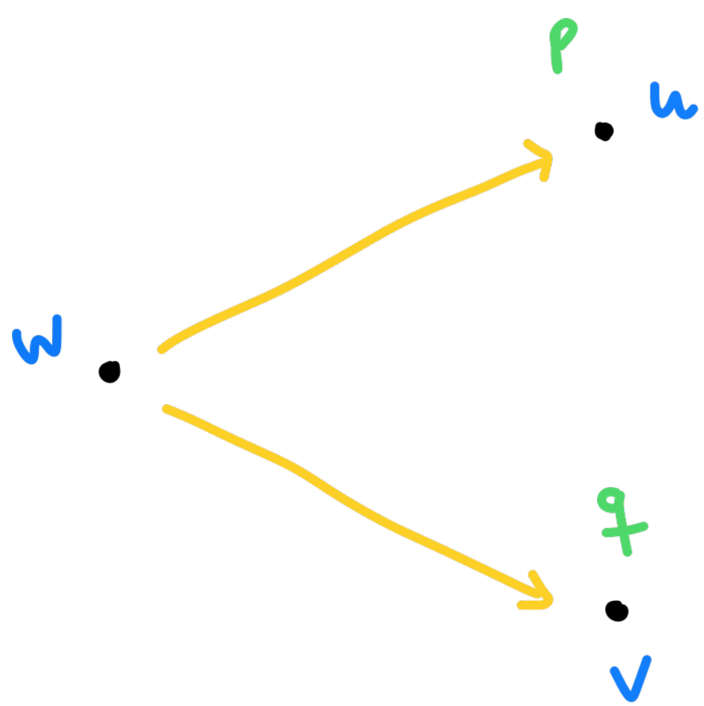
- How about proving that a formula is *not* valid?
    - Then we need to give a *counterexample* where the formula is not true
    - This shows that the formula is not always true in any model at any state

Example: Disproving Validity

We return to an earlier example:

Prove that the formula $(\Diamond p \land \Diamond q) \to \Diamond(p \land q)$ is not valid

Proof.



Because $M, w \Vdash \Diamond p \land \Diamond q$
and $M, w \nVdash \Diamond(p \land q)$

$\downarrow$

Because
$M, w \nVdash (\Diamond p \land \Diamond q) \to \Diamond(p \land q)$

$$M, w \nVdash \Diamond(p \land q) \leftrightarrow (\Diamond p \land \Diamond q)$$

In homework and on the exam, remember
to explain your solution!

Frame Definability

- There is a deep connection between valid formulas and frame properties
- Some formulas are valid on a frame if and only if the frame has a particular property

Definition (Definability)

A modal formula ϕ **defines,** or **characterizes** a property K if for all frames F, F has property K if and only if $F \Vdash \phi$.

Proving Frame Definability

- How to we prove that a formula ϕ defines a property K?
- We want to prove that, for any F, $F \Vdash \phi$ if and only if F has property K
    - Therefore, we have to prove both directions:
        - Fix an arbitrary F, assume that $F \Vdash \phi$ and then prove that F has property K
        - And fix an arbitrary F, assume that it has property K, and prove that $F \Vdash \phi$

Example: Proving Frame Definability

Prove that $F \Vdash \Diamond\Diamond p \to \Diamond p$ if and only if F is transitive

Recall that transitivity is the property such that for any $w, v, u \in W$: if Rwv and Rvu, then Rwu

Proof.

($\Leftarrow$) Fix an arbitrary frame $F = (W, R)$, and assume that F is transitive.

We want to prove that $F \Vdash \Diamond\Diamond p \to \Diamond p$.

Fix an arbitrary valuation V on F and name $M = (F, V)$.

Fix an arbitrary $w \in W$, and assume that $M, w \Vdash \Diamond\Diamond p$.

By the semantics, we have that there exists $v \in W$ such that Rwv and $M, v \Vdash \Diamond p$.

By the semantics, we have that there exists $u \in W$ such that Rvu and $M, u \Vdash p$.

Since Rwv and Rvu and the underlying frame F is transitive, then Rwu.

Then by the semantics, $M, w \Vdash \Diamond p$ and therefore $M, w \Vdash \Diamond\Diamond p \to \Diamond p$.

Since we fixed arbitrary F, V and w, it follows that $F \Vdash \Diamond\Diamond p \to \Diamond p$.

Example: Proving Frame Definability

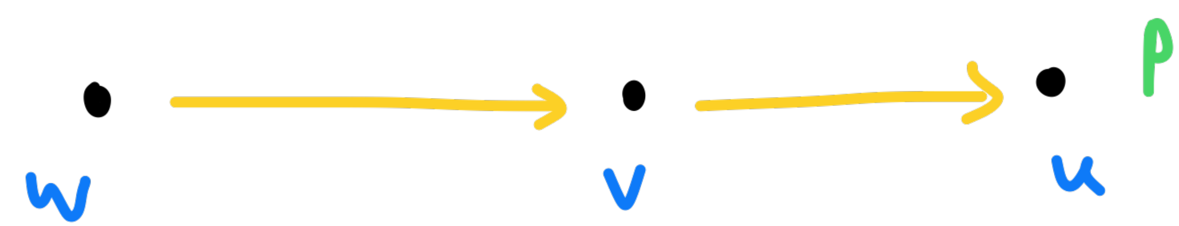
Prove that $F \Vdash \Diamond\Diamond p \to \Diamond p$ if and only if F is transitive

Recall that transitivity is the property such that for any $w, v, u \in W$: if Rwv and Rvu, then Rwu

Proof.

($\Rightarrow$) **Proof by contraposition: We assume that F is not transitive, and we want to prove that $F \nVdash \Diamond\Diamond p \to \Diamond p$.**

M



In this model, $M, w \Vdash \Diamond\Diamond p$, but $M, w \nVdash \Diamond p$.

Thus $M, w \nVdash \Diamond\Diamond p \to \Diamond p$.

Therefore $F \nVdash \Diamond\Diamond p \to \Diamond p$.

Coming up next: modal logic and multi-agent systems…

Additional Reading Material

- For additional reading material, I advise:
    - Johan van Benthem: *Modal Logic for Open Minds*. Center for the Study of Language and Information, 2010.
    - (For more mathematical content):
        - Patrick Blackburn, Maarten de Rijke and Yde Venema: *Modal Logic*. Cambridge University Press, 2001.