An Invitation to Modal Logic: Lecture 2

Philosophy 150

Eric Pacuit

Stanford University

ai.stanford.edu/~epacuit

November 28, 2007

Plan

✓ Motivating Examples

11/28: Formalizing the muddy children puzzle, Basic Modal Logic I

11/30: Basic Modal Logic II

12/3: Basic Modal Logic III

12/5: Dynamics in Logic I

12/7: Dynamics in Logic II

Three children are outside playing. Two of them get mud on their forehead. They cannot see or feel the mud on their own foreheads, but can see who is dirty.

Their mother enters the room and says "At least one of you have mud on your forehead".

Then the children are repeatedly asked "do you know if you have mud on your forehead?"

What happens?

Claim: After first question, the children answer "I don't know",

Three children are outside playing. Two of them get mud on their forehead. They cannot see or feel the mud on their own foreheads, but can see who is dirty.

Their mother enters the room and says "At least one of you have mud on your forehead".

Then the children are repeatedly asked "do you know if you have mud on your forehead?"

What happens?

Claim: After first question, the children answer "I don't know", after the second question the muddy children answer "I have mud on my forehead!" (but the clean child is still in the dark).

Three children are outside playing. Two of them get mud on their forehead. They cannot see or feel the mud on their own foreheads, but can see who is dirty.

Their mother enters the room and says "At least one of you have mud on your forehead".

Then the children are repeatedly asked "do you know if you have mud on your forehead?"

What happens?

Claim: After first question, the children answer "I don't know", after the second question the muddy children answer "I have mud on my forehead!" (but the clean child is still in the dark). Then the clean child says, "Oh, I must be clean."

Muddy Children

Assume:

- There are three children: Ann, Bob and Charles.
- (Only) Ann and Bob have mud on their forehead.

Muddy Children

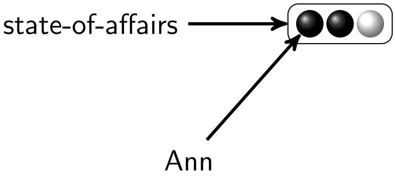
Assume:

- There are three children: Ann, Bob and Charles.
- (Only) Ann and Bob have mud on their forehead.

Muddy Children

Assume:

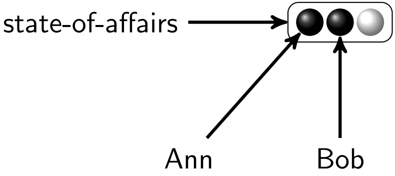
- There are three children: Ann, Bob and Charles.
- (Only) Ann and Bob have mud on their forehead.



Muddy Children

Assume:

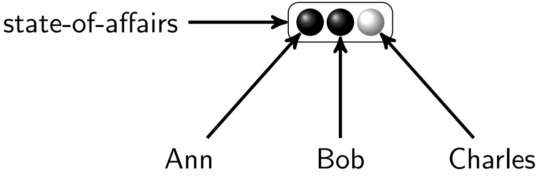
- There are three children: Ann, Bob and Charles.
- (Only) Ann and Bob have mud on their forehead.



Muddy Children

Assume:

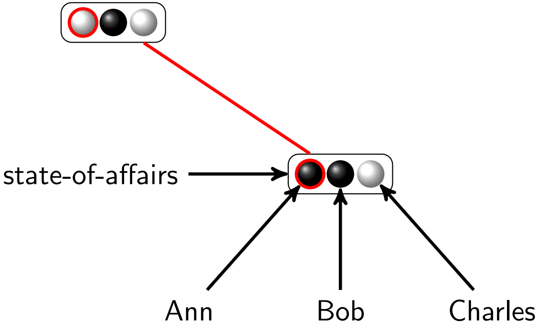
- There are three children: Ann, Bob and Charles.
- (Only) Ann and Bob have mud on their forehead.



Muddy Children

Assume:

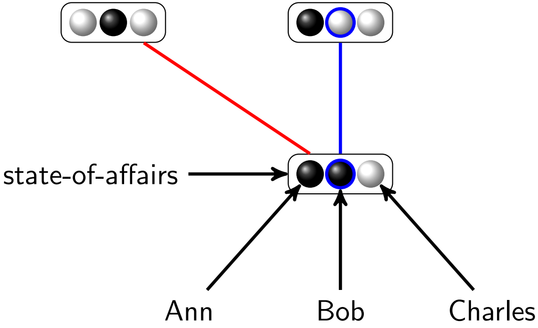
- There are three children: Ann, Bob and Charles.
- (Only) Ann and Bob have mud on their forehead.



Muddy Children

Assume:

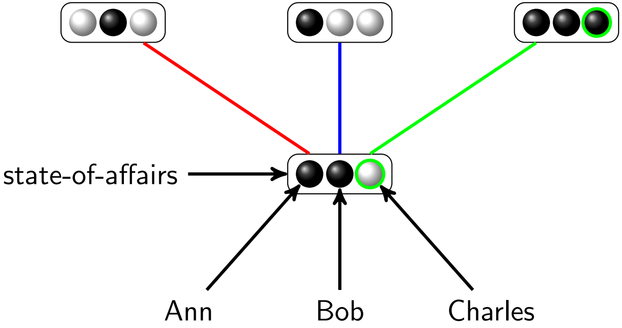
- There are three children: Ann, Bob and Charles.
- (Only) Ann and Bob have mud on their forehead.



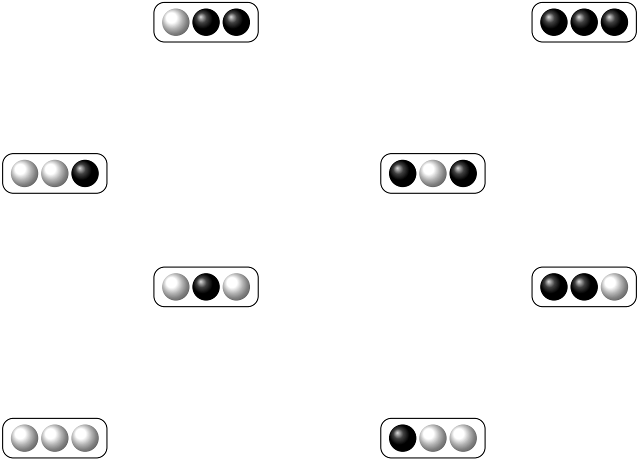
Muddy Children

Assume:

- There are three children: Ann, Bob and Charles.
- (Only) Ann and Bob have mud on their forehead.

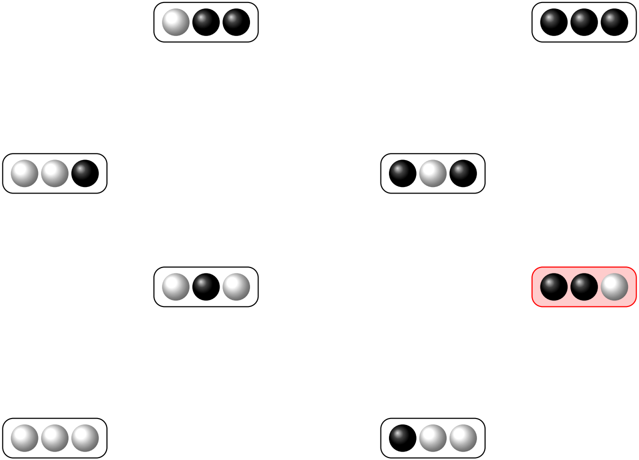


Muddy Children



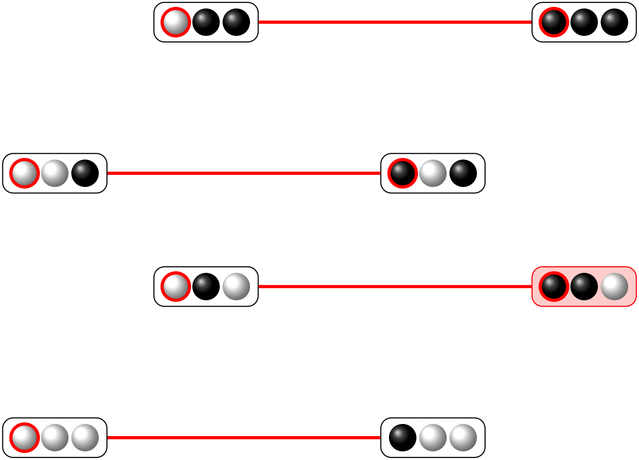
All 8 possible situations

Muddy Children



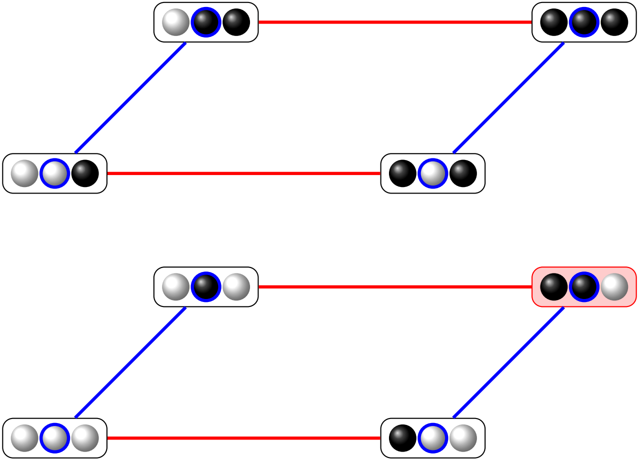
The actual situation

Muddy Children



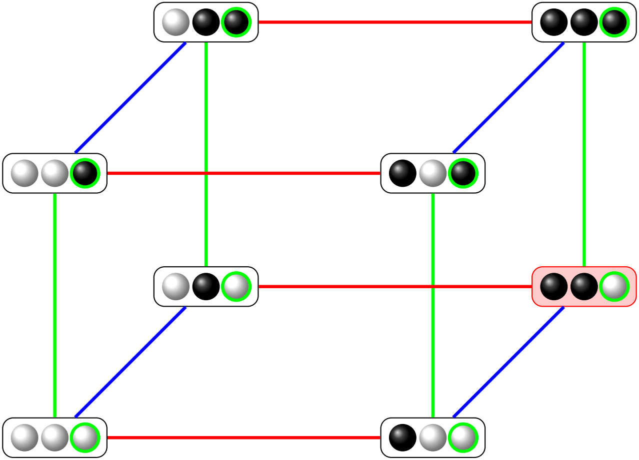
Ann's uncertainty

Muddy Children



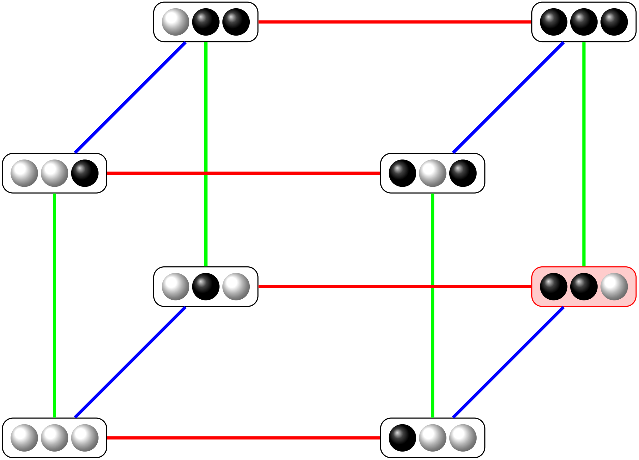
Bob's uncertainty

Muddy Children

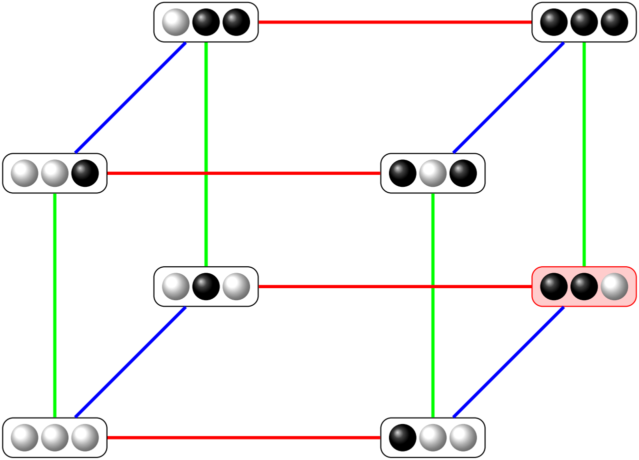


Charles' uncertainty

Muddy Children

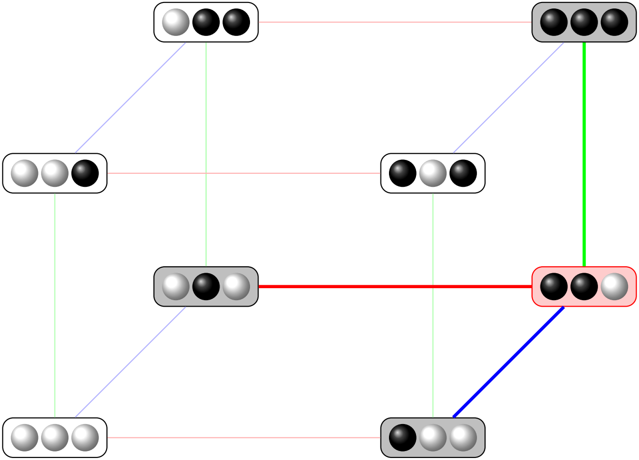


Muddy Children



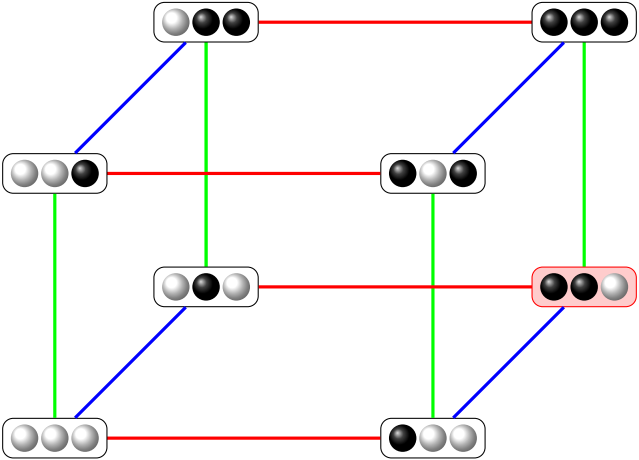
None of the children know if they are muddy

Muddy Children



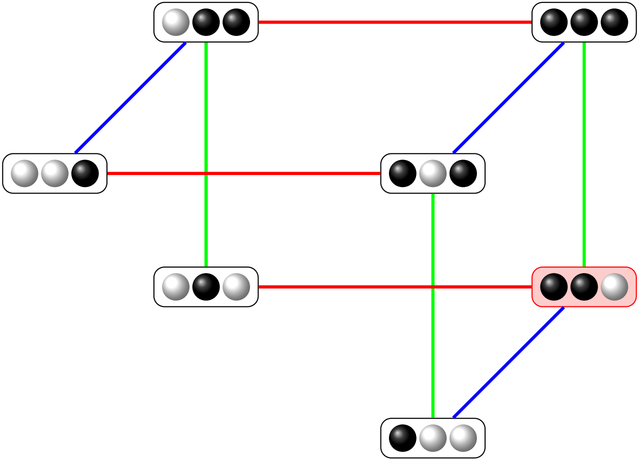
None of the children know if they are muddy

Muddy Children



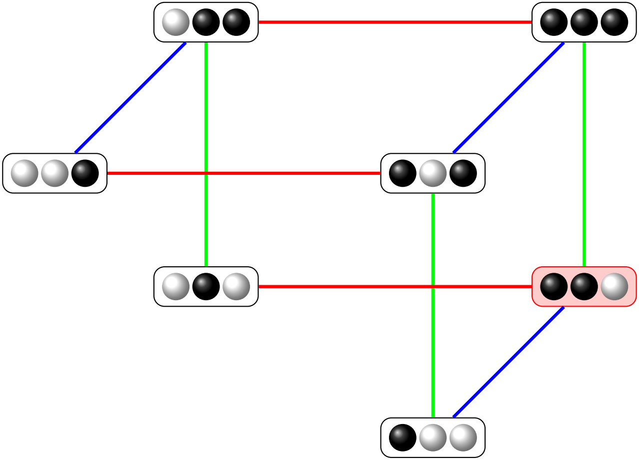
"At least one has mud on their forehead."

Muddy Children



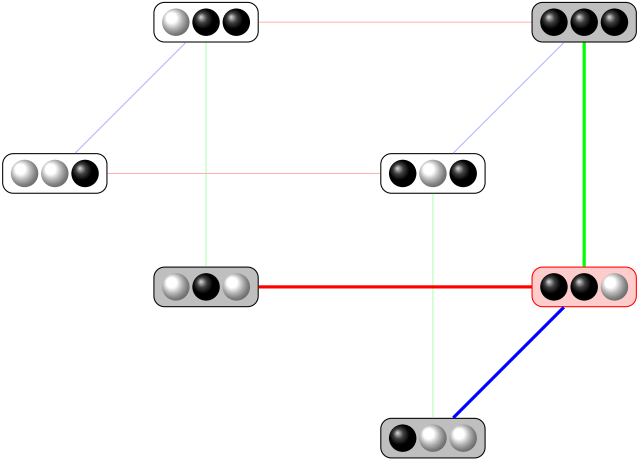
"At least one has mud on their forehead."

Muddy Children



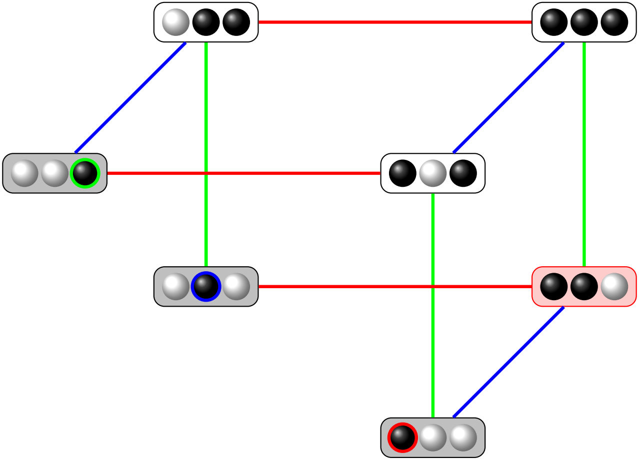
"Who has mud on their forehead?"

Muddy Children



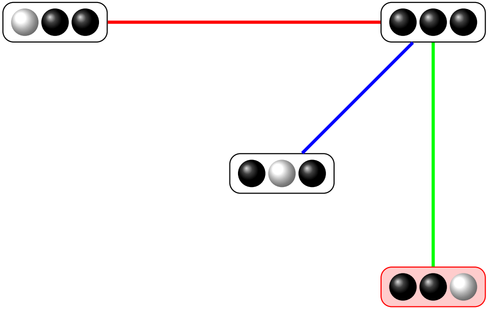
"Who has mud on their forehead?"

Muddy Children



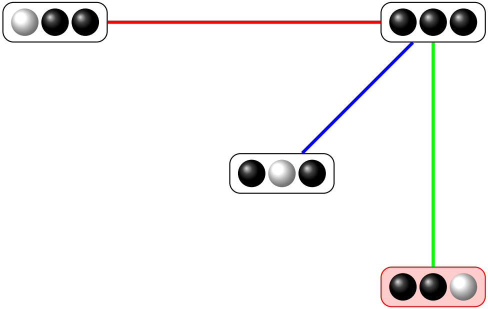
No one steps forward.

Muddy Children



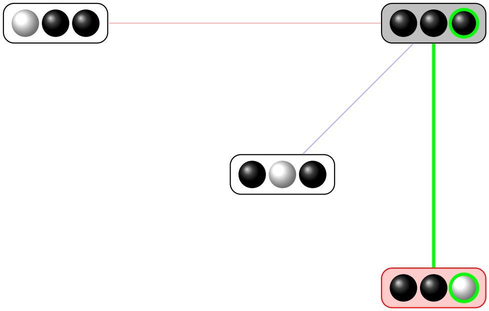
No one steps forward.

Muddy Children



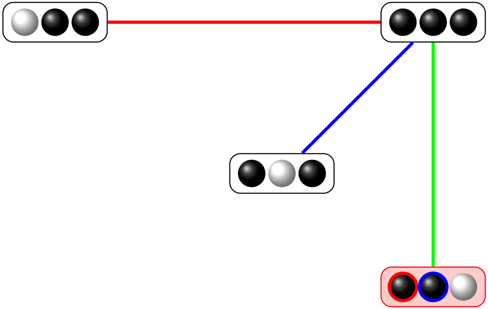
“Who has mud on their forehead?”

Muddy Children



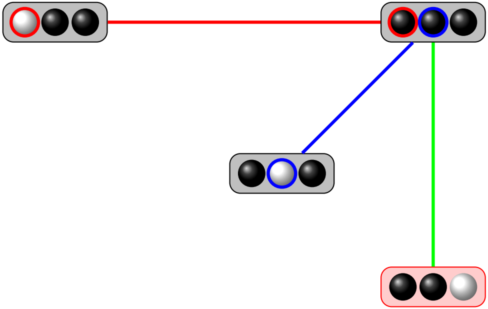
Charles does not know he is clean.

Muddy Children



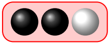
Ann and Bob step forward.

Muddy Children



Now, Charles knows he is clean.

Muddy Children



Now, Charles knows he is clean.

Recall:

A wff of **Propositional Logic** is defined *inductively*:

- Any atomic propositional variable is a wff
- If P and Q are wff, then so are $\neg P$, $P \wedge Q$, $P \vee Q$ and $P \rightarrow Q$

A wff of **Modal Logic** is defined *inductively*:

1. Any atomic propositional variable is a wff
2. If P and Q are wff, then so are $\neg P$, $P \wedge Q$, $P \vee Q$ and $P \rightarrow Q$
3. If P is a wff, then so is $\Box P$ and $\Diamond P$

One formal language, many interpretations

Alethic
$\Box P$ is intended to mean P is **necessary**

One formal language, many interpretations

Alethic
$\Box P$ is intended to mean P is **necessary**

Deontic
$\Box P$ is intended to mean P is **obligatory**

One formal language, many interpretations

Alethic
$\Box P$ is intended to mean P is **necessary**

Deontic
$\Box P$ is intended to mean P is **obligatory**

Epistemic
$\Box P$ is intended to mean P is **known**

One formal language, many interpretations

Alethic
$\Box P$ is intended to mean P is **necessary**

Deontic
$\Box P$ is intended to mean P is **obligatory**

Epistemic
$\Box P$ is intended to mean P is **known**

Doxastic
$\Box P$ is intended to mean P is **believed**

One formal language, many interpretations

Alethic
$\Box P$ is intended to mean P is **necessary**

Deontic
$\Box P$ is intended to mean P is **obligatory**

Epistemic
$\Box P$ is intended to mean P is **known**

Doxastic
$\Box P$ is intended to mean P is **believed**

Temporal
$\Box P$ is intended to mean P will **always** be true (at every point in the future)

Some Questions:

Some Questions:

1. Is $(A \rightarrow B) \vee (B \rightarrow A)$ **true** or **false**?

Some Questions:

1. Is $(A \rightarrow B) \vee (B \rightarrow A)$ **true** or **false**? true.

Some Questions:

1. Is $(A \rightarrow B) \vee (B \rightarrow A)$ **true** or **false**? true.
2. Is $A \wedge \neg(B \vee A)$ **true** or **false**?

Some Questions:

1. Is $(A \rightarrow B) \vee (B \rightarrow A)$ **true** or **false**? true.
2. Is $A \wedge \neg(B \vee A)$ **true** or **false**? false.

Some Questions:

1. Is $(A \rightarrow B) \vee (B \rightarrow A)$ **true** or **false**? true.
2. Is $A \wedge \neg(B \vee A)$ **true** or **false**? false.
3. Is $A \rightarrow (B \vee C)$ **true** or **false**?

Some Questions:

1. Is $(A \rightarrow B) \vee (B \rightarrow A)$ **true** or **false**? true.
2. Is $A \wedge \neg(B \vee A)$ **true** or **false**? false.
3. Is $A \rightarrow (B \vee C)$ **true** or **false**? It depends!

Some Questions:

1. Is $(A \to B) \vee (B \to A)$ **true** or **false**? true.
2. Is $A \wedge \neg(B \vee A)$ **true** or **false**? false.
3. Is $A \to (B \vee C)$ **true** or **false**? It depends!
4. Is $\Box A \to (B \to \Box A)$ **true** or **false**?

Some Questions:

1. Is $(A \rightarrow B) \vee (B \rightarrow A)$ **true** or **false**? true.
2. Is $A \wedge \neg(B \vee A)$ **true** or **false**? false.
3. Is $A \rightarrow (B \vee C)$ **true** or **false**? It depends!
4. Is $\Box A \rightarrow (B \rightarrow \Box A)$ **true** or **false**? true.

Some Questions:

1. Is $(A \rightarrow B) \vee (B \rightarrow A)$ **true** or **false**? true.
2. Is $A \wedge \neg(B \vee A)$ **true** or **false**? false.
3. Is $A \rightarrow (B \vee C)$ **true** or **false**? It depends!
4. Is $\Box A \rightarrow (B \rightarrow \Box A)$ **true** or **false**? true.
5. Is $\neg\Box A \wedge \neg(\Diamond B \vee \neg\Box A)$ **true** or **false**?

Some Questions:

1. Is $(A \rightarrow B) \vee (B \rightarrow A)$ **true** or **false**? true.
2. Is $A \wedge \neg(B \vee A)$ **true** or **false**? false.
3. Is $A \rightarrow (B \vee C)$ **true** or **false**? It depends!
4. Is $\Box A \rightarrow (B \rightarrow \Box A)$ **true** or **false**? true.
5. Is $\neg\Box A \wedge \neg(\Diamond B \vee \neg\Box A)$ **true** or **false**? false.

Some Questions:

1. Is $(A \to B) \vee (B \to A)$ **true** or **false**? true.
2. Is $A \wedge \neg(B \vee A)$ **true** or **false**? false.
3. Is $A \to (B \vee C)$ **true** or **false**? It depends!
4. Is $\Box A \to (B \to \Box A)$ **true** or **false**? true.
5. Is $\neg\Box A \wedge \neg(\Diamond B \vee \neg\Box A)$ **true** or **false**? false.
6. Is $\neg\Box A \wedge \neg(\Diamond B \vee \Diamond\neg A)$ **true** or **false**?

Some Questions:

1. Is $(A \to B) \vee (B \to A)$ **true** or **false**? true.
2. Is $A \wedge \neg(B \vee A)$ **true** or **false**? false.
3. Is $A \to (B \vee C)$ **true** or **false**? It depends!
4. Is $\Box A \to (B \to \Box A)$ **true** or **false**? true.
5. Is $\neg\Box A \wedge \neg(\Diamond B \vee \neg\Box A)$ **true** or **false**? false.
6. Is $\neg\Box A \wedge \neg(\Diamond B \vee \Diamond\neg A)$ **true** or **false**? false.
   (**tricky:** $\Box A$ **is equivalent to** $\neg\Diamond\neg A$**.**)

Some Questions:

1. Is $(A \rightarrow B) \vee (B \rightarrow A)$ **true** or **false**? true.
2. Is $A \wedge \neg(B \vee A)$ **true** or **false**? false.
3. Is $A \rightarrow (B \vee C)$ **true** or **false**? It depends!
4. Is $\Box A \rightarrow (B \rightarrow \Box A)$ **true** or **false**? true.
5. Is $\neg\Box A \wedge \neg(\Diamond B \vee \neg\Box A)$ **true** or **false**? false.
6. Is $\neg\Box A \wedge \neg(\Diamond B \vee \Diamond\neg A)$ **true** or **false**? false.
   (**tricky: $\Box A$ is equivalent to $\neg\Diamond\neg A$.**)
7. Is $\Box A \rightarrow A$ **true** or **false**?

Some Questions:

1. Is $(A \rightarrow B) \vee (B \rightarrow A)$ **true** or **false**? true.
2. Is $A \wedge \neg(B \vee A)$ **true** or **false**? false.
3. Is $A \rightarrow (B \vee C)$ **true** or **false**? It depends!
4. Is $\Box A \rightarrow (B \rightarrow \Box A)$ **true** or **false**? true.
5. Is $\neg\Box A \wedge \neg(\Diamond B \vee \neg\Box A)$ **true** or **false**? false.
6. Is $\neg\Box A \wedge \neg(\Diamond B \vee \Diamond\neg A)$ **true** or **false**? false.
   (**tricky:** $\Box A$ **is equivalent to** $\neg\Diamond\neg A$**.**)
7. Is $\Box A \rightarrow A$ **true** or **false**? It depends!

Can we give find a natural *semantics* for the basic modal language?

Can we give find a natural *semantics* for the basic modal language?

What about truth tables?

Can we give find a natural *semantics* for the basic modal language?

What about truth tables? Won't work ! (Why?)

Proof (from the board): There are four possible truth tables:

P	$\Box P$		P	$\Box P$		P	$\Box P$		P	$\Box P$	
T	T	$\mathcal{T}_1$	T	F	$\mathcal{T}_2$	T	F	$\mathcal{T}_3$	T	T	$\mathcal{T}_4$
F	T		F	F		F	T		F	F	

Suppose we want $\Box P \rightarrow P$ to be *valid* (i.e., true regardless of the interpretation of P), but allow for the possibility that both $\neg\Box P$ and $P \rightarrow \Box P$ are false. (This is natural on an epistemic reading: it is a principle that knowledge of P entails the truth of P. Further it is possible that P is known ($\neg\Box P$ is false), and it is false that *if P is true then P is known* ($P \rightarrow \Box P$ is false).)

Assuming $\Box P \rightarrow P$ is true under *all* interpretations means we have to rule out all truth tables that contain a row with $\Box P$ assigned T but P assigned F. Hence, we throw out $\mathcal{T}_1$ and $\mathcal{T}_3$.

Now in order to make $P \rightarrow \Box P$ false, there must at least one row in which P is assigned T, but $\Box P$ is assigned F. Hence we throw out $\mathcal{T}_4$.

This leaves us with truth table $\mathcal{T}_2$, but here $\neg\Box P$ is always true (i.e., $\Box P$ is always assigned F). Q.E.D.

Can we give find a natural *semantics* for the basic modal language?

What about truth tables? Won't work ! (Why?)

The solution was provided by the American philosopher Saul Kripke (see also the work of Hintikka, McKinsey and Tarski, and others).

The main idea:

- ‘Currently, it is sunny outside.’ is true

The main idea:

- 'Currently, it is sunny outside.' is true, but it is not necessary (for example, if we were in Amsterdam).

The main idea:

- 'Currently, it is sunny outside.' is true, but it is not necessary (for example, if we were in Amsterdam).

- We say P is **necessary** provided P is true in all (relevant) situations (states, worlds, possibilities).

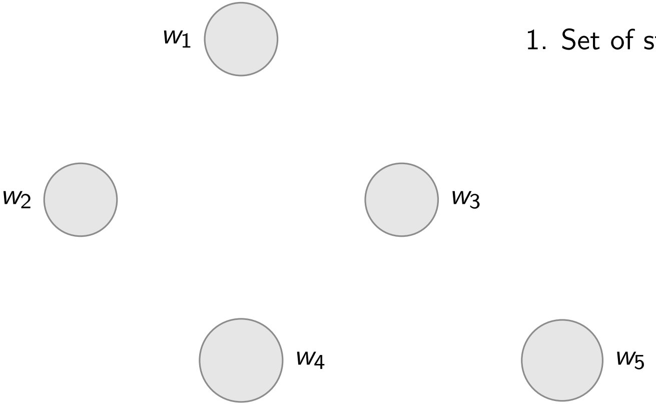
The main idea:

- 'Currently, it is sunny outside.' is true, but it is not necessary (for example, if we were in Amsterdam).

- We say P is **necessary** provided P is true in all (relevant) situations (states, worlds, possibilities).

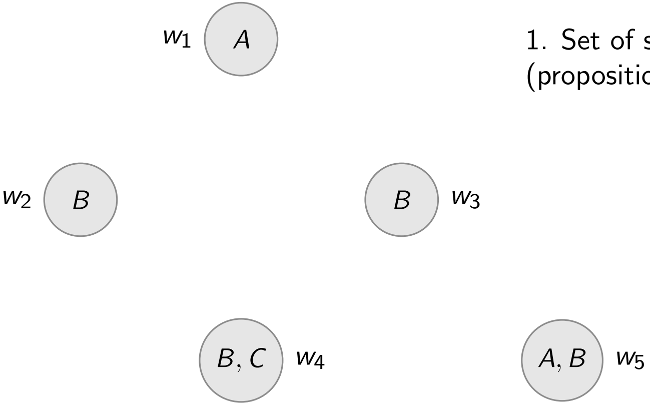
- A **Kripke structure** is
  1. A set of states, or worlds (each world specifies the truth value of all propositional variables)
  2. A **relation** on the set of states (specifying the "relevant situations")

A Kripke Structure



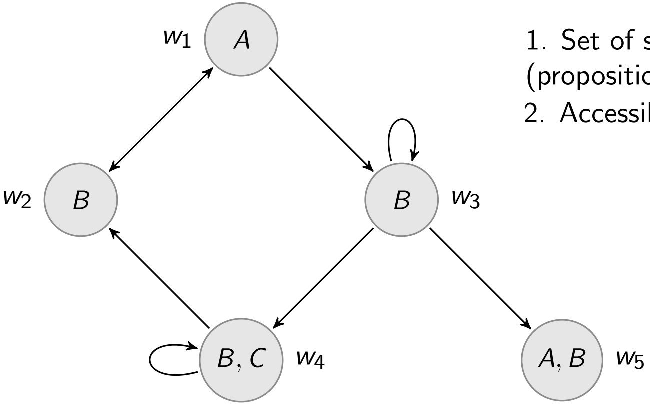
1. Set of states

A Kripke Structure



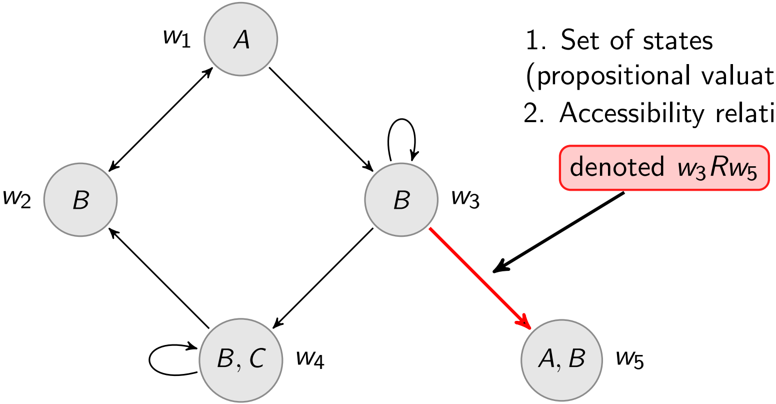
1. Set of states (propositional valuations)

A Kripke Structure



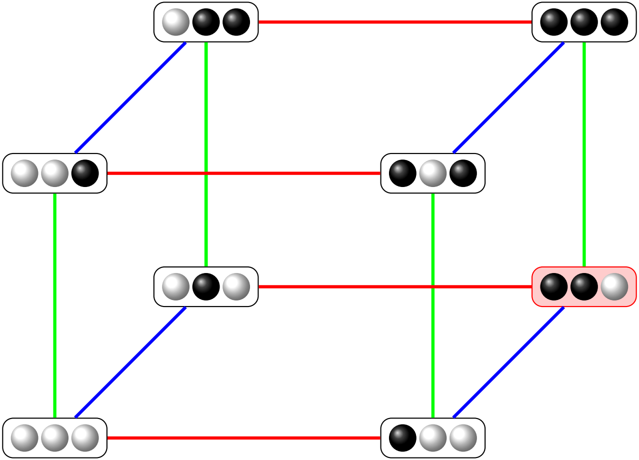
1. Set of states (propositional valuations)
2. Accessibility relation

A Kripke Structure



1. Set of states (propositional valuations)
2. Accessibility relation

A More Concrete Example of a Kripke Structure



Truth of Modal Formulas

We interpret formulas at states in a Kripke structure: $w \models P$ means P is true at state w.

Truth of Modal Formulas

We interpret formulas at states in a Kripke structure: $w \models P$ means P is true at state w.

We write wRv is v is accessible from state w.

Truth of Modal Formulas

We interpret formulas at states in a Kripke structure: $w \models P$ means P is true at state w.

We write wRv is v is accessible from state w.

1. $\Box P$ is true at state w iff P is true in all accessible worlds.

Truth of Modal Formulas

We interpret formulas at states in a Kripke structure: $w \models P$ means P is true at state w.

We write wRv is v is accessible from state w.

1. $\Box P$ is true at state w iff P is true in all accessible worlds.
   $w \models \Box P$ iff for all v, if wRv then $v \models P$

Truth of Modal Formulas

We interpret formulas at states in a Kripke structure: $w \models P$ means P is true at state w.

We write wRv is v is accessible from state w.

1. $\Box P$ is true at state w iff P is true in all accessible worlds.
   $w \models \Box P$ iff for all v, if wRv then $v \models P$
2. $\Diamond P$ is true at state w iff P is true at some accessible world.

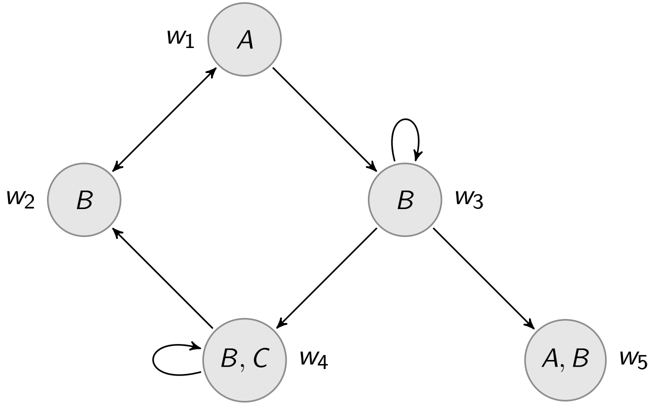
Truth of Modal Formulas

We interpret formulas at states in a Kripke structure: $w \models P$ means P is true at state w.

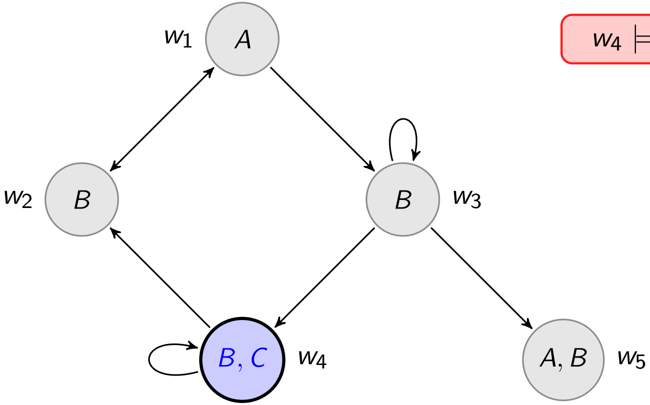
We write wRv is v is accessible from state w.

1. $\Box P$ is true at state w iff P is true in all accessible worlds.
   $w \models \Box P$ iff for all v, if wRv then $v \models P$
2. $\Diamond P$ is true at state w iff P is true at some accessible world.
   $w \models \Diamond P$ iff there exists v such that wRv and $v \models P$.

Example

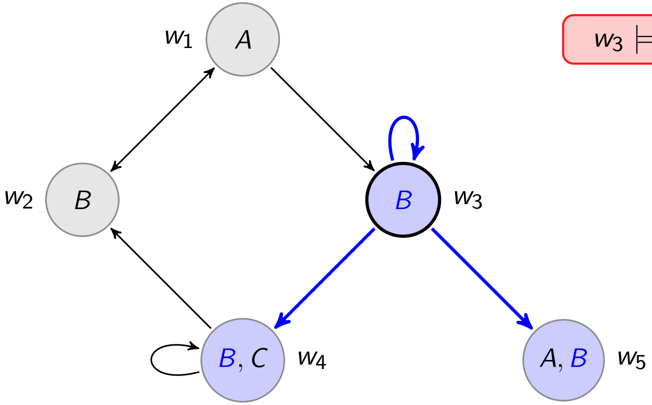


Example



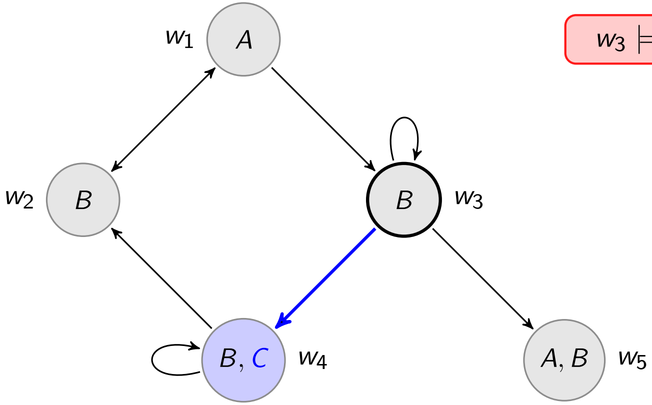
$w_4 \models B \wedge C$

Example



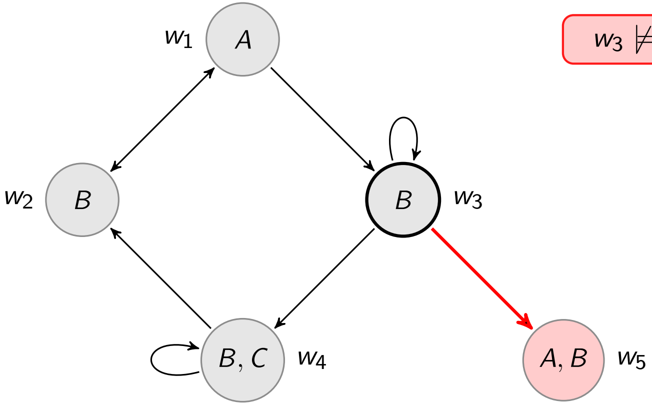
$w_3 \models \Box B$

Example



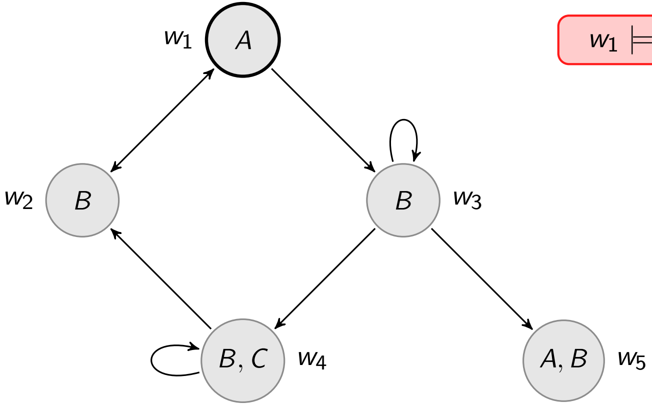
$w_3 \models \Diamond C$

Example



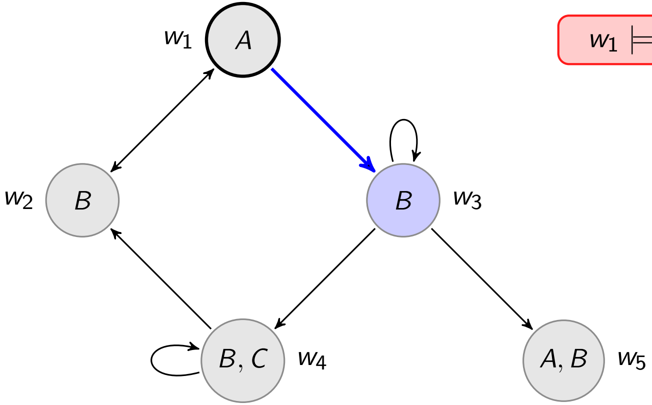
$w_3 \not\models \Box C$

Example



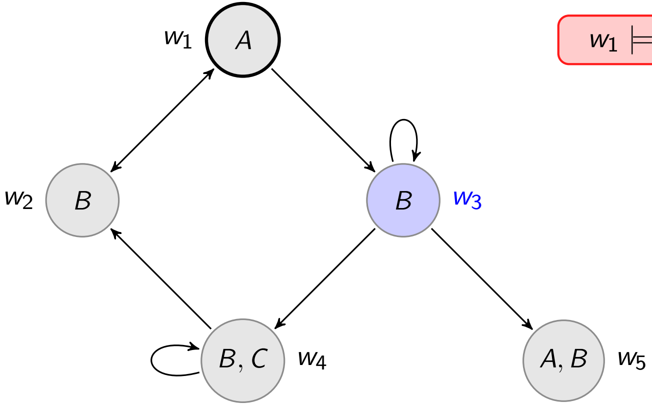
$w_1 \models \Diamond\Box B$

Example



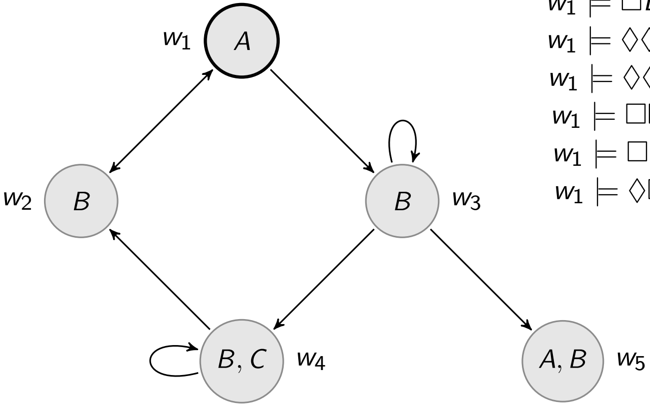
$w_1 \models \Diamond\Box B$

Example



$w_1 \models \Diamond\Box B$

Example



$w_1 \models \Box B \wedge B$?

$w_1 \models \Diamond\Diamond B$?

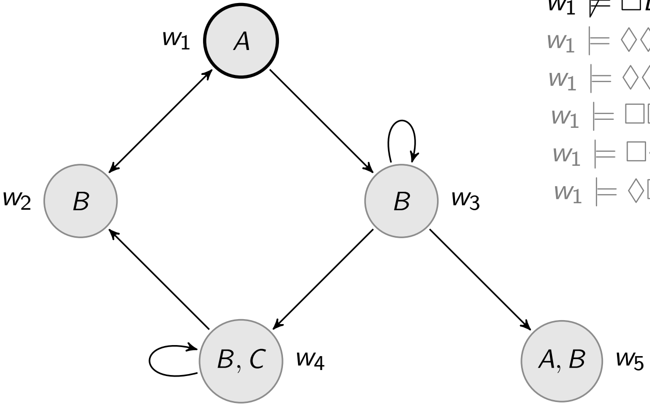
$w_1 \models \Diamond\Diamond\Diamond B$?

$w_1 \models \Box\Box B$?

$w_1 \models \Box\Diamond C$?

$w_1 \models \Diamond\Box A$?

Example



$w_1 \not\models \Box B \wedge B$

$w_1 \models \Diamond\Diamond B$?

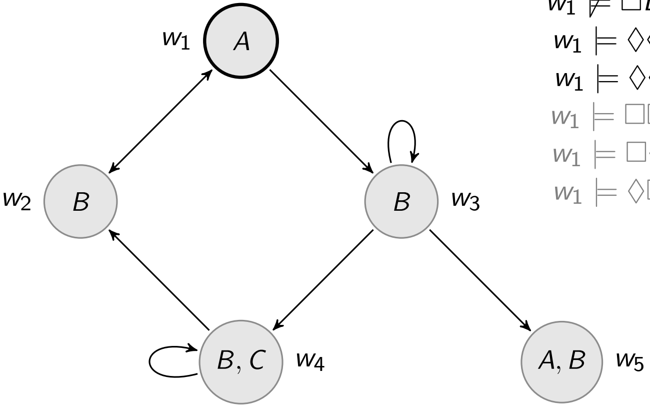
$w_1 \models \Diamond\Diamond\Diamond B$?

$w_1 \models \Box\Box B$?

$w_1 \models \Box\Diamond C$?

$w_1 \models \Diamond\Box A$?

Example



$w_1 \not\models \Box B \wedge B$

$w_1 \models \Diamond\Diamond B$

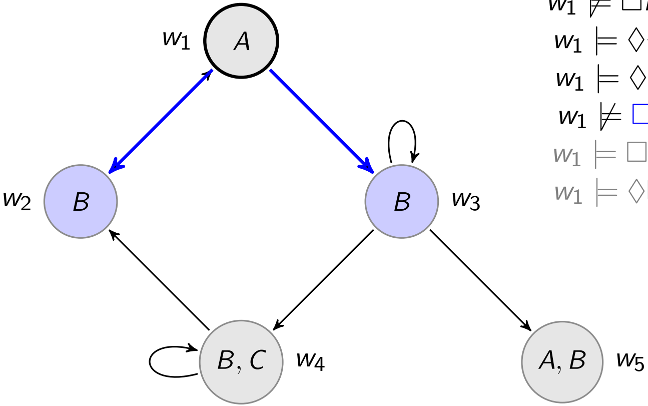
$w_1 \models \Diamond\Diamond\Diamond B$

$w_1 \models \Box\Box B$?

$w_1 \models \Box\Diamond C$?

$w_1 \models \Diamond\Box A$?

Example



$w_1 \not\models \Box B \wedge B$

$w_1 \models \Diamond\Diamond B$

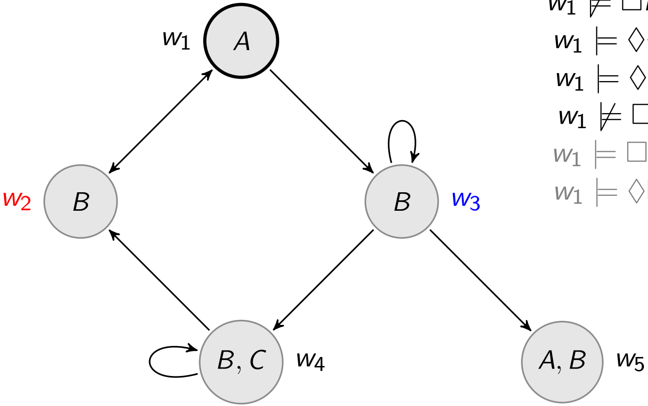
$w_1 \models \Diamond\Diamond\Diamond B$

$w_1 \not\models \Box\Box B$

$w_1 \models \Box\Diamond C?$

$w_1 \models \Diamond\Box A?$

Example



$w_1 \not\models \Box B \wedge B$

$w_1 \models \Diamond\Diamond B$

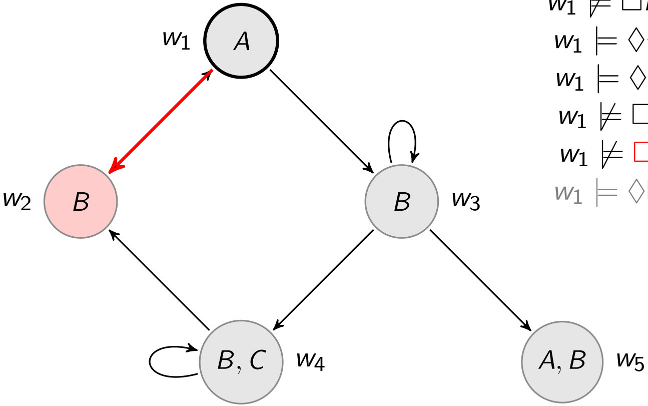
$w_1 \models \Diamond\Diamond\Diamond B$

$w_1 \not\models \Box\Box B$

$w_1 \models \Box\Diamond C$?

$w_1 \models \Diamond\Box A$?

Example



$w_1 \not\models \Box B \wedge B$

$w_1 \models \Diamond\Diamond B$

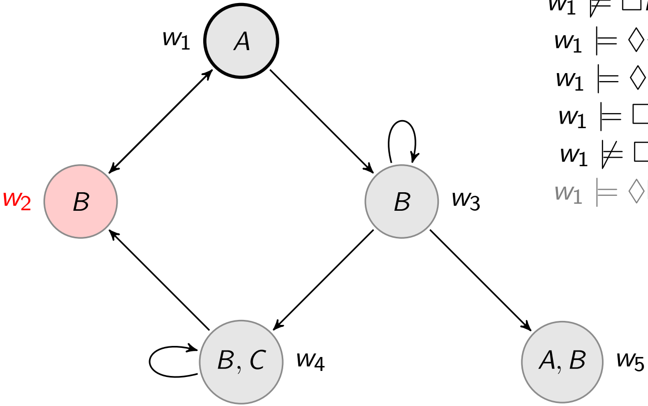
$w_1 \models \Diamond\Diamond\Diamond B$

$w_1 \not\models \Box\Box B$

$w_1 \not\models \Box\Diamond C$

$w_1 \models \Diamond\Box A$?

Example



$w_1 \not\models \Box B \wedge B$

$w_1 \models \Diamond\Diamond B$

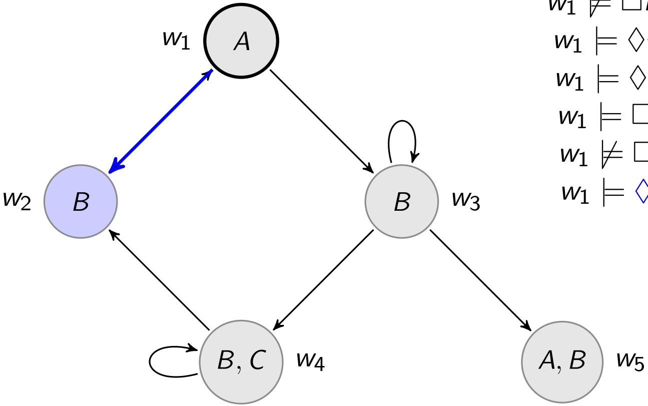
$w_1 \models \Diamond\Diamond\Diamond B$

$w_1 \models \Box\Box B$

$w_1 \not\models \Box\Diamond C$

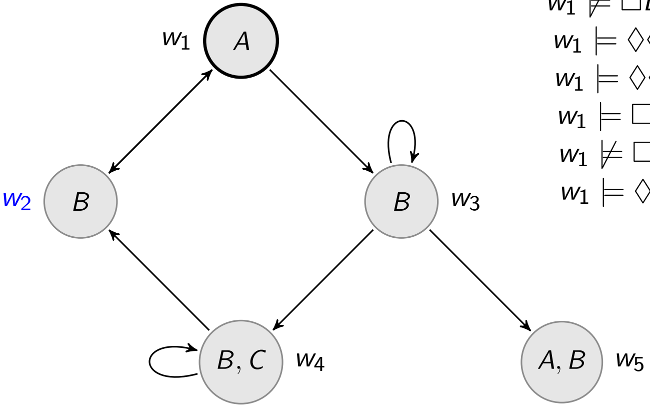
$w_1 \models \Diamond\Box A$?

Example



$w_1 \not\models \Box B \wedge B$
$w_1 \models \Diamond\Diamond B$
$w_1 \models \Diamond\Diamond\Diamond B$
$w_1 \models \Box\Box B$
$w_1 \not\models \Box\Diamond C$
$w_1 \models \Diamond\Box A$

Example



$w_1 \not\models \Box B \wedge B$

$w_1 \models \Diamond\Diamond B$

$w_1 \models \Diamond\Diamond\Diamond B$

$w_1 \models \Box\Box B$

$w_1 \not\models \Box\Diamond C$

$w_1 \models \Diamond\Box A$

Some Facts

- $\Box P \vee \neg\Box P$ is always true (i.e., true at any state in any Kripke structure), but what about $\Box P \vee \Box\neg P$?

Some Facts

- $\Box P \vee \neg\Box P$ is always true (i.e., true at any state in any Kripke structure), but what about $\Box P \vee \Box\neg P$?

- $\Box P \wedge \Box Q \rightarrow \Box(P \wedge Q)$ is true at any state in any Kripke structure.

Some Facts

- $\Box P \vee \neg\Box P$ is always true (i.e., true at any state in any Kripke structure), but what about $\Box P \vee \Box\neg P$?

- $\Box P \wedge \Box Q \rightarrow \Box(P \wedge Q)$ is true at any state in any Kripke structure. What about $\Box(P \vee Q) \rightarrow \Box(P \vee \Box Q)$?

Some Facts

- $\Box P \vee \neg\Box P$ is always true (i.e., true at any state in any Kripke structure), but what about $\Box P \vee \Box\neg P$?
- $\Box P \wedge \Box Q \rightarrow \Box(P \wedge Q)$ is true at any state in any Kripke structure. What about $\Box(P \vee Q) \rightarrow \Box(P \vee \Box Q)$?
- $\Box P \leftrightarrow \neg\Diamond\neg P$ is true at any state in any Kripke structure.

Next time: continue our discussion of modal logic.

Homework: available on the course website.

Questions?
Email: epacuit@stanford.edu
Website: ai.stanford.edu/~epacuit
Office: Gates 258