# CS 295A/395D: Artificial Intelligence

Epistemic Logic: Knowledge & Belief Prof. Emma Tosch 23 March 2022



The University of Vermont

### **"Modes" of truth**

Recall: logic with respect to a knowledge base

Contextualizes statements – need abstract notion of context

- Informally, time: when is something true?
  - When must something be true? Refers not to time, but "in what context?"
- Formally more abstract than time: time has a specific sequential meaning, but we may want a broader definition

## **Recap: Semantics**

Recall: a formal semantics is a mapping from a surface string (syntax) to an underlying structure that gives a surface string meaning.

**Propositional logic example**: Given assignment  $\mathcal{A} = \{p : 1, q : 0, r : 1\}, \qquad \qquad \mathcal{A} \models (p \lor q) \rightarrow r$ 

**Predicate logic example**: Given  $\mathcal{U} = \mathbb{N}$  and structure  $\mathcal{M} = (\mathcal{F} = \{+\}, \mathcal{P} = \{=\}), \mathcal{M} \models \forall n \exists m (m = n + 1)$ 

Finding an assignment that models a formula == searching for a satisfying assignment (SAT)

Finding a structure to model a predicate formula  $\rightarrow$  not emphasized

Take everything from propositional logic and add:

- $\Box$  ("box" = "necessity"  $\rightarrow$  like  $\forall$ )
- $\diamond$  ("diamond" = "possibility"  $\rightarrow$  like  $\exists$ )

Take everything from propositional logic and add:

- $\Box$  ("box" = "necessity"  $\rightarrow$  like  $\forall$ )
- $\diamond$  ("diamond" = "possibility"  $\rightarrow$  like  $\exists$ )

p	$\neg p$	$q \wedge p$	p  ightarrow q	$p \lor q$

Take everything from propositional logic and add:

- $\Box$  ("box" = "necessity"  $\rightarrow$  like  $\forall$ )
- $\diamond$  ("diamond" = "possibility"  $\rightarrow$  like  $\exists$ )

p	$\neg p$	$q \wedge p$	p  ightarrow q	$p \lor q$
$\Box p$				

Take everything from propositional logic and add:

- $\Box$  ("box" = "necessity"  $\rightarrow$  like  $\forall$ )
- $\diamond$  ("diamond" = "possibility"  $\rightarrow$  like  $\exists$ )

p	$\neg p$	$q \wedge p$	ho  ightarrow q	$p \lor q$
$\Box p$	$\diamond \neg p$			

Take everything from propositional logic and add:

- $\Box$  ("box" = "necessity"  $\rightarrow$  like  $\forall$ )
- $\diamond$  ("diamond" = "possibility"  $\rightarrow$  like  $\exists$ )

p	$\neg p$	$q \wedge p$	ho  ightarrow q	$p \lor q$
$\Box p$	$\Diamond \neg p$	$\Box(q \land p)$	$\Diamond(p  ightarrow q)$	

Take everything from propositional logic and add:

- $\Box$  ("box" = "necessity"  $\rightarrow$  like  $\forall$ )
- $\diamond$  ("diamond" = "possibility"  $\rightarrow$  like  $\exists$ )

p	$\neg p$	$q \wedge p$	ho  ightarrow q	$p \lor q$
$\Box p$	$\Diamond \neg p$	$\Box(q \land p)$	$\diamond(p  ightarrow q)$	$q \lor \Box p$

Take everything from propositional logic and add:

- $\Box$  ("box" = "necessity"  $\rightarrow$  like  $\forall$ )
- $\diamond$  ("diamond" = "possibility"  $\rightarrow$  like  $\exists$ )

p	$\neg p$	$q \wedge p$	p  ightarrow q	$p \lor q$
$\Box p$	$\diamond \neg p$	$\Box(q \land p)$	$\Diamond(p  ightarrow q)$	$q \lor \Box p$
$\Box \Box p$				

Take everything from propositional logic and add:

- $\Box$  ("box" = "necessity"  $\rightarrow$  like  $\forall$ )
- $\diamond$  ("diamond" = "possibility"  $\rightarrow$  like  $\exists$ )

p	$\neg p$	$q \wedge p$	p  ightarrow q	$p \lor q$
$\Box p$	$\diamond \neg p$	$\Box(q \land p)$	$\Diamond(p  ightarrow q)$	$q \lor \Box p$
$\Box \Box p$	$\neg \diamondsuit p$			

Take everything from propositional logic and add:

- $\Box$  ("box" = "necessity"  $\rightarrow$  like  $\forall$ )
- $\diamond$  ("diamond" = "possibility"  $\rightarrow$  like  $\exists$ )

	p	$\neg p$	$q \wedge p$	p  ightarrow q	$p \lor q$
[	] <i>p</i>	$\Diamond \neg p$	$\Box(q \land p)$	$\Diamond(p  ightarrow q)$	$q \lor \Box p$
	$\Box p$	$\neg \diamondsuit p$	$\Box q \land \Diamond p$	$\Diamond(p \rightarrow \Box q)$	

Take everything from propositional logic and add:

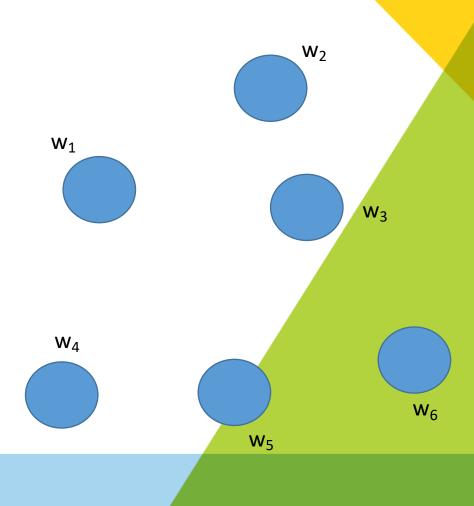
- $\Box$  ("box" = "necessity"  $\rightarrow$  like  $\forall$ )
- $\diamond$  ("diamond" = "possibility"  $\rightarrow$  like  $\exists$ )

p	$\neg p$	$q \wedge p$	ho  ightarrow q	$p \lor q$
$\Box p$	$\diamond \neg p$	$\Box(q \land p)$	$\Diamond(p  ightarrow q)$	$q \lor \Box p$
$\Box \Box p$	$\neg \diamondsuit p$	$\Box q \land \Diamond p$	$\Diamond(p \rightarrow \Box q)$	$\Box \diamondsuit (q \lor \Box p)$

## **Basic modal logic: Model specification**

Model structure:  $\mathcal{M} = (W, R, L)$ 

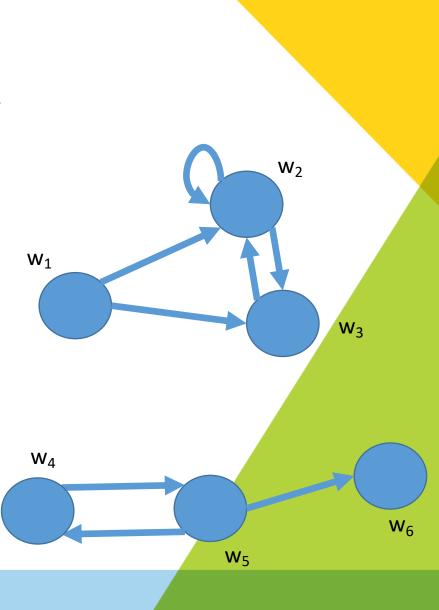
• W: set of "worlds" (nodes in a graph),



## **Basic modal logic: Model specification**

Model structure:  $\mathcal{M} = (W, R, L)$ 

- W: set of "worlds" (nodes in a graph),
- R: binary "accessibility relation" on W (edges in a graph),

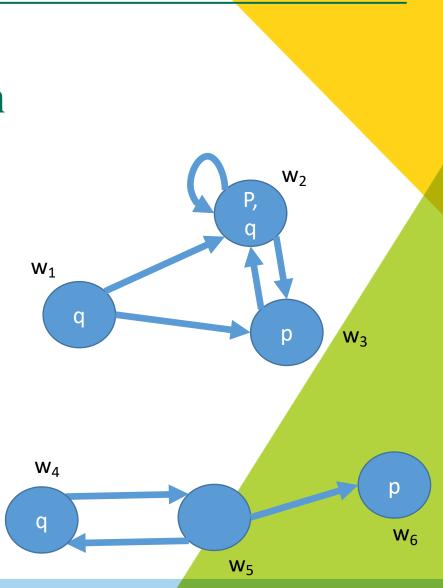


## **Basic modal logic: Model specification**

Model structure:  $\mathcal{M} = (W, R, L)$ 

- W: set of "worlds" (nodes in a graph),
- R: binary "accessibility relation" on W (edges in a graph),
- L: "labeling function" from each world to a subset of atoms,

Where the set of atoms is the set of propositions.



 $w \Vdash \Diamond \varphi \quad \text{iff} \exists w' \in W (R(w, w') \land w' \Vdash \psi)$ 

 $w \Vdash \Box \varphi \quad \text{iff} \; \forall \; \mathsf{w'} \in \mathsf{W} \; (\mathsf{R}(\mathsf{w}, \mathsf{w'}) \to \mathsf{w'} \Vdash \psi)$ 

 $w \Vdash \varphi \rightarrow \psi$  iff  $w \Vdash \varphi$  whenever  $w \Vdash \psi$ 

 $w \Vdash \varphi \lor \psi$  iff at least one of  $w \Vdash \varphi$  or  $w \Vdash \psi$ 

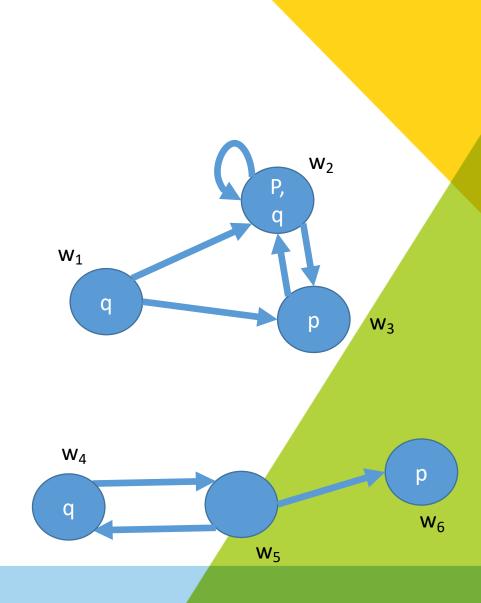
 $w \Vdash \varphi \land \psi$  iff  $w \Vdash \varphi$  and  $w \Vdash \psi$ 

 $w \Vdash a \text{ iff } a \in L(w)$  $w \Vdash \neg \varphi \text{ iff } w \Vdash \varphi$ 

*w* ⊩ ⊤

w⊮⊥⊥

Given Model structure:  $\mathcal{M} = (W, R, L)$ . Let  $w \in W$ . Then:



 $w \Vdash \Diamond \varphi \quad \text{iff} \exists w' \in W (\mathbb{R}(w, w') \land w' \Vdash \psi)$ 

 $w \Vdash \Box \varphi \text{ iff } \forall w' \in W (R(w, w') \to w' \Vdash \psi)$ 

 $w\Vdash \varphi \to \psi \text{ iff } w\Vdash \varphi \text{ whenever } w\Vdash \psi$ 

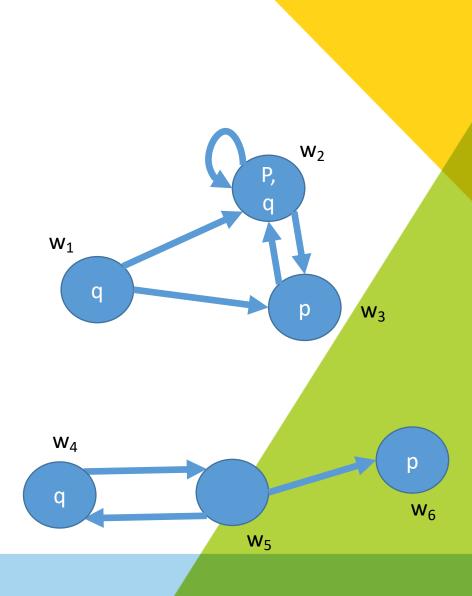
 $w \Vdash \varphi \lor \psi$  iff at least one of  $w \Vdash \varphi$  or  $w \Vdash \psi$ 

 $w \Vdash \varphi \land \psi$  iff  $w \Vdash \varphi$  and  $w \Vdash \psi$ 

 $w \Vdash a \text{ iff } a \in L(w)$  $w \Vdash \neg \varphi \text{ iff } w \Vdash \varphi$ 

*w* ⊪ ⊤ *w* ⊮⊥

Given Model structure:  $\mathcal{M} = (W, R, L)$ . Let  $w \in W$ . Then:



 $w \Vdash \Diamond \varphi \quad \text{iff} \exists w' \in W (\mathbb{R}(w, w') \land w' \Vdash \psi)$ 

 $w \Vdash \Box \varphi \quad \text{iff} \; \forall \; \mathsf{w'} \in \mathsf{W} \; (\mathsf{R}(\mathsf{w}, \mathsf{w'}) \to \mathsf{w'} \Vdash \psi)$ 

 $w \Vdash \varphi \to \psi \text{ iff } w \Vdash \varphi \text{ whenever } w \Vdash \psi$ 

 $w \Vdash \varphi \lor \psi$  iff at least one of  $w \Vdash \varphi$  or  $w \Vdash \psi$ 

 $w \Vdash \varphi \land \psi$  iff  $w \Vdash \varphi$  and  $w \Vdash \psi$ 

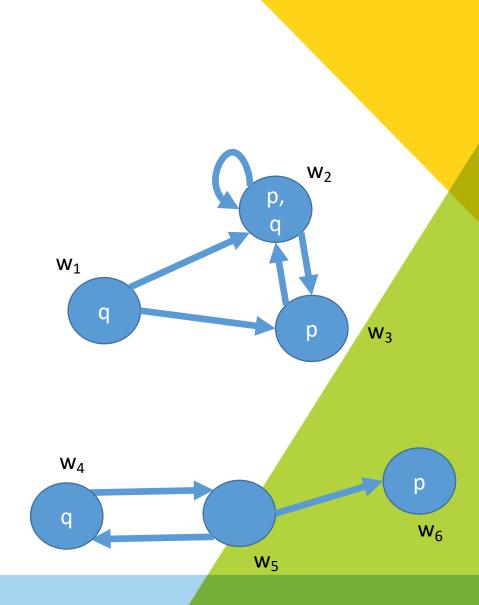
 $w \Vdash \neg \varphi$  iff  $w \nvDash \varphi$ 

 $w \Vdash p \text{ iff } p \in L(w)$ 

 $w \Vdash \mathsf{T}$ 

w⊮⊬⊥

Given Model structure:  $\mathcal{M} = (W, R, L)$ . Let  $w \in W$ . Then:



Given Model structure:  $\mathcal{M} = (W, R, L)$ . Let  $w \in W$ . Then:

 $w \Vdash \mathsf{T}$ 

w⊮⊬⊥

 $w \Vdash p \text{ iff } p \in L(w)$ 

 $w \Vdash \neg \varphi$  iff  $w \nvDash \varphi$ 

 $w \Vdash \varphi \land \psi$  iff  $w \Vdash \varphi$  and  $w \Vdash \psi$ 

 $w \Vdash \Box \varphi \quad \text{iff} \; \forall \; \mathsf{w'} \in \mathsf{W} \; (\mathsf{R}(\mathsf{w}, \mathsf{w'}) \to \mathsf{w'} \Vdash \psi)$ 

 $w \Vdash \Diamond \varphi \text{ iff } \exists w' \in W (\mathsf{R}(w, w') \land w' \Vdash \psi)$ 

 $w \Vdash \varphi \lor \psi$  iff at least one of  $w \Vdash \varphi$  or  $w \Vdash \psi$ 

 $w \Vdash \varphi \rightarrow \psi$  iff  $w \Vdash \varphi$  whenever  $w \Vdash \psi$ 

 $W_2$ Ρ,  $W_1$ W<sub>3</sub>  $W_4$ Q  $W_6$  $W_5$ 

Let  $\varphi$  = q and  $\psi$  = p

 $w \Vdash \Diamond \varphi$  iff  $\exists w' \in W (R(w, w') \land w' \Vdash \psi)$ 

 $w \Vdash \Box \varphi \quad \text{iff} \; \forall \; \mathsf{w'} \in \mathsf{W} \; (\mathsf{R}(\mathsf{w}, \mathsf{w'}) \to \mathsf{w'} \Vdash \psi)$ 

 $w \Vdash \varphi \rightarrow \psi$  iff  $w \Vdash \varphi$  whenever  $w \Vdash \psi$ 

 $w \Vdash \varphi \lor \psi$  iff at least one of  $w \Vdash \varphi$  or  $w \Vdash \psi$ 

 $w \Vdash \varphi \land \psi$  iff  $w \Vdash \varphi$  and  $w \Vdash \psi$ 

 $w \Vdash \neg \varphi$  iff  $w \nvDash \varphi$ 

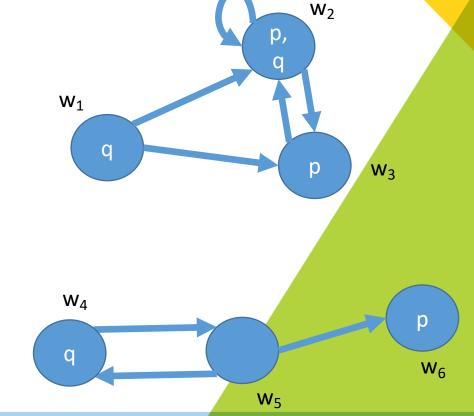
 $w \Vdash \mathsf{T}$ 

w⊮⊬⊥

 $w \Vdash p \text{ iff } p \in L(w)$ 

**Basic modal logic: Semantics** 

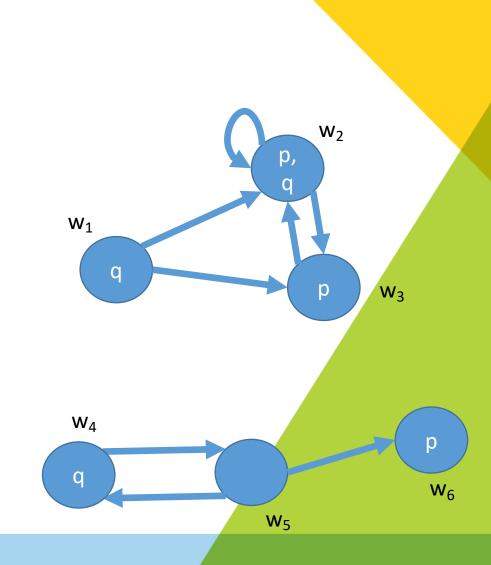
Given Model structure:  $\mathcal{M} = (W, R, L)$ . Let  $w \in W$ . Then:



 $w \Vdash \downarrow \bot$   $w \Vdash p \text{ iff } p \in L(w)$   $w \Vdash \neg \varphi \text{ iff } w \Vdash \varphi$   $w \Vdash \varphi \land \psi \text{ iff } w \Vdash \varphi \text{ and } w \Vdash \psi$   $w \Vdash \varphi \land \psi \text{ iff } at \text{ least one of } w \Vdash \varphi \text{ or } w \Vdash \psi$   $w \Vdash \varphi \rightarrow \psi \text{ iff } w \Vdash \varphi \text{ whenever } w \vdash \psi$   $w \Vdash \varphi \rightarrow \psi \text{ iff } w \Vdash \varphi \text{ whenever } w \vdash \psi$   $w \Vdash \Box \varphi \text{ iff } \forall w' \in W (R(w, w')) \rightarrow w' \Vdash \psi)$   $w \Vdash \Diamond \varphi \text{ iff } \exists w' \in W (R(w, w') \land w' \Vdash \psi)$ 

 $w \Vdash \mathsf{T}$ 

Given Model structure:  $\mathcal{M} = (W, R, L)$ . Let  $w \in W$ . Then:



Given Model structure:  $\mathcal{M} = (W, R, L)$ . Let  $w \in W$ . Then:

 $w \Vdash \top$ 

 $w \Vdash \!\!\! \perp$ 

 $w \Vdash p \text{ iff } p \in L(w)$ 

 $w\Vdash \neg \varphi \text{ iff } w \nvDash \varphi$ 

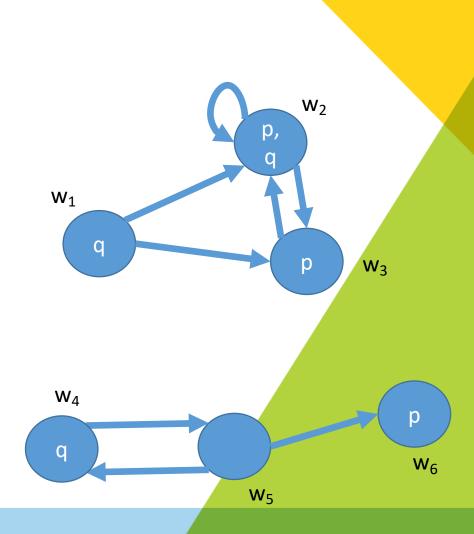
 $w \Vdash \varphi \land \psi$  iff  $w \Vdash \varphi$  and  $w \Vdash \psi$ 

 $w \Vdash \varphi \lor \psi$  iff at least one of  $w \Vdash \varphi$  or  $w \Vdash \psi$ 

 $w\Vdash \varphi \to \psi \text{ iff } w\Vdash \varphi \text{ whenever } w\Vdash \psi$ 

```
w \Vdash \Box \varphi \text{ iff } \forall w' \in \mathbb{W} (\mathbb{R}(w, w') \to w' \Vdash \psi)
```

 $w \Vdash \Diamond \varphi \text{ iff } \exists \ \mathsf{w'} \in \mathsf{W} \ (\mathsf{R}(\mathsf{w}, \mathsf{w'}) \land \ \mathsf{w'} \Vdash \psi)$ 



Find a world w such that  $w \Vdash \Box p \lor \Box q$ 

Given Model structure:  $\mathcal{M} = (W, R, L)$ . Let  $w \in W$ . Then:

 $w \Vdash \top$ 

 $w \Vdash \!\!\! \perp$ 

 $w \Vdash p \text{ iff } p \in L(w)$ 

 $w\Vdash \neg \varphi \text{ iff } w \nVdash \varphi$ 

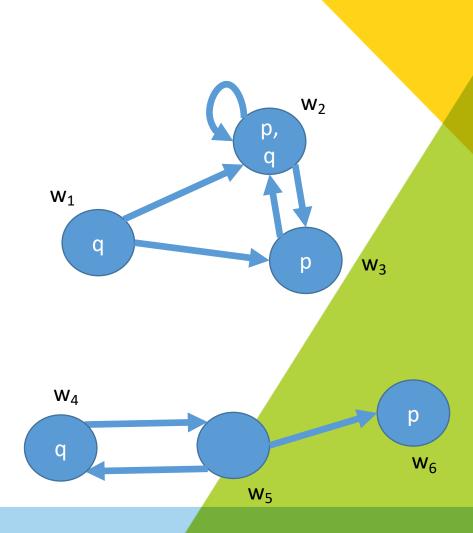
 $w \Vdash \varphi \land \psi$  iff  $w \Vdash \varphi$  and  $w \Vdash \psi$ 

 $w \Vdash \varphi \lor \psi$  iff at least one of  $w \Vdash \varphi$  or  $w \Vdash \psi$ 

 $w\Vdash \varphi \to \psi \text{ iff } w\Vdash \varphi \text{ whenever } w \Vdash \psi$ 

```
w \Vdash \Box \varphi \text{ iff } \forall w' \in \mathbb{W} (\mathbb{R}(w, w') \to w' \Vdash \psi)
```

 $w \Vdash \Diamond \varphi \text{ iff } \exists \ \mathsf{w'} \in \mathsf{W} \ (\mathsf{R}(\mathsf{w}, \mathsf{w'}) \land \ \mathsf{w'} \Vdash \psi)$ 



Find a world w such that  $w \Vdash \Box (p \lor q)$ 

Given Model structure:  $\mathcal{M} = (W, R, L)$ . Let  $w \in W$ . Then:

 $w \Vdash \top$ 

 $w \Vdash \!\!\! \perp$ 

 $w \Vdash p \text{ iff } p \in L(w)$ 

 $w\Vdash \neg \varphi \text{ iff } w \nvDash \varphi$ 

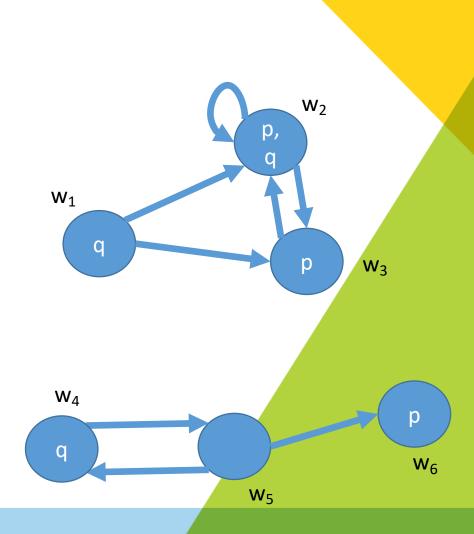
 $w \Vdash \varphi \land \psi$  iff  $w \Vdash \varphi$  and  $w \Vdash \psi$ 

 $w \Vdash \varphi \lor \psi$  iff at least one of  $w \Vdash \varphi$  or  $w \Vdash \psi$ 

 $w\Vdash \varphi \to \psi \text{ iff } w\Vdash \varphi \text{ whenever } w \Vdash \psi$ 

```
w \Vdash \Box \varphi \text{ iff } \forall w' \in \mathbb{W} (\mathbb{R}(w, w') \to w' \Vdash \psi)
```

 $w \Vdash \Diamond \varphi \text{ iff } \exists \ \mathsf{w'} \in \mathsf{W} \ (\mathsf{R}(\mathsf{w}, \mathsf{w'}) \land \ \mathsf{w'} \Vdash \psi)$ 



Find a world *w* such that  $w \Vdash \Box p \rightarrow p$ 

Given Model structure:  $\mathcal{M} = (W, R, L)$ . Let  $w \in W$ . Then:

 $w \Vdash \top$ 

 $w \Vdash \!\!\! \perp$ 

 $w \Vdash p \text{ iff } p \in L(w)$ 

 $w\Vdash \neg \varphi \text{ iff } w \nvDash \varphi$ 

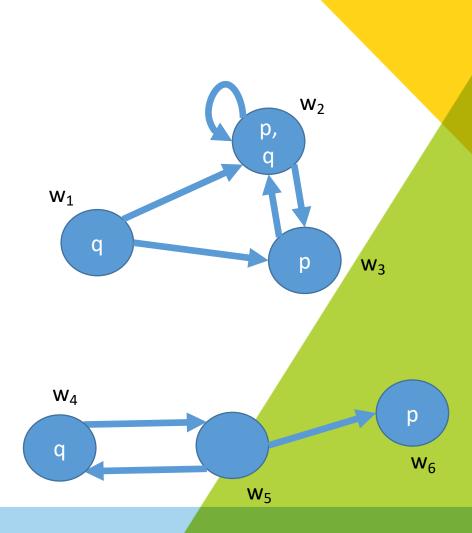
 $w \Vdash \varphi \land \psi$  iff  $w \Vdash \varphi$  and  $w \Vdash \psi$ 

 $w \Vdash \varphi \lor \psi$  iff at least one of  $w \Vdash \varphi$  or  $w \Vdash \psi$ 

 $w\Vdash \varphi \to \psi \text{ iff } w\Vdash \varphi \text{ whenever } w \Vdash \psi$ 

```
w \Vdash \Box \varphi \text{ iff } \forall w' \in \mathbb{W} (\mathbb{R}(w, w') \to w' \Vdash \psi)
```

 $w \Vdash \Diamond \varphi \quad \text{iff} \exists w' \in W (\mathsf{R}(w, w') \land w' \Vdash \psi)$ 



Find a world w such that  $w \Vdash \diamond \top$ 

Given Model structure:  $\mathcal{M} = (W, R, L)$ . Let  $w \in W$ . Then:

 $w \Vdash \top$ 

 $w \Vdash \bot$ 

 $w \Vdash p \text{ iff } p \in L(w)$ 

 $w\Vdash \neg \varphi \text{ iff } w \nVdash \varphi$ 

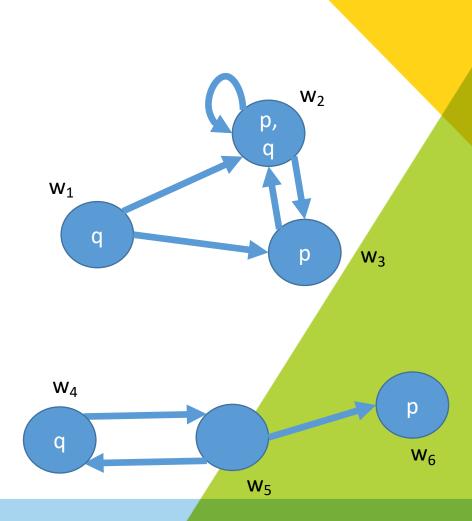
 $w \Vdash \varphi \land \psi$  iff  $w \Vdash \varphi$  and  $w \Vdash \psi$ 

 $w \Vdash \varphi \lor \psi$  iff at least one of  $w \Vdash \varphi$  or  $w \Vdash \psi$ 

 $w\Vdash \varphi \to \psi \text{ iff } w\Vdash \varphi \text{ whenever } w \Vdash \psi$ 

```
w \Vdash \Box \varphi \text{ iff } \forall w' \in \mathbb{W} (\mathbb{R}(w, w') \to w' \Vdash \psi)
```

 $w \Vdash \Diamond \varphi \text{ iff } \exists \ w' \in \mathbb{W} \ (\mathbb{R}(w, w') \land w' \Vdash \psi)$ 



#### Find a world *w* such that $w \Vdash \Box \bot$

New rules/equivalences

DeMorgan's:Distributive:Tautology, contradiction: $\neg \Box \varphi = \diamond \neg \varphi$  $\Box (\varphi \land \psi) = \Box \varphi \lor \Box \psi$  $\Box T = T$  $\neg \diamond \varphi = \Box \neg \varphi$  $\diamond (\varphi \lor \psi) = \diamond \varphi \land \diamond \psi$  $\Box T \neq \diamond T$ Connective equivalence: $\neg \Box \neg \varphi = \diamond \varphi$  $\diamond \bot \neq \Box \bot$ 

## **Basic modal logic: Stacked modals**

Given Model structure:  $\mathcal{M} = (W, R, L)$ . Let  $w \in W$ . Then:

 $w \Vdash \top$ 

 $w \Vdash \bot$ 

 $w \Vdash p \text{ iff } p \in L(w)$ 

 $w\Vdash \neg \varphi \text{ iff } w \nvDash \varphi$ 

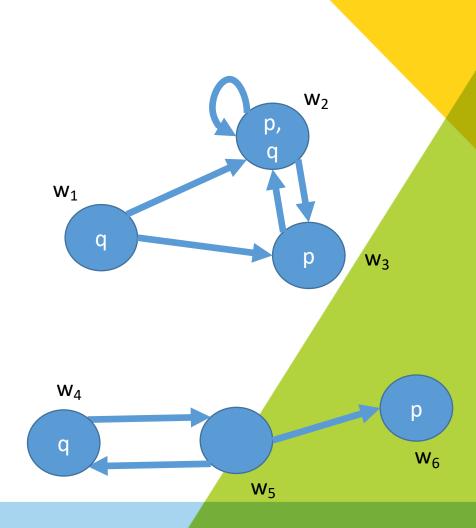
 $w \Vdash \varphi \land \psi$  iff  $w \Vdash \varphi$  and  $w \Vdash \psi$ 

 $w \Vdash \varphi \lor \psi$  iff at least one of  $w \Vdash \varphi$  or  $w \Vdash \psi$ 

 $w\Vdash \varphi \to \psi \text{ iff } w\Vdash \varphi \text{ whenever } w\Vdash \psi$ 

```
w \Vdash \Box \varphi \text{ iff } \forall w' \in W (R(w, w') \rightarrow w' \Vdash \psi)
```

 $w \Vdash \Diamond \varphi \text{ iff } \exists \ \mathsf{w'} \in \mathsf{W} \ (\mathsf{R}(\mathsf{w}, \mathsf{w'}) \land \ \mathsf{w'} \Vdash \psi)$ 



Find a world w such that  $w \Vdash \Box \diamond q$ 

## **Basic modal logic: Stacked modals**

Given Model structure:  $\mathcal{M} = (W, R, L)$ . Let  $w \in W$ . Then:

 $w \Vdash \top$ 

 $w \Vdash \bot$ 

 $w \Vdash p \text{ iff } p \in L(w)$ 

 $w\Vdash \neg \varphi \text{ iff } w \nVdash \varphi$ 

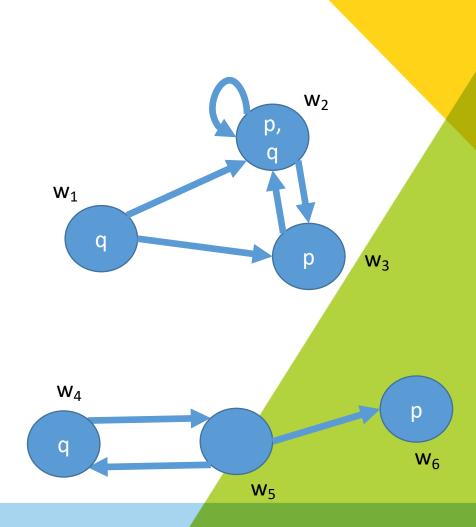
 $w \Vdash \varphi \land \psi$  iff  $w \Vdash \varphi$  and  $w \Vdash \psi$ 

 $w \Vdash \varphi \lor \psi$  iff at least one of  $w \Vdash \varphi$  or  $w \Vdash \psi$ 

 $w\Vdash \varphi \to \psi \text{ iff } w\Vdash \varphi \text{ whenever } w \Vdash \psi$ 

```
w \Vdash \Box \varphi \text{ iff } \forall w' \in \mathbb{W} (\mathbb{R}(w, w') \to w' \Vdash \psi)
```

 $w \Vdash \Diamond \varphi \text{ iff } \exists \ \mathsf{w'} \in \mathsf{W} \ (\mathsf{R}(\mathsf{w}, \mathsf{w'}) \land \ \mathsf{w'} \Vdash \psi)$ 



#### Find a world w such that $w \Vdash \diamond \diamond \diamond \Box \perp$

## Model satisfiability

So far: world safisfiability (e.g.,  $w \Vdash p$ ).

Let  $\mathcal{M} = (W, R, L)$ . If every world  $w \in W$  satisfies a formula  $\varphi$ , then we say  $\mathcal{M} \models \varphi$ .

## Valid formulas: Propositional logic

Recall: a formula is valid iff every possible assignment/structure makes it true.

Equivalently: a formula is valid iff there does not exist an assignment that could make it false.

Examples:

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

Valid formulas: Basic Modal Logic

All the same valid formulas, plus "K":

 $(\Box(\varphi \to \psi) \land \Box \varphi) \to \Box \psi$ 

Also written:

$$\Box(\varphi \to \psi) \to (\Box \varphi \to \Box \psi)$$

Work through derivation on the board

## New interpretation of modes: Knowledge, Belief

Necessity, possibility  $\rightarrow$  abstract concept

Can provide alternative interpretations of "It is necessarily true that  $\varphi$ " for the string  $\Box \varphi$ :

- "It will always be true that  $\varphi$ " (temporal logic  $\rightarrow$  coming up!)
- "It ought to be true that" (deontological  $\rightarrow$  law)
- "Agent Q believes that" (belief)
- "Agent Q knows that" (knowledge)

## New interpretation of modes: Knowledge, Belief

Necessity, possibility  $\rightarrow$  abstract concept

Can provide alternative interpretations of "It is necessarily true that  $\varphi$ " for the string  $\Box \varphi$ :

- "It will always be true that  $\varphi$ " (temporal logic  $\rightarrow$  coming up!)
- "It ought to be true that" (deontological  $\rightarrow$  law)  $\Box \varphi$
- "Agent Q believes that" (belief)
- "Agent Q knows that" (knowledge)

 $\label{eq:phi} \begin{array}{l} & \diamond \, \varphi \\ \\ " \varphi \text{ is consistent with Q's beliefs"} \end{array}$ 

## New interpretation of modes: Knowledge, Belief

Necessity, possibility  $\rightarrow$  abstract concept

Can provide alternative interpretations of "It is necessarily true that  $\varphi$ " for the string  $\Box \varphi$ :

- "It will always be true that  $\varphi$ " (temporal logic  $\rightarrow$  coming up!)
- "It ought to be true that" (deontological  $\rightarrow$  law)
- "Agent Q believes that" (belief)
- "Agent Q knows that" (knowledge)  $\Box \varphi$

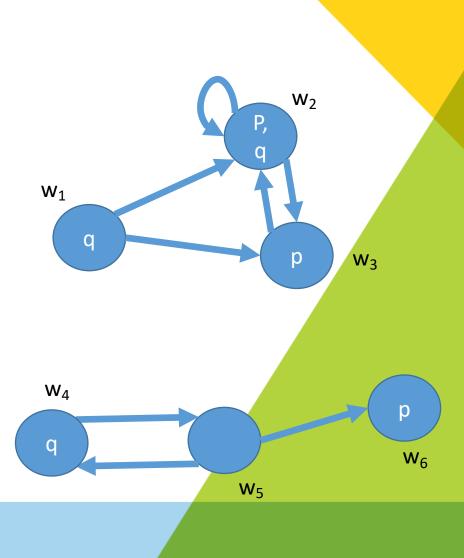
•  $\varphi$ "For all Q knows,  $\varphi$ "

Both interpretations of possible (diamond) are more intuitive under  $\neg \Box \neg \varphi = \diamond \varphi$ 

## **Capturing domain-specific axioms**

Returning to the totally abstracted case:

Is  $\Box p \rightarrow p$  valid?



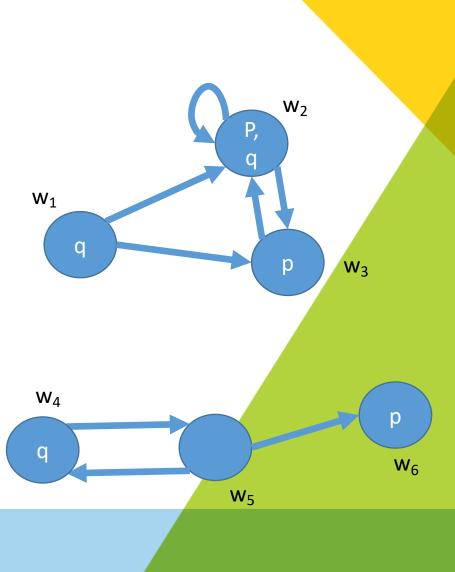
## **Capturing domain-specific axioms**

Returning to the totally abstracted case:

Is  $\Box p \rightarrow p$  valid?

If  $\Box$  represents **belief**, is  $\Box p \rightarrow p$ ?

(Recall:  $\Box p$  means "Agent Q believes p to be true")



## **Capturing domain-specific axioms**

Returning to the totally abstracted case:

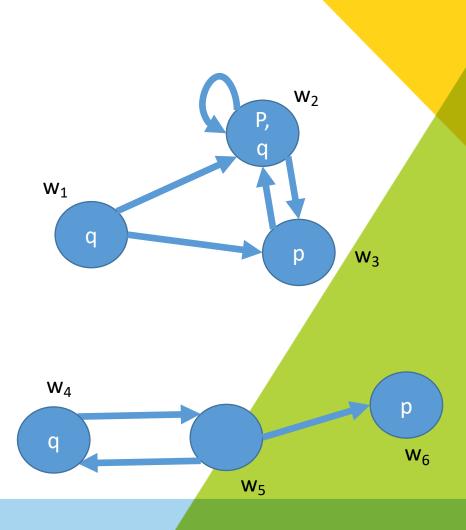
Is  $\Box p \rightarrow p$  valid?

If  $\Box$  represents belief, is  $\Box p \rightarrow p$ ?

(Recall: □p means "Agent Q believes p to be true")

If  $\Box$  represents **knowledge**, is  $\Box p \rightarrow p$ ?

(Recall:  $\Box p$  means "Agent Q knows p to be true")



## Sometimes we want to assert a formula schema

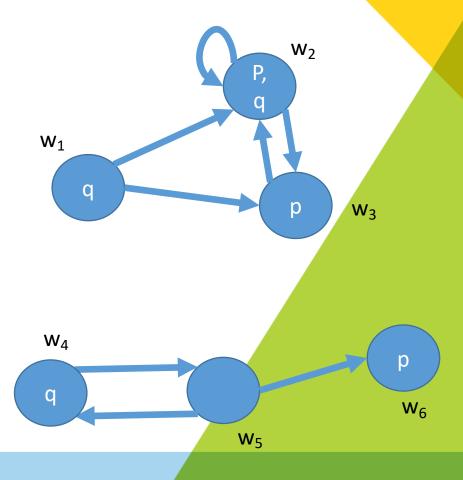
Formula schema: high-level pattern (we've seen this before)

e.g.,  $\varphi \lor \neg \varphi$  is a schema;

 $p \lor \neg p$  and  $(p \to q) \lor \neg (p \to q)$  are instances

We want  $\Box p \rightarrow p$  to be true when talking about knowledge, but not belief (even though it isn't valid generally).

When we assert a formula schema, we introduce it as an **axiom**.



Axiom	Knowledge	Belief	
$\Box p \to p$	Agent Q only knows true things	NOT A SUITABLE AXIOM	

Axiom	Knowledge	Belief	
$\Box p \to p$	Agent Q only knows true things	NOT A SUITABLE AXIOM	
$\Box \varphi \to \Box \Box \varphi$	Agent Q knows what it knows (introspection)	Agent Q believes what it believes	

Axiom	Knowledge	Belief	
$\Box p \to p$	Agent Q only knows true things	NOT A SUITABLE AXIOM	
$\Box \varphi \to \Box \Box \varphi$	Agent Q knows what it knows (introspection)	Agent Q believes what it believes	
$\circ \varphi \rightarrow \circ \Box \varphi$	If agent Q doesn't know something, it doesn't know what it doesn't know	If agent Q doesn't believe something, it doesn't believe what it doesn't believe	

Axiom	Knowledge	Belief	
$\Box p \to p$	Agent Q only knows true things	NOT A SUITABLE AXIOM	
$\Box \varphi \to \Box \Box \varphi$	Agent Q knows what it knows (introspection)	Agent Q believes what it believes	
$\circ \varphi \rightarrow \circ \Box \varphi$	If agent Q doesn't know something, it doesn't know what it doesn't know	If agent Q doesn't believe something, it doesn't believe what it doesn't believe	
¢Т	Agent Q doesn't know contradictions	Agent Q doesn't believe contradictions	

Axiom	Knowledge	Belief	
$\Box p \to p$	Agent Q only knows true things	NOT A SUITABLE AXIOM	
$\Box \varphi \to \Box \Box \varphi$	Agent Q knows what it knows (introspection)	Agent Q believes what it believes	
$\circ \varphi \rightarrow \Box \circ \varphi$	If agent Q doesn't know something, it doesn't know what it doesn't know	If agent Q doesn't believe something, it doesn't believe what it doesn't believe	
¢Τ	Agent Q doesn't know contradictions	Agent Q doesn't believe contradictions	
$\Box \varphi \to \diamond \varphi$	Agent Q can chain knowledge (true things)	Agent Q can chain belief (true things)	

Axiom	Knowledge	Belief	Property (axiom name)	R(w, w')
$\Box p \to p$	Agent Q only knows true things	NOT A SUITABLE AXIOM	Reflexive (T)	
$\Box \varphi \to \Box \Box \varphi$	Agent Q knows what it knows (introspection)	Agent Q believes what it believes		
$\circ \varphi \rightarrow \Box \circ \varphi$	If agent Q doesn't know something, it doesn't know what it doesn't know	If agent Q doesn't believe something, it doesn't believe what it doesn't believe		
<b>«</b> Т	Agent Q doesn't know contradictions	Agent Q doesn't believe contradictions		
$\Box \varphi \to \diamond \varphi$	Agent Q can chain knowledge (true things)	Agent Q can chain belief (true things)		

Axiom	Knowledge	Belief	Property (axiom name)	R(w, w')
$\Box p \to p$	Agent Q only knows true things	NOT A SUITABLE AXIOM	Reflexive (T)	
$\Box \varphi \to \Box \Box \varphi$	Agent Q knows what it knows (introspection)	Agent Q believes what it believes	Transitive (4)	$ \forall (w, w', w'') \in (W \times W \times W), (R(w, w') \land R(w', w'') \rightarrow R(w, w'')) $
$\circ \varphi \rightarrow \Box \circ \varphi$	If agent Q doesn't know something, it doesn't know what it doesn't know	If agent Q doesn't believe something, it doesn't believe what it doesn't believe		
¢Τ	Agent Q doesn't know contradictions	Agent Q doesn't believe contradictions		
$\Box \varphi \to \diamond \varphi$	Agent Q can chain knowledge (true things)	Agent Q can chain belief (true things)		

Axiom	Knowledge	Belief	Property (axiom name)	R(w, w')
$\Box p \to p$	Agent Q only knows true things	NOT A SUITABLE AXIOM	Reflexive (T)	
$\Box \varphi \to \Box \Box \varphi$	Agent Q knows what it knows (introspection)	Agent Q believes what it believes	Transitive (4)	$ \forall (w, w', w'') \in (W \times W \times W), (R(w, w') \land R(w', w'') \rightarrow R(w, w'')) $
$\circ \varphi \rightarrow \Box \circ \varphi$	If agent Q doesn't know something, it doesn't know what it doesn't know	If agent Q doesn't believe something, it doesn't believe what it doesn't believe	Euclidean (5)	$\forall (w, w', w'') \in (W \times W \times W),$ $(R(w, w') \wedge R(w, w'') \rightarrow R(w', w''))$
¢Τ	Agent Q doesn't know contradictions	Agent Q doesn't believe contradictions		
$\Box \varphi \to \diamond \varphi$	Agent Q can chain knowledge (true things)	Agent Q can chain belief (true things)		

Axiom	Knowledge	Belief	Property (axiom name)	R(w, w')
$\Box p \to p$	Agent Q only knows true things	NOT A SUITABLE AXIOM	Reflexive (T)	
$\Box \varphi \to \Box \Box \varphi$	Agent Q knows what it knows (introspection)	Agent Q believes what it believes	Transitive (4)	$ \forall (w, w', w'') \in (W \times W \times W), \\ (R(w, w') \land R(w', w'') \rightarrow R(w, w'')) $
$\circ \varphi \rightarrow \Box \circ \varphi$	If agent Q doesn't know something, it doesn't know what it doesn't know	If agent Q doesn't believe something, it doesn't believe what it doesn't believe	Euclidean (5)	$\forall (w, w', w'') \in (W \times W \times W),$ $(R(w, w') \wedge R(w, w'') \rightarrow R(w', w''))$
¢Τ	Agent Q doesn't know contradictions	Agent Q doesn't believe contradictions	Serial (D)	$\forall w \in W(\exists w' \in W(R(w, w'))$
$\Box \varphi \to \diamond \varphi$	Agent Q can chain knowledge (true things)	Agent Q can chain belief (true things)	Serial (D)	$\forall w \in W(\exists w' \in W(R(w, w'))$