COMPUTATIONAL LEARNING IN DYNAMIC LOGICS

DAY 1: INTRODUCTION TO LEARNING AND EPISTEMIC LOGIC

Nina Gierasimczuk and Caleb Schultz Kisby

@NASSLLI, June 2025

Course Homepage:

https://sites.google.com/view/nasslli25-learning-in-del

THE STRUCTURE OF THIS COURSE

Lecture 1. Introduction to Learning and Epistemic Logic

Lecture 2. Dynamic Epistemic Logic and Belief Revision

Lecture 3. Dynamic Logic over Neural Networks

Lecture 4. Iterated Updates and Learnability

Lecture 5. Current Topics on Learnability

PLAN FOR TODAY

1. Inductive Inference: The Eleusis Game

2. Learning Paradigms and Perspectives

3. Introduction to Epistemic Logic

PLAN FOR TODAY

1. Inductive Inference: The Eleusis Game
2. Learning Paradigms and Perspectives
3. Introduction to Epistemic Logic

INDUCTIVE INFERENCE: CARD GAME

What is the rule behind this sequence of cards?

A♠ Q♠ 3♠ A♠ Q♠ 4♡ . . .

INDUCTIVE INFERENCE: CARD GAME

What is the rule behind this sequence of cards?

A♠ Q♠ 3♠ A♠ Q♠ 4♡ . . .

INDUCTIVE INFERENCE: CARD GAME

What is the rule behind this sequence of cards?

A♠ Q♠ 3♠ A♠ Q♠ 4♡ . . .

INDUCTIVE INFERENCE: CARD GAME

What is the rule behind this sequence of cards?

A♠ Q♠ 3♠ A♠ Q♠ 4♡ . . .

INDUCTIVE INFERENCE: CARD GAME

What is the rule behind this sequence of cards?

A♠ Q♠ 3♠ A♠ Q♠ 4♡ . . .

INDUCTIVE INFERENCE: CARD GAME

What is the rule behind this sequence of cards?

A♠ Q♠ 3♠ A♠ Q♠ 4♡ . . .

INDUCTIVE INFERENCE: CARD GAME

What is the rule behind this sequence of cards?

A♠ Q♠ 3♠ A♠ Q♠ 4♡ . . .

INDUCTIVE INFERENCE: CARD GAME

What is the rule behind this sequence of cards?

A♠ Q♠ 3♠ A♠ Q♠ 4♡ . . .

HOW MANY DIFFERENT ABSTRACT SCENARIOS?

Assume we have at our disposal unlimited amount of playing cards.

1. How many different (kinds of) playing cards do we have?
2. How many different beginnings of length 1?
3. How many different beginnings of length 2?
4. How many different infinite sequences?

HOW MANY DIFFERENT ABSTRACT SCENARIOS?

Assume we have at our disposal unlimited amount of playing cards.

1. How many different (kinds of) playing cards do we have?
2. How many different beginnings of length 1?
3. How many different beginnings of length 2?
4. How many different infinite sequences?

HOW MANY DIFFERENT ABSTRACT SCENARIOS?

Assume we have at our disposal unlimited amount of playing cards.

1. How many different (kinds of) playing cards do we have?
2. How many different beginnings of length 1?
3. How many different beginnings of length 2?
4. How many different infinite sequences?

HOW MANY DIFFERENT ABSTRACT SCENARIOS?

Assume we have at our disposal unlimited amount of playing cards.

1. How many different (kinds of) playing cards do we have?
2. How many different beginnings of length 1?
3. How many different beginnings of length 2?
4. How many different infinite sequences?

HOW MANY DIFFERENT ABSTRACT SCENARIOS?

Assume we have at our disposal unlimited amount of playing cards.

1. How many different (kinds of) playing cards do we have?
2. How many different beginnings of length 1?
3. How many different beginnings of length 2?
4. How many different infinite sequences?

THE INFINITIES OF INDUCTIVE INFERENCE

1.	A♠	A♠	A♠	A♠	A♠	A♠	...
2.	A♣	A♣	A♣	A♣	A♣	A♣	...
3.	A♡	A♡	A♡	A♡	A♡	A♡	...
4.	A♢	Q♠	3♠	8♡	2♡	5♠	...
5.	A♠	Q♠	7♠	J♠	5♠	5♠	...
...							
m.	A♣	A♡	A♣	A♡	A♣	A♢	...
...							

THE INFINITIES OF INDUCTIVE INFERENCE

1.	A♠	A♠	A♠	A♠	A♠	A♠	...
2.	A♣	A♣	A♣	A♣	A♣	A♣	...
3.	A♡	A♡	A♡	A♡	A♡	A♡	...
4.	A♢	Q♠	3♠	8♡	2♡	5♠	...
5.	A♠	Q♠	7♠	J♠	5♠	5♠	...
...							
m.	A♣	A♡	A♣	A♡	A♣	A♢	...
...							

THE INFINITIES OF INDUCTIVE INFERENCE

1.	A♠	A♠	A♠	A♠	A♠	A♠	...
2.	A♣	A♣	A♣	A♣	A♣	A♣	...
3.	A♡	A♡	A♡	A♡	A♡	A♡	...
4.	A♢	Q♠	3♠	8♡	2♡	5♠	...
5.	A♠	Q♠	7♠	J♠	5♠	5♠	...
...							
m.	A♣	A♡	A♣	A♡	A♣	A♢	...
...							

THE INFINITIES OF INDUCTIVE INFERENCE

1.	A♠	A♠	A♠	A♠	A♠	A♠	...
2.	A♣	A♣	A♣	A♣	A♣	A♣	...
3.	A♡	A♡	A♡	A♡	A♡	A♡	...
4.	A♢	Q♠	3♠	8♡	2♡	5♠	...
5.	A♠	Q♠	7♠	J♠	5♠	5♠	...
...							
m.	A♣	A♡	A♣	A♡	A♣	A♢	...
...							

THE INFINITIES OF INDUCTIVE INFERENCE

1.	A♠	A♠	A♠	A♠	A♠	A♠	...
2.	A♣	A♣	A♣	A♣	A♣	A♣	...
3.	A♡	A♡	A♡	A♡	A♡	A♡	...
4.	A♢	Q♠	3♠	8♡	2♡	5♠	...
5.	A♠	Q♠	7♠	J♠	5♠	5♠	...
...							
m.	A♣	A♡	A♣	A♡	A♣	A♢	...
...							

THE INFINITIES OF INDUCTIVE INFERENCE

1.	A♠	A♠	A♠	A♠	A♠	A♠	...
2.	A♣	A♣	A♣	A♣	A♣	A♣	...
3.	A♡	A♡	A♡	A♡	A♡	A♡	...
4.	A♢	Q♠	3♠	8♡	2♡	5♠	...
5.	A♠	Q♠	7♠	J♠	5♠	5♠	...
...							
m.	A♣	A♡	A♣	A♡	A♣	A♢	...
...							

THE INFINITIES OF INDUCTIVE INFERENCE

1.	A$\spadesuit$	A$\spadesuit$	A$\spadesuit$	A$\spadesuit$	A$\spadesuit$	A$\spadesuit$	...	
2.	A$\clubsuit$	A$\clubsuit$	A$\clubsuit$	A$\clubsuit$	A$\clubsuit$	A$\clubsuit$	...	
3.	A$\heartsuit$	A$\heartsuit$	A$\heartsuit$	A$\heartsuit$	A$\heartsuit$	A$\heartsuit$	...	
4.	A$\diamondsuit$	Q$\spadesuit$	3$\spadesuit$	8$\heartsuit$	2$\heartsuit$	5$\spadesuit$	...	
5.	A$\spadesuit$	Q$\spadesuit$	7$\spadesuit$	J$\spadesuit$	5$\spadesuit$	5$\spadesuit$	...	
...								
m.	A$\clubsuit$	A$\heartsuit$	A$\clubsuit$	A$\heartsuit$	A$\clubsuit$	A$\diamondsuit$	...	m-th
...								

THE INFINITIES OF INDUCTIVE INFERENCE

1.	A♠	A♠	A♠	A♠	A♠	A♠	...	
2.	A♣	A♣	A♣	A♣	A♣	A♣	...	
3.	A♡	A♡	A♡	A♡	A♡	A♡	...	
4.	A♢	Q♠	3♠	8♡	2♡	5♠	...	
5.	A♠	Q♠	7♠	J♠	5♠	5♠	...	
...								
m.	A♣	A♡	A♣	A♡	A♣	A♢	...	m-th
...								
								...

HOW MANY POSSIBLE RULES ARE THERE?

1. In principle...
2. Rule written down on a piece of paper.
3. Rule expressed by a natural language sentence.
4. Rule described by a theory that fills a 300 pages book.
5. Rule encoded by a Turing Machine program.

Descriptions are finite, and there are countably many of them.

HOW MANY POSSIBLE RULES ARE THERE?

1. In principle...
2. Rule written down on a piece of paper.
3. Rule expressed by a natural language sentence.
4. Rule described by a theory that fills a 300 pages book.
5. Rule encoded by a Turing Machine program.

Descriptions are finite, and there are countably many of them.

HOW MANY POSSIBLE RULES ARE THERE?

1. In principle...
2. Rule written down on a piece of paper.
3. Rule expressed by a natural language sentence.
4. Rule described by a theory that fills a 300 pages book.
5. Rule encoded by a Turing Machine program.

Descriptions are finite, and there are countably many of them.

HOW MANY POSSIBLE RULES ARE THERE?

1. In principle...
2. Rule written down on a piece of paper.
3. Rule expressed by a natural language sentence.
4. Rule described by a theory that fills a 300 pages book.
5. Rule encoded by a Turing Machine program.

Descriptions are finite, and there are countably many of them.

HOW MANY POSSIBLE RULES ARE THERE?

1. In principle...
2. Rule written down on a piece of paper.
3. Rule expressed by a natural language sentence.
4. Rule described by a theory that fills a 300 pages book.
5. Rule encoded by a Turing Machine program.

Descriptions are finite, and there are countably many of them.

HOW MANY SEQUENCES COMPLY TO ONE RULE?

1. The sequence has solely A♠-cards.
2. The sequence has solely ♠-cards.
3. The sequence has ♡-cards on even places.
4. The sequence is definable in first-order logic.
5. etc...

HOW MANY SEQUENCES COMPLY TO ONE RULE?

1. The sequence has solely A♠-cards.
2. The sequence has solely ♠-cards.
3. The sequence has ♡-cards on even places.
4. The sequence is definable in first-order logic.
5. etc...

HOW MANY SEQUENCES COMPLY TO ONE RULE?

1. The sequence has solely A♠-cards.
2. The sequence has solely ♠-cards.
3. The sequence has ♡-cards on even places.
4. The sequence is definable in first-order logic.
5. etc...

HOW MANY SEQUENCES COMPLY TO ONE RULE?

1. The sequence has solely A♠-cards.
2. The sequence has solely ♠-cards.
3. The sequence has ♡-cards on even places.
4. The sequence is definable in first-order logic.
5. etc...

DIFFERENT HYPOTHESIS SPACES

1. $\{(\text{all cards are } \spadesuit), (\text{all cards are } \diamondsuit)\}$
2. $\{(\spadesuit \text{ at the 4-th position}), \neg(\spadesuit \text{ at the 4-th position})\}$
3. $\underline{\{(\text{exactly } n \text{ cards are } \heartsuit) \mid n \in \mathbb{N}\}}$
4. $\{(\text{exactly } n \text{ cards are } \heartsuit) \mid n \in \mathbb{N}\} \cup \{(\infty \text{ cards are } \heartsuit)\}$

A♠ Q♠ 3♠ A♠ Q♠ 4♡ Q♠ Q♡...

DIFFERENT HYPOTHESIS SPACES

1. $\{(\text{all cards are } \spadesuit), (\text{all cards are } \diamondsuit)\}$
2. $\{(\spadesuit \text{ at the 4-th position}), \neg(\spadesuit \text{ at the 4-th position})\}$
3. $\underline{\{(\text{exactly } n \text{ cards are } \heartsuit) \mid n \in \mathbb{N}\}}$
4. $\{(\text{exactly } n \text{ cards are } \heartsuit) \mid n \in \mathbb{N}\} \cup \{(\infty \text{ cards are } \heartsuit)\}$

A♠ Q♠ 3♠ A♠ Q♠ 4♡ Q♠ Q♡...

DIFFERENT HYPOTHESIS SPACES

1. $\{(\text{all cards are } \spadesuit), (\text{all cards are } \diamondsuit)\}$
2. $\{(\spadesuit \text{ at the 4-th position}), \neg(\spadesuit \text{ at the 4-th position})\}$
3. $\underline{\{(\text{exactly } n \text{ cards are } \heartsuit) \mid n \in \mathbb{N}\}}$
4. $\{(\text{exactly } n \text{ cards are } \heartsuit) \mid n \in \mathbb{N}\} \cup \{(\infty \text{ cards are } \heartsuit)\}$

A♠ Q♠ 3♠ A♠ Q♠ 4♡ Q♠ Q♡...

DIFFERENT HYPOTHESIS SPACES

1. $\{(\text{all cards are } \spadesuit), (\text{all cards are } \diamondsuit)\}$
2. $\{(\spadesuit \text{ at the 4-th position}), \neg(\spadesuit \text{ at the 4-th position})\}$
3. $\underline{\{(\text{exactly } n \text{ cards are } \heartsuit) \mid n \in \mathbb{N}\}}$
4. $\{(\text{exactly } n \text{ cards are } \heartsuit) \mid n \in \mathbb{N}\} \cup \{(\infty \text{ cards are } \heartsuit)\}$

A♠ Q♠ 3♠ A♠ Q♠ 4♡ Q♠ Q♡...

DIFFERENT HYPOTHESIS SPACES

1. $\{(\text{all cards are } \spadesuit), (\text{all cards are } \diamondsuit)\}$
2. $\{(\spadesuit \text{ at the 4-th position}), \neg(\spadesuit \text{ at the 4-th position})\}$
3. $\underline{\{(\text{exactly } n \text{ cards are } \heartsuit) \mid n \in \mathbb{N}\}}$
4. $\{(\text{exactly } n \text{ cards are } \heartsuit) \mid n \in \mathbb{N}\} \cup \{(\infty \text{ cards are } \heartsuit)\}$

A♠ Q♠ 3♠ A♠ Q♠ 4♡ Q♠ Q♡...

DIFFERENT HYPOTHESIS SPACES

1. $\{(\text{all cards are } \spadesuit), (\text{all cards are } \diamondsuit)\}$
2. $\{(\spadesuit \text{ at the 4-th position}), \neg(\spadesuit \text{ at the 4-th position})\}$
3. $\underline{\{(\text{exactly } n \text{ cards are } \heartsuit) \mid n \in \mathbb{N}\}}$
4. $\{(\text{exactly } n \text{ cards are } \heartsuit) \mid n \in \mathbb{N}\} \cup \{(\infty \text{ cards are } \heartsuit)\}$

A♠ Q♠ 3♠ A♠ Q♠ 4♡ Q♠ Q♡...

DIFFERENT HYPOTHESIS SPACES

1. $\{(\text{all cards are } \spadesuit), (\text{all cards are } \diamondsuit)\}$
2. $\{(\spadesuit \text{ at the 4-th position}), \neg(\spadesuit \text{ at the 4-th position})\}$
3. $\underline{\{(\text{exactly } n \text{ cards are } \heartsuit) \mid n \in \mathbb{N}\}}$
4. $\{(\text{exactly } n \text{ cards are } \heartsuit) \mid n \in \mathbb{N}\} \cup \{(\infty \text{ cards are } \heartsuit)\}$

A♠ Q♠ 3♠ A♠ Q♠ 4♡ Q♠ Q♡. . .

DIFFERENT HYPOTHESIS SPACES

1. $\{(\text{all cards are } \spadesuit), (\text{all cards are } \diamondsuit)\}$
2. $\{(\spadesuit \text{ at the 4-th position}), \neg(\spadesuit \text{ at the 4-th position})\}$
3. $\underline{\{(\text{exactly } n \text{ cards are } \heartsuit) \mid n \in \mathbb{N}\}}$
4. $\{(\text{exactly } n \text{ cards are } \heartsuit) \mid n \in \mathbb{N}\} \cup \{(\infty \text{ cards are } \heartsuit)\}$

A♠ Q♠ 3♠ A♠ Q♠ 4♡ Q♠ Q♡. . .

DIFFERENT HYPOTHESIS SPACES

1. $\{(\text{all cards are } \spadesuit), (\text{all cards are } \diamondsuit)\}$
2. $\{(\spadesuit \text{ at the 4-th position}), \neg(\spadesuit \text{ at the 4-th position})\}$
3. $\underline{\{(\text{exactly } n \text{ cards are } \heartsuit) \mid n \in \mathbb{N}\}}$
4. $\{(\text{exactly } n \text{ cards are } \heartsuit) \mid n \in \mathbb{N}\} \cup \{(\infty \text{ cards are } \heartsuit)\}$

A♠ Q♠ 3♠ A♠ Q♠ 4♡ Q♠ Q♡. . .

PLAN FOR TODAY

1. Inductive Inference: The Eleusis Game
2. Learning Paradigms and Perspectives
3. Introduction to Epistemic Logic

WHAT DO WE MEAN BY 'LEARNING'?

In this course, we will present a general **qualitative** model of (exact) learning:

- An agent receives incoming data consistent with an underlying concept
- She **learns** something about the underlying concept, e.g., she achieves a desired **type of knowledge or belief** about the underlying concept.

Our perspective generalizes three different learning paradigms:

- **Set Learning:** paradigm of Computational Learning Theory
- **Function Learning:** paradigm of Machine Learning, Bayesian Learning, & Reinforcement Learning
- **Epistemic Learning:** paradigm of Belief Revision Theory & Dynamic Epistemic Logic

LEARNING PARADIGMS

1. Possible realities:

2. Hypotheses:

3. Information accessible to the learner:

4. Learner:

5. Success criterion:

Set Learning

1. Possible realities:

   Sets of integers

2. Hypotheses:

   Names of sets

3. Information accessible to the learner:

   Sequences of numbers

4. Learner:

   Function that takes a sequence and outputs a hypothesis

5. Success criterion:

   After finite number of outputs stabilize on a correct answer

Function Learning

1. Possible realities:

   Functions

2. Hypotheses:

   Names or implementations of functions

3. Information accessible to the learner:

   Sequences of input-output pairs (arguments, value)

4. Learner:

   Function that takes a sequence and outputs a hypothesis

5. Success criterion:

   After finite number of outputs stabilize on a correct answer

Epistemic Learning

1. Possible realities:

   Models over a given logic (language & semantics)

2. Hypotheses:

   Formulas in the logic

3. Information accessible to the learner:

   Sequences of atomic formulas and negations thereof

4. Learner:

   Function that takes a sequence and outputs a hypothesis

5. Success criterion:

   After finite number of outputs stabilize on a correct answer

ADDITIONAL NOTES ON PARADIGM SPECIFICATION

- Hypotheses are systematic descriptions of possible realities.
- The hypotheses are finite descriptions of sets / functions / formulas
- e.g., Turing machines, grammars, programs, logical formulas, neural network weights, etc.

ADDITIONAL NOTES ON PARADIGM SPECIFICATION

- In interesting cases the data available at a given step presents only partial information about a possible reality.
- The character of data is determined by the setting, e.g. in language learning one might consider only positive or positive and negative information about a possible reality.
- In the basic setting, data is “passively” presented to the learner. In some paradigms the learner can actively request or give attention to particular information.

ADDITIONAL NOTES ON PARADIGM SPECIFICATION

- Finite identifiability
- Identifiability in the limit
- Gradual identifiability

We will fix the success criterion to be:

After a finite time the learner's answers stabilize to the correct answer.

THE GAME OF LEARNING IN THE LIMIT

Just like our card game, you can think of learning in general as a game played between a **learner** and **nature**.

- A class of possible worlds (available to both players).
- Nature chooses one of them (learner does not know which).
- Nature generates data about the world.
- From inductively given data learner draws her conjectures.
- After each new input, learner can answer with an updated hypothesis.
- Learner succeeds if **she stabilizes to a correct hypothesis**.

Her success depends on the problem, but also on her **learning strategy**.

THE GAME OF LEARNING IN THE LIMIT

Just like our card game, you can think of learning in general as a game played between a **learner** and **nature**.

- A class of possible worlds (available to both players).
- Nature chooses one of them (learner does not know which).
- Nature generates data about the world.
- From inductively given data learner draws her conjectures.
- After each new input, learner can answer with an updated hypothesis.
- Learner succeeds if **she stabilizes to a correct hypothesis**.

Her success depends on the problem, but also on her **learning strategy**.

PLAN FOR TODAY

1. Inductive Inference: The Eleusis Game
2. Learning Paradigms and Perspectives
3. Introduction to Epistemic Logic

KNOWLEDGE AND POSSIBLE WORLDS

- Besides of the current state of affairs,
- there is a number of other **possible states of affairs** or “worlds”.

An agent knows a fact φ if φ is true at all the worlds she **considers possible**.

SEATTLE EXAMPLE

Ann is walking the streets of Copenhagen on a sunny day. She has no information at all about the weather in Seattle.

Thus, in all the worlds that she considers possible, it is sunny in Copenhagen.

Since she has no information about the weather in Seattle, there are worlds she considers possible in which it is sunny in Seattle, and others in which not.

Thus, this agent knows that it is sunny in Copenhagen, but she does not know whether it is sunny in Seattle.

SEATTLE EXAMPLE

Ann is walking the streets of Copenhagen on a sunny day. She has no information at all about the weather in Seattle.

Thus, in all the worlds that she considers possible, it is sunny in Copenhagen.

Since she has no information about the weather in Seattle, there are worlds she considers possible in which it is sunny in Seattle, and others in which not.

Thus, this agent knows that it is sunny in Copenhagen, but she does not know whether it is sunny in Seattle.

SEATTLE EXAMPLE

Ann is walking the streets of Copenhagen on a sunny day. She has no information at all about the weather in Seattle.

Thus, in all the worlds that she considers possible, it is sunny in Copenhagen.

Since she has no information about the weather in Seattle, there are worlds she considers possible in which it is sunny in Seattle, and others in which not.

Thus, this agent knows that it is sunny in Copenhagen, but she does not know whether it is sunny in Seattle.

SEATTLE EXAMPLE

Ann is walking the streets of Copenhagen on a sunny day. She has no information at all about the weather in Seattle.

Thus, in all the worlds that she considers possible, it is sunny in Copenhagen.

Since she has no information about the weather in Seattle, there are worlds she considers possible in which it is sunny in Seattle, and others in which not.

Thus, this agent knows that it is sunny in Copenhagen, but she does not know whether it is sunny in Seattle.

SEATTLE EXAMPLE

Ann is walking the streets of Copenhagen on a sunny day. She has no information at all about the weather in Seattle.

Thus, in all the worlds that she considers possible, it is sunny in Copenhagen.

Since she has no information about the weather in Seattle, there are worlds she considers possible in which it is sunny in Seattle, and others in which not.

Thus, this agent knows that it is sunny in Copenhagen, but she does not know whether it is sunny in Seattle.

SEATTLE EXAMPLE

If the agent acquires additional information from a reliable source:

It is currently sunny in Seattle.

She would no longer consider possible worlds in which it is raining in Seattle.

Intuitively, **the fewer worlds = less uncertainty, and more knowledge**.

SEATTLE EXAMPLE

If the agent acquires additional information from a reliable source:

It is currently sunny in Seattle.

She would no longer consider possible worlds in which it is raining in Seattle.

Intuitively, **the fewer worlds = less uncertainty, and more knowledge**.

SEATTLE EXAMPLE

If the agent acquires additional information from a reliable source:

It is currently sunny in Seattle.

She would no longer consider possible worlds in which it is raining in Seattle.

Intuitively, **the fewer worlds = less uncertainty, and more knowledge**.

SEATTLE EXAMPLE

If the agent acquires additional information from a reliable source:

It is currently sunny in Seattle.

She would no longer consider possible worlds in which it is raining in Seattle.

Intuitively, **the fewer worlds = less uncertainty, and more knowledge**.

EPISTEMIC LOGIC: BRIEF HISTORY

Epistemic logic was introduced as a modal logic in **1962** by **Jaakko Hintikka**.

In his logic both knowledge and belief are introduced as two separate concepts. His logic had two modal operators K and B (for knowledge and belief) to represent the two attitudes separately.

SYNTAX: THE LANGUAGE OF EPISTEMIC LOGIC

Definition (Language of Epistemic Logic)
Prop is a (countable) set of propositions, with $p \in Prop$, and $\mathcal{A} = \{1, \ldots, n\}$ is a set of agents.

$$\varphi := \top \;\Big|\; p \;\Big|\; \neg\varphi \;\Big|\; \varphi \wedge \varphi \;\Big|\; K_i\varphi$$

where $\top$ is a special symbol and $i \in \mathcal{A}$ is the name of some agent.

In case we are only dealing with one agent, we can also omit the index.

SYNTAX: THE LANGUAGE OF EPISTEMIC LOGIC

$K\varphi$: I know that φ.

$\neg K\varphi$: I don't know that φ.

$K\neg\varphi$: I know that not φ.

SYNTAX: THE LANGUAGE OF EPISTEMIC LOGIC

$K\varphi$: I know that φ.

$\neg K\varphi$: I don't know that φ.

$K\neg\varphi$: I know that not φ.

SYNTAX: THE LANGUAGE OF EPISTEMIC LOGIC

$K\varphi$: I know that φ.

$\neg K\varphi$: I don't know that φ.

$K\neg\varphi$: I know that not φ.

SYNTAX: THE LANGUAGE OF EPISTEMIC LOGIC

$K\varphi$: I know that φ.

$\neg K\varphi$: I don't know that φ.

$K\neg\varphi$: I know that not φ.

SYNTAX: THE LANGUAGE OF EPISTEMIC LOGIC

$K\varphi$: I know that φ.

$\neg K\varphi$: I don't know that φ.

$K\neg\varphi$: I know that not φ.

SYNTAX: THE LANGUAGE OF EPISTEMIC LOGIC

$K\varphi$: I know that φ.

$\neg K\varphi$: I don't know that φ.

$K\neg\varphi$: I know that not φ.

SYNTAX: THE LANGUAGE OF EPISTEMIC LOGIC

$K\varphi$: I know that φ.

$\neg K\varphi$: I don't know that φ.

$K\neg\varphi$: I know that not φ.

Definition (Possible world model or epistemic model)

A **possible world model** M for n agents over *Prop* is $(S, \mathcal{K}_1, \ldots, \mathcal{K}_n, v)$, where:

1. S is a non-empty set states (or worlds);
2. $v : Prop \to \wp(S)$ is a valuation;
3. for each agent i, $\mathcal{K}_i$ is a binary relation on S.

ADDITIONAL EXPLANATION

1. v tells us whether a proposition is true or false in state.
2. $\mathcal{K}_i$ captures the possibility relation according to agent i, i.e.,

   $(s,t) \in \mathcal{K}_i$ if agent i considers world t possible, given her information in world s.

   $\mathcal{K}_i$ is a possibility (or accessibility, or indistinguishability) relation; it says what worlds agent i considers possible (or can access) in any given world.

EQUIVALENCE POSSIBILITY RELATION

$\mathcal{K}_i$ is an **equivalence** relation on S, i.e., it is a binary relation that is:

1. reflexive: for all $s \in S$, we have $(s, s) \in \mathcal{K}_i$,
2. symmetric: for all $s, t \in S$, we have $(s, t) \in \mathcal{K}_i$ iff $(t, s) \in \mathcal{K}_i$,
3. transitive: for all $s, t, u \in S$, we have that if $(s, t) \in \mathcal{K}_i$ and $(t, u) \in \mathcal{K}_i$, then $(s, u) \in \mathcal{K}_i$.

WHEN IS A FORMULA TRUE IN A SITUATION?

We write $(M, s) \vDash \varphi$ to say that φ is true at s in M.

Definition

$(M, s) \vDash \top$		always
$(M, s) \vDash p$	iff	$s \in v(p)$
$(M, s) \vDash \neg\varphi$	iff	it is not the case that: $(M, s) \vDash \varphi$
$(M, s) \vDash \varphi \wedge \psi$	iff	$(M, s) \vDash \varphi$ and $(M, s) \vDash \psi$
$(M, s) \vDash K_i\varphi$	**iff**	**for all** v **with** $(s, v) \in \mathcal{K}_i$**,** $(M, v) \vDash \varphi$

We use $(M, s) \nvDash \varphi$ to express that φ is false at s in M.

$K_i\varphi$ is false at state s when there a t such that $(s, t) \in \mathcal{K}_i$ and φ is false at v.

WHEN IS A FORMULA TRUE IN A SITUATION?

We write $(M, s) \models \varphi$ to say that φ is true at s in M.

Definition

$(M, s) \models \top$		always
$(M, s) \models p$	iff	$s \in v(p)$
$(M, s) \models \neg\varphi$	iff	it is not the case that: $(M, s) \models \varphi$
$(M, s) \models \varphi \wedge \psi$	iff	$(M, s) \models \varphi$ and $(M, s) \models \psi$
$(M, s) \models K_i\varphi$	**iff**	**for all** v **with** $(s, v) \in \mathcal{K}_i$**,** $(M, v) \models \varphi$

We use $(M, s) \not\models \varphi$ to express that φ is false at s in M.

$K_i\varphi$ is false at state s when there a t such that $(s, t) \in \mathcal{K}_i$ and φ is false at v.

PRACTICE: EPISTEMIC LOGIC

See Day 1 practice sheet

What are the properties of K?

How well does the K operator model knowledge?

We will try to answer this question

by looking at formulas about knowledge that are always true

in a given kind of possible world models.

What are the properties of K?

How well does the K operator model knowledge?

We will try to answer this question

by looking at formulas about knowledge that are always true

in a given kind of possible world models.

VALIDITY AND SATISFIABILITY

Definition

Given a model $M = (S, \mathcal{K}_1, \ldots, \mathcal{K}_n, v)$, we say that:

- φ is **valid in** M, $M \models \varphi$, if $(M, s) \models \varphi$ for every state $s \in S$.
- φ is **satisfiable in** M, if $(M, s) \models \varphi$ for some state $s \in S$.
- φ is **valid**, $\models \varphi$, if φ is valid in all models.
- φ is **satisfiable**, if φ is satisfiable in a model.

VALID LAWS OF KNOWLEDGE (THE S5 SYSTEM)

The following formulas are valid whenever $\mathcal{K}_i$ is an equivalence relation:

Distribution of Knowledge: $K_i\varphi \wedge K_i(\varphi \rightarrow \psi) \rightarrow K_i\psi$

Each agent knows all the logical consequences of her knowledge.

Knowledge Generalization: For all models M, if $M \models \varphi$ then $M \models K_i\varphi$

Each agent knows all the formulas that are **valid** in a given model.

Truthfulness of Knowledge: $\mathrm{K}_i\varphi \rightarrow \varphi$

Agents can only know facts. (Contrast this with belief)

Pos. and Neg. Introspection: $\mathrm{K}_i\varphi \rightarrow \mathrm{K}_i\mathrm{K}_i\varphi$ and $\neg\mathrm{K}_i\varphi \rightarrow \mathrm{K}_i\neg\mathrm{K}_i\varphi$

Agents know what they know and what they do not know.

VALID LAWS OF KNOWLEDGE (THE S5 SYSTEM)

The following formulas are valid whenever $\mathcal{K}_i$ is an equivalence relation:

Distribution of Knowledge: $K_i\varphi \wedge K_i(\varphi \rightarrow \psi) \rightarrow K_i\psi$

Each agent knows all the logical consequences of her knowledge.

Knowledge Generalization: For all models M, if $M \vDash \varphi$ then $M \vDash K_i\varphi$

Each agent knows all the formulas that are **valid** in a given model.

Truthfulness of Knowledge: $\mathrm{K}_i\varphi \rightarrow \varphi$

Agents can only know facts. (Contrast this with belief)

Pos. and Neg. Introspection: $\mathrm{K}_i\varphi \rightarrow \mathrm{K}_i\mathrm{K}_i\varphi$ and $\neg\mathrm{K}_i\varphi \rightarrow \mathrm{K}_i\neg\mathrm{K}_i\varphi$

Agents know what they know and what they do not know.

VALID LAWS OF KNOWLEDGE (THE S5 SYSTEM)

The following formulas are valid whenever $\mathcal{K}_i$ is an equivalence relation:

Distribution of Knowledge: $K_i\varphi \wedge K_i(\varphi \rightarrow \psi) \rightarrow K_i\psi$

Each agent knows all the logical consequences of her knowledge.

Knowledge Generalization: For all models M, if $M \vDash \varphi$ then $M \vDash K_i\varphi$

Each agent knows all the formulas that are **valid** in a given model.

Truthfulness of Knowledge: $\mathrm{K}_i\varphi \rightarrow \varphi$

Agents can only know facts. (Contrast this with belief)

Pos. and Neg. Introspection: $\mathrm{K}_i\varphi \rightarrow \mathrm{K}_i\mathrm{K}_i\varphi$ and $\neg\mathrm{K}_i\varphi \rightarrow \mathrm{K}_i\neg\mathrm{K}_i\varphi$

Agents know what they know and what they do not know.

VALID LAWS OF KNOWLEDGE (THE S5 SYSTEM)

The following formulas are valid whenever $\mathcal{K}_i$ is an equivalence relation:

Distribution of Knowledge: $K_i\varphi \wedge K_i(\varphi \rightarrow \psi) \rightarrow K_i\psi$

Each agent knows all the logical consequences of her knowledge.

Knowledge Generalization: For all models M, if $M \vDash \varphi$ then $M \vDash K_i\varphi$

Each agent knows all the formulas that are **valid** in a given model.

Truthfulness of Knowledge: $\mathbf{K}_i\varphi \rightarrow \varphi$

Agents can only know facts. (Contrast this with belief)

Pos. and Neg. Introspection: $\mathbf{K}_i\varphi \rightarrow \mathbf{K}_i\mathbf{K}_i\varphi$ and $\neg\mathbf{K}_i\varphi \rightarrow \mathbf{K}_i\neg\mathbf{K}_i\varphi$

Agents know what they know and what they do not know.

VALID LAWS OF KNOWLEDGE (THE S5 SYSTEM)

The following formulas are valid whenever $\mathcal{K}_i$ is an equivalence relation:

Distribution of Knowledge: $K_i\varphi \wedge K_i(\varphi \rightarrow \psi) \rightarrow K_i\psi$

Each agent knows all the logical consequences of her knowledge.

Knowledge Generalization: For all models M, if $M \vDash \varphi$ then $M \vDash K_i\varphi$

Each agent knows all the formulas that are **valid** in a given model.

Truthfulness of Knowledge: $\mathbf{K}_i\varphi \rightarrow \varphi$

Agents can only know facts. (Contrast this with belief)

Pos. and Neg. Introspection: $\mathbf{K}_i\varphi \rightarrow \mathbf{K}_i\mathbf{K}_i\varphi$ and $\neg\mathbf{K}_i\varphi \rightarrow \mathbf{K}_i\neg\mathbf{K}_i\varphi$

Agents know what they know and what they do not know.

EXAMPLE OF VALIDITY ARGUMENTS: POSITIVE INTROSPECTION

Proposition

$\vDash K_i\varphi \to K_iK_i\varphi$ in the class of models with equivalence possibility relations.

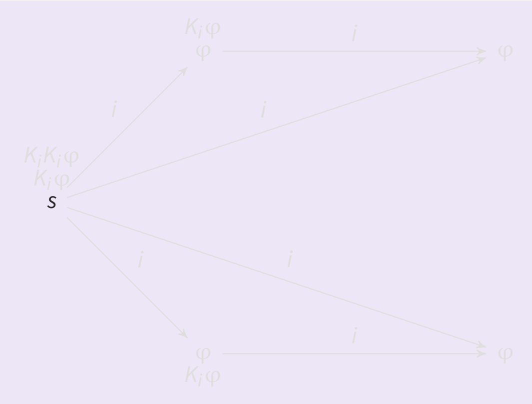
EXAMPLE OF VALIDITY ARGUMENTS: POSITIVE INTROSPECTION

Proposition

$\models K_i\varphi \to K_iK_i\varphi$ in the class of models with equivalence possibility relations.

Proposition

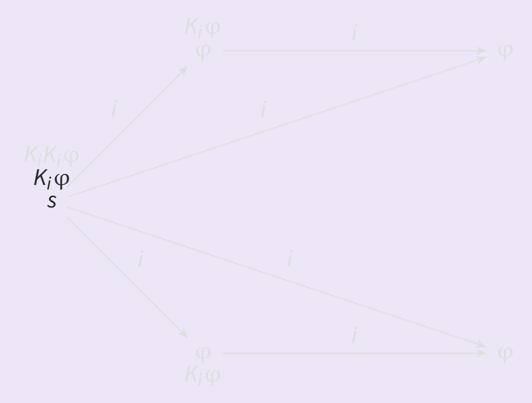
$\vDash K_i\varphi \to K_iK_i\varphi$ in the class of models with equivalence possibility relations.



EXAMPLE OF VALIDITY ARGUMENTS: POSITIVE INTROSPECTION

Proposition

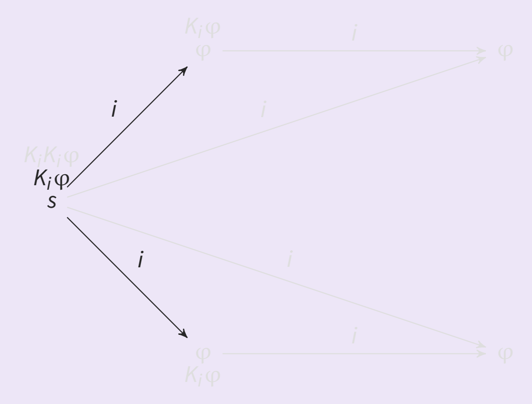
$\vDash K_i\varphi \to K_iK_i\varphi$ in the class of models with equivalence possibility relations.



EXAMPLE OF VALIDITY ARGUMENTS: POSITIVE INTROSPECTION

Proposition

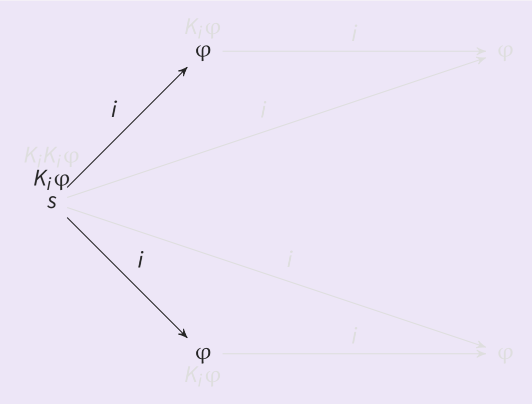
$\models K_i\varphi \to K_iK_i\varphi$ in the class of models with equivalence possibility relations.



EXAMPLE OF VALIDITY ARGUMENTS: POSITIVE INTROSPECTION

Proposition

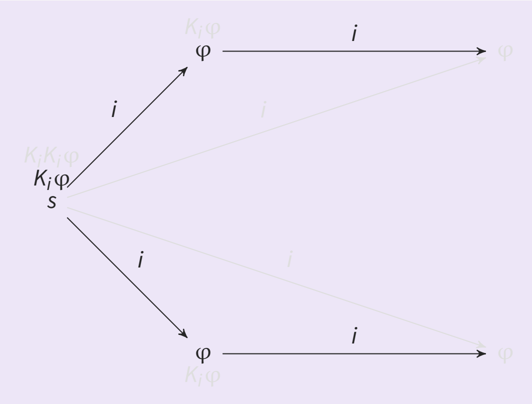
$\vDash K_i\varphi \to K_iK_i\varphi$ in the class of models with equivalence possibility relations.



EXAMPLE OF VALIDITY ARGUMENTS: POSITIVE INTROSPECTION

Proposition

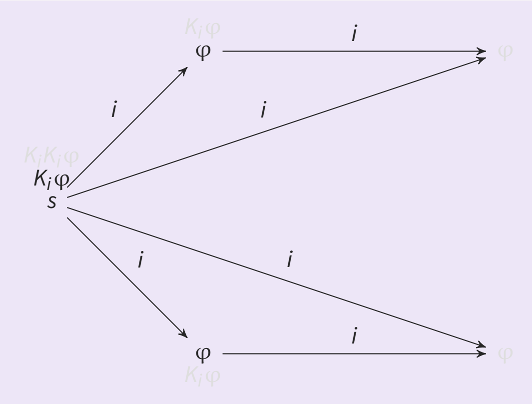
$\vDash K_i\varphi \to K_iK_i\varphi$ in the class of models with equivalence possibility relations.



EXAMPLE OF VALIDITY ARGUMENTS: POSITIVE INTROSPECTION

Proposition

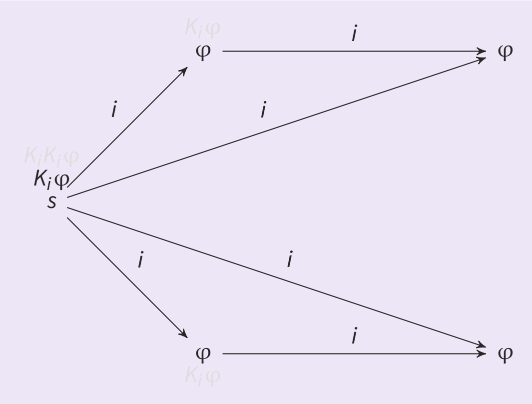
$\models K_i\varphi \to K_iK_i\varphi$ in the class of models with equivalence possibility relations.



EXAMPLE OF VALIDITY ARGUMENTS: POSITIVE INTROSPECTION

Proposition

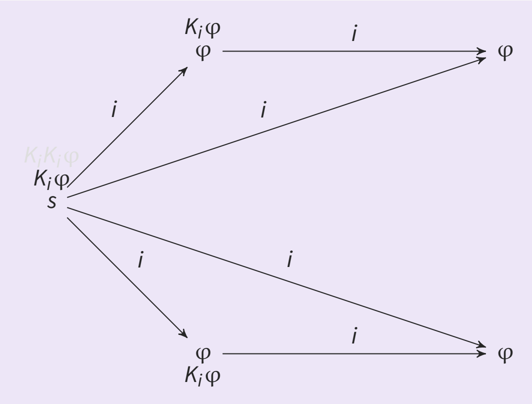
$\models K_i\varphi \to K_iK_i\varphi$ in the class of models with equivalence possibility relations.



EXAMPLE OF VALIDITY ARGUMENTS: POSITIVE INTROSPECTION

Proposition

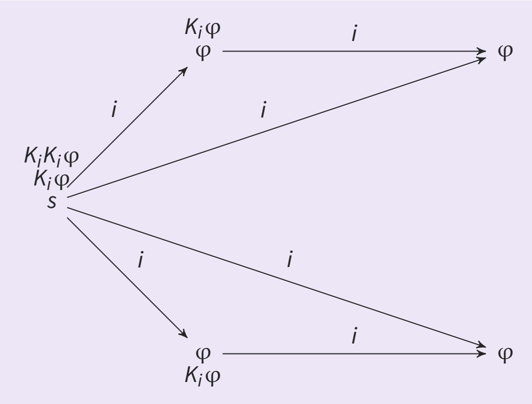
$\vDash K_i\varphi \rightarrow K_iK_i\varphi$ in the class of models with equivalence possibility relations.



EXAMPLE OF VALIDITY ARGUMENTS: POSITIVE INTROSPECTION

Proposition

$\models K_i\varphi \rightarrow K_iK_i\varphi$ in the class of models with equivalence possibility relations.



AXIOMATIC SYSTEM

An **axiomatic system** consists of:

- a set of formulas called **axioms** and
- a set of **rules of inference**.

Together they are used to infer (derive) **theorems**.

PROOF IN AN AXIOMATIC SYSTEM

A **proof** of a formula ψ is a sequence of formulas $\varphi_1, \ldots, \varphi_n$, with $\varphi_n = \psi$, such that each φ_k is either an axiom or it is derived from previous formulas by rules of inference.

When such a proof exists, we say that ψ is a **theorem** (of the system) and that ψ is **provable** (in the system), denoted by:

$$\vdash \psi$$

We can use substitution instances of axioms and inference rules.

E.g., the formula $(p \vee q) \vee \neg(p \vee q)$ is an instance of the tautology $\varphi \vee \neg\varphi$.

PROOF IN AN AXIOMATIC SYSTEM

A **proof** of a formula ψ is a sequence of formulas $\varphi_1, \dots, \varphi_n$, with $\varphi_n = \psi$, such that each φ_k is either an axiom or it is derived from previous formulas by rules of inference.

When such a proof exists, we say that ψ is a **theorem** (of the system) and that ψ is **provable** (in the system), denoted by:

$$\vdash \psi$$

We can use substitution instances of axioms and inference rules.

E.g., the formula $(p \vee q) \vee \neg(p \vee q)$ is an instance of the tautology $\varphi \vee \neg\varphi$.

PROOF IN AN AXIOMATIC SYSTEM

A **proof** of a formula ψ is a sequence of formulas $\varphi_1, \ldots, \varphi_n$, with $\varphi_n = \psi$, such that each φ_k is either an axiom or it is derived from previous formulas by rules of inference.

When such a proof exists, we say that ψ is a **theorem** (of the system) and that ψ is **provable** (in the system), denoted by:

$$\vdash \psi$$

We can use substitution instances of axioms and inference rules.

E.g., the formula $(p \vee q) \vee \neg(p \vee q)$ is an instance of the tautology $\varphi \vee \neg\varphi$.

SYSTEM *S5* FOR EPISTEMIC LOGIC

A1. All tautologies of propositional logic

A2. $(K_i\varphi \land K_i(\varphi \to \psi)) \to K_i\psi, i \in \{1, \ldots, n\}$

A3. $K_i\varphi \to \varphi, i \in \{1, \ldots, n\}$

A4. $K_i\varphi \to K_iK_i\varphi, i \in \{1, \ldots, n\}$

A5. $\neg K_i\varphi \to K_i\neg K_i\varphi, i \in \{1, \ldots, n\}$

$$\frac{\vdash \varphi \qquad \vdash (\varphi \to \psi)}{\psi}\ \text{R1} \qquad\qquad \frac{\vdash \varphi}{K_i\varphi,\ \text{for each } i \in \{1, \ldots, n\}}\ \text{R2}$$

LANGUAGES AND MODELS

Take *Prop* to be a set of propositions.

- Let $\mathcal{L}_n(Prop)$ be the set of formulas that can be built up starting from the primitive propositions in *Prop*, using $\wedge$, $\neg$, and $K_1, \ldots, K_n$.
- Let $\mathcal{M}_n(Prop)$ be the class of all possible world models for n agents over *Prop* (with no restrictions on the $\mathcal{K}_i$ relations).
- $\mathcal{M}_n(Prop)$ can be restricted by specifying the $\mathcal{K}_i$ relations, e.g.:

  for $\mathcal{M}_n^{rst}(Prop)$, $\mathcal{K}_i$ relations are reflexive, symmetric, and transitive.

Note: *Prop* is fixed from now on and we suppress it from the notation.

LANGUAGES AND MODELS

Take *Prop* to be a set of propositions.

- Let $\mathcal{L}_n(Prop)$ be the set of formulas that can be built up starting from the primitive propositions in *Prop*, using $\wedge$, $\neg$, and $K_1, \ldots, K_n$.
- Let $\mathcal{M}_n(Prop)$ be the class of all possible world models for n agents over *Prop* (with no restrictions on the $\mathcal{K}_i$ relations).
- $\mathcal{M}_n(Prop)$ can be restricted by specifying the $\mathcal{K}_i$ relations, e.g.:

  for $\mathcal{M}_n^{rst}(Prop)$, $\mathcal{K}_i$ relations are reflexive, symmetric, and transitive.

Note: *Prop* is fixed from now on and we suppress it from the notation.

VALIDITY WITH RESPECT TO A CLASS OF MODELS

Definition

We say that φ is valid with respect to $\mathcal{M}_n$, and write $\mathcal{M}_n \vDash \varphi$, if φ is valid in all the structures in $\mathcal{M}_n$.

- If $\mathcal{M}$ is some subclass of $\mathcal{M}_n$, φ is valid with respect to $\mathcal{M}$, $\mathcal{M} \vDash \varphi$, if φ is valid in all the structures in $\mathcal{M}$.
- If $\mathcal{M}$ is some subclass of $\mathcal{M}_n$, φ is satisfiable with respect to $\mathcal{M}$, if φ is satisfied in some structure in $\mathcal{M}$.

SOUNDNESS AND COMPLETENESS

What is the ideal relationship between

provability (in a given axiomatic system)

and

validity (in a given class of models)?

Definition

1. An axiom system AX is **sound** for a language $\mathcal{L}$ wrt a class $\mathcal{M}$ of structures if every formula in $\mathcal{L}$ provable in AX is valid wrt $\mathcal{M}$.
2. An axiom system AX is **complete** for a language $\mathcal{L}$ wrt a class $\mathcal{M}$ of structures if every formula in $\mathcal{L}$ that is valid wrt $\mathcal{M}$ is provable in AX.

***AX* characterizes the class $\mathcal{M}$ if it is sound and complete axiomatization of $\mathcal{M}$**

in other words

SOUNDNESS AND COMPLETENESS

What is the ideal relationship between
provability (in a given axiomatic system)
and
validity (in a given class of models)?

Definition

1. An axiom system *AX* is **sound** for a language $\mathcal{L}$ wrt a class $\mathcal{M}$ of structures if every formula in $\mathcal{L}$ provable in *AX* is valid wrt $\mathcal{M}$.
2. An axiom system *AX* is **complete** for a language $\mathcal{L}$ wrt a class $\mathcal{M}$ of structures if every formula in $\mathcal{L}$ that is valid wrt $\mathcal{M}$ is provable in *AX*.

***AX* characterizes the class $\mathcal{M}$ if it is sound and complete axiomatization of $\mathcal{M}$**

in other words

SOUNDNESS AND COMPLETENESS

What is the ideal relationship between

provability (in a given axiomatic system)

and

validity (in a given class of models)?

Definition

1. An axiom system *AX* is **sound** for a language $\mathcal{L}$ wrt a class $\mathcal{M}$ of structures if every formula in $\mathcal{L}$ provable in *AX* is valid wrt $\mathcal{M}$.
2. An axiom system *AX* is **complete** for a language $\mathcal{L}$ wrt a class $\mathcal{M}$ of structures if every formula in $\mathcal{L}$ that is valid wrt $\mathcal{M}$ is provable in *AX*.

***AX* characterizes the class $\mathcal{M}$ if it is sound and complete axiomatization of $\mathcal{M}$**

in other words

What is the ideal relationship between
provability (in a given axiomatic system)
and
validity (in a given class of models)?

Definition

1. An axiom system *AX* is **sound** for a language $\mathcal{L}$ wrt a class $\mathcal{M}$ of structures if every formula in $\mathcal{L}$ provable in *AX* is valid wrt $\mathcal{M}$.
2. An axiom system *AX* is **complete** for a language $\mathcal{L}$ wrt a class $\mathcal{M}$ of structures if every formula in $\mathcal{L}$ that is valid wrt $\mathcal{M}$ is provable in *AX*.

***AX* characterizes the class $\mathcal{M}$ if it is sound and complete axiomatization of $\mathcal{M}$**

in other words

SOUNDNESS AND COMPLETENESS OF $\mathcal{K}_n$ AND $\mathcal{S}5_n$

Theorem

$\mathcal{K}_n$ is sound and complete with respect to $\mathcal{M}_n$ for the language $\mathcal{L}_n$.

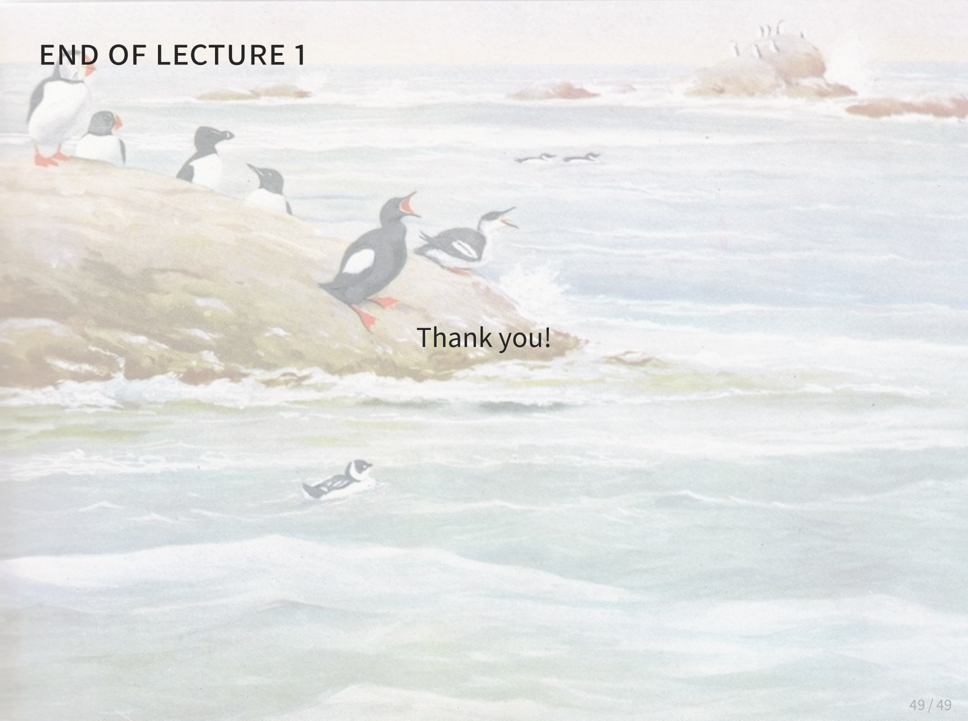
Theorem

$\mathcal{S}5_n$ is sound and complete with respect to $\mathcal{M}_n^{rst}$ for the language $\mathcal{L}_n$.

OVERVIEW OF COMPLETENESS RESULTS

K	the class of all frames
K4	the class of transitive frames
T	the class of reflexive frames
B	the class of symmetric frames
KD	the class of right-unbounded frames
S4	the class of reflexive, transitive frames
S5	the class of frames whose relation is an equivalence relation
K4.3	the class of transitive frames with no branching to the right
S4.3	the class of reflexive, transitive frames with no branching to the right
KL	the class of finite transitive trees (*weak* completeness only)

END OF LECTURE 1



Thank you!