Logic and Learning

Day 2: Truth-tracking by Belief Revision

Nina Gierasimczuk

DTU

Tsinghua Logic Summer School
June 30th–July 3rd, 2026

Outline

Outline

Introduction

Learnability in Epistemic Spaces

Constructive, Order-driven Learning: Belief Revision

Bias and Belief Revision

Belief revision and learning

We investigate order-driven belief revision methods:

- in particular, some aspects of their long term behaviour:
- **how good are they as learning methods?**
- under what conditions are they reliable?

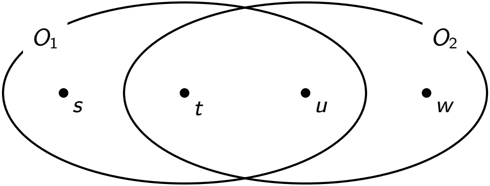
We can then use the same apparatus for:

- a topological characterisation of learnability and solvability;
- a modal dynamic logic of learnability;
- **collective learning.**

Outline

Epistemic Spaces and Observables

Definition
An **epistemic space** is a pair $\mathbb{S} = (S, \mathcal{O})$ consisting of a state space S and a set of observables $\mathcal{O} \subseteq \mathcal{P}(S)$, both at most countable.



Learning: Streams of Observables

Definition

Let $\mathbb{S} = (S, \mathcal{O})$ be an epistemic space.

- A **data stream** is an infinite sequence $\vec{O} = (O_0, O_1, \ldots)$ of data from $\mathcal{O}$.
- A **data sequence** is a finite initial segment of an $\vec{O}$;
  such a finite sequence of length $n + 1$ is denoted by $\vec{O}[n] = (O_0, \ldots, O_n)$.

Definition

Take $\mathbb{S} = (S, \mathcal{O})$ and $s \in S$. A data stream $\vec{O}$ is:

- **sound with respect to** s iff every element listed in $\vec{O}$ is true in s.
- **complete with respect to** s iff every observable true in s is listed in $\vec{O}$.

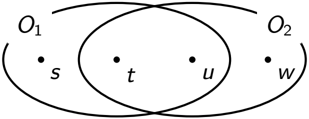
We assume that data streams are sound and complete.

Learning: Learners and Conjectures

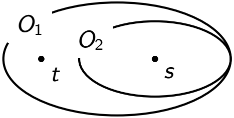
Definition

Let $\mathbb{S} = (S, \mathcal{O})$ be an epistemic space and let $O_0, \ldots, O_n \in \mathcal{O}$.
A **learner** is a function L that on the input of $\mathbb{S}$ and data sequence $(O_0, \ldots, O_n)$ outputs some set of worlds $L(\mathbb{S}, (O_0, \ldots, O_n)) \subseteq S$, called a **conjecture**.

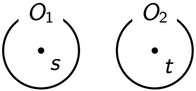
An intuition about separability by observations



(A) t and u are not separable



(B) weakly separated space $T0$



(C) strongly separated space $T1$

Learnability

Definition

$\mathbb{S} = (S, \mathcal{O})$ is **learnable by** L if for every state $s \in S$

Learnability

Definition

$\mathbb{S} = (S, \mathcal{O})$ is **learnable by** L if for every state $s \in S$ and for every sound and complete data stream $\vec{O}$ for s,

Learnability

Definition

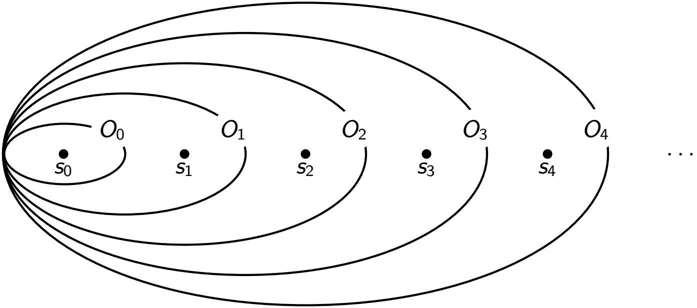
$\mathbb{S} = (S, \mathcal{O})$ is **learnable by** L if for every state $s \in S$ and for every sound and complete data stream $\vec{O}$ for s, there is $n \in \mathbb{N}$ such that:

$$L(\mathbb{S}, \vec{O}[k]) = \{s\} \text{ for all } k \geq n.$$

An epistemic space $\mathbb{S}$ is **learnable** if it is learnable by a learner L.

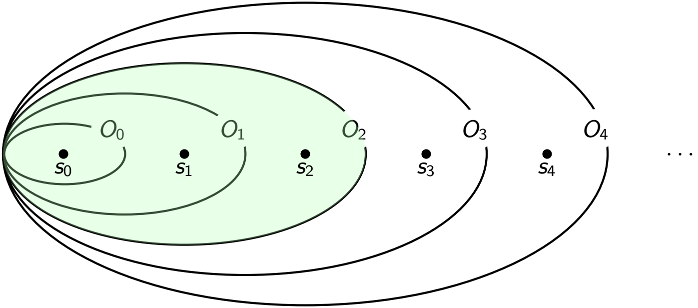
Example of a Learnable Space

Let $\mathbb{S} = (S, \mathcal{O})$ such that $S = \{s_n \mid n \in \mathbb{N}\}$, $\mathcal{O} = \{O_i \mid i \in \mathbb{N}\}$, and for any $k \in \mathbb{N}$, $O_k = \{s_i \mid 0 \leq i \leq k\}$. $\mathbb{S}$ is learnable.



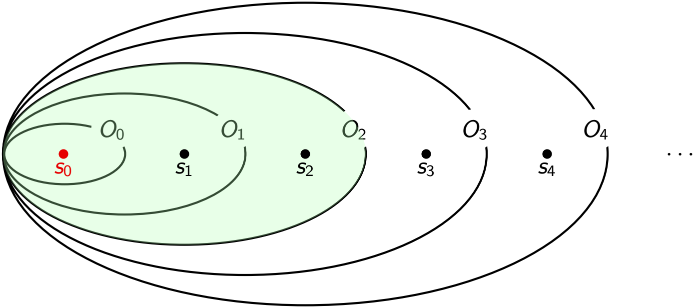
Example of a Learnable Space

Let $\mathbb{S} = (S, \mathcal{O})$ such that $S = \{s_n \mid n \in \mathbb{N}\}$, $\mathcal{O} = \{O_i \mid i \in \mathbb{N}\}$, and for any $k \in \mathbb{N}$, $O_k = \{s_i \mid 0 \leq i \leq k\}$. $\mathbb{S}$ is learnable.



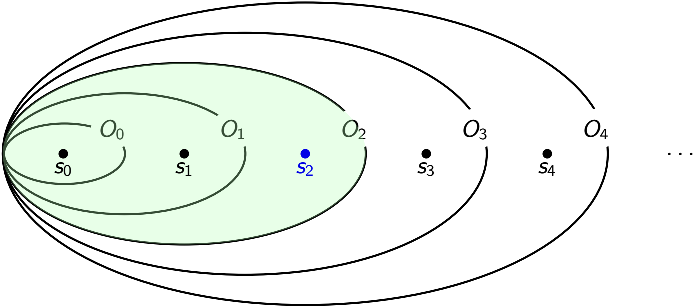
Example of a Learnable Space

Let $\mathbb{S} = (S, \mathcal{O})$ such that $S = \{s_n \mid n \in \mathbb{N}\}$, $\mathcal{O} = \{O_i \mid i \in \mathbb{N}\}$, and for any $k \in \mathbb{N}$, $O_k = \{s_i \mid 0 \leq i \leq k\}$. $\mathbb{S}$ is learnable.



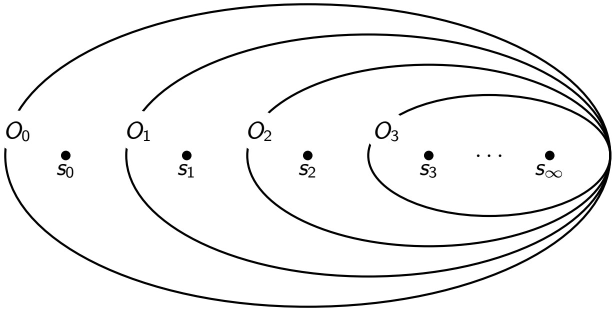
Example of a Learnable Space

Let $\mathbb{S} = (S, \mathcal{O})$ such that $S = \{s_n \mid n \in \mathbb{N}\}$, $\mathcal{O} = \{O_i \mid i \in \mathbb{N}\}$, and for any $k \in \mathbb{N}$, $O_k = \{s_i \mid 0 \leq i \leq k\}$. $\mathbb{S}$ is learnable.



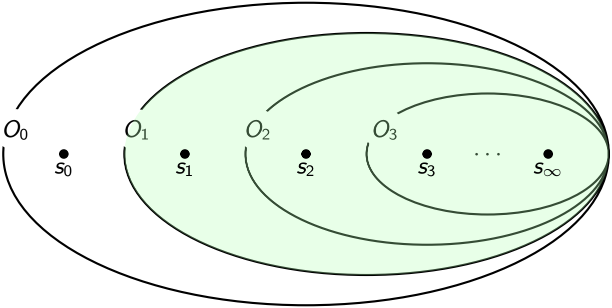
Example of a Non-Learnable Space

Consider $\mathbb{S} = (S, \mathcal{O})$, where $S := \{s_n \mid n \in \mathbb{N}\} \cup \{s_\infty\}$, and $\mathcal{O} = \{O_i \mid i \in \mathbb{N}\}$, and for any $k \in \mathbb{N}$, $O_k := \{s_k, s_{k+1}, \ldots\} \cup \{s_\infty\}$. $\mathbb{S}$ is not learnable.



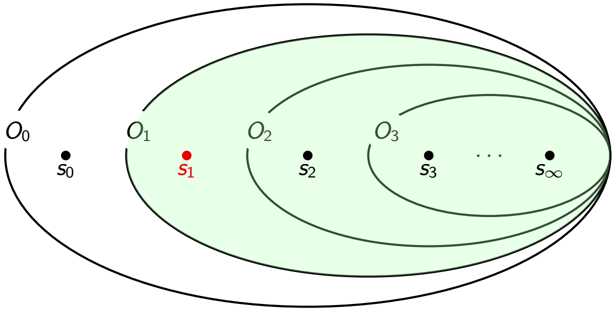
Example of a Non-Learnable Space

Consider $\mathbb{S} = (S, \mathcal{O})$, where $S := \{s_n \mid n \in \mathbb{N}\} \cup \{s_\infty\}$, and $\mathcal{O} = \{O_i \mid i \in \mathbb{N}\}$, and for any $k \in \mathbb{N}$, $O_k := \{s_k, s_{k+1}, \dots\} \cup \{s_\infty\}$. $\mathbb{S}$ is not learnable.



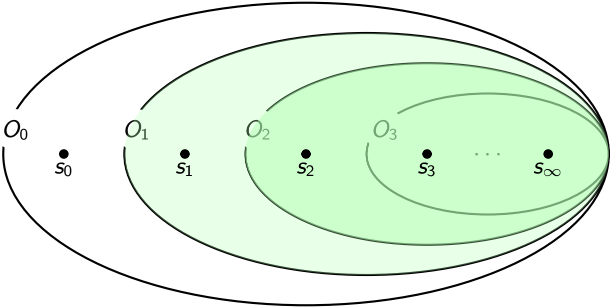
Example of a Non-Learnable Space

Consider $\mathbb{S} = (S, \mathcal{O})$, where $S := \{s_n \mid n \in \mathbb{N}\} \cup \{s_\infty\}$, and $\mathcal{O} = \{O_i \mid i \in \mathbb{N}\}$, and for any $k \in \mathbb{N}$, $O_k := \{s_k, s_{k+1}, \ldots\} \cup \{s_\infty\}$. $\mathbb{S}$ is not learnable.



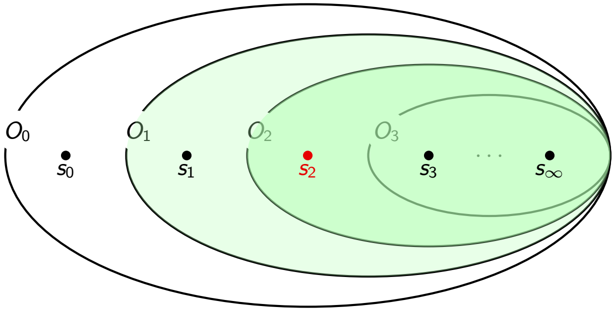
Example of a Non-Learnable Space

Consider $\mathbb{S} = (S, \mathcal{O})$, where $S := \{s_n \mid n \in \mathbb{N}\} \cup \{s_\infty\}$, and $\mathcal{O} = \{O_i \mid i \in \mathbb{N}\}$, and for any $k \in \mathbb{N}$, $O_k := \{s_k, s_{k+1}, \ldots\} \cup \{s_\infty\}$. $\mathbb{S}$ is not learnable.



Example of a Non-Learnable Space

Consider $\mathbb{S} = (S, \mathcal{O})$, where $S := \{s_n \mid n \in \mathbb{N}\} \cup \{s_\infty\}$, and $\mathcal{O} = \{O_i \mid i \in \mathbb{N}\}$, and for any $k \in \mathbb{N}$, $O_k := \{s_k, s_{k+1}, \ldots\} \cup \{s_\infty\}$. $\mathbb{S}$ is not learnable.



Outline

Introduction

Learnability in Epistemic Spaces

Constructive, Order-driven Learning: Belief Revision

Bias and Belief Revision

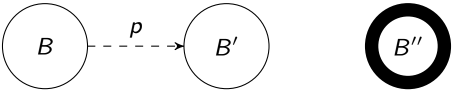
Background

- Learning and belief revision go their separate ways,
- conjecture dynamics is a common theme.
- What are the principles of this dynamics?

$$B \xrightarrow{p} B'$$

Background

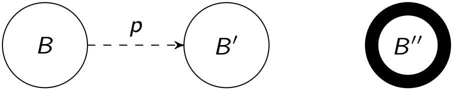
- Learning and belief revision go their separate ways,
- conjecture dynamics is a common theme.
- What are the principles of this dynamics?



Truth-tracking!

Background

- Learning and belief revision go their separate ways,
- conjecture dynamics is a common theme.
- What are the principles of this dynamics?



Truth-tracking!

Order-driven learning: Motivation

- Belief Revision: minimal states give beliefs.
- Computational Learning Theory: co-learning, learning by erasing.
- Philosophy of Science: Ockham's razor.

Plausibility Spaces

A **plausibility space**, $\mathbb{B}_{\mathbb{S}} = (S, \mathcal{O}, \preceq)$, consists of an epistemic space $\mathbb{S} = (S, \mathcal{O})$ and a plausibility preorder $\preceq \subseteq S \times S$.

Plausibility Spaces

A **plausibility space**, $\mathbb{B}_{\mathbb{S}} = (S, \mathcal{O}, \preceq)$, consists of an epistemic space $\mathbb{S} = (S, \mathcal{O})$ and a plausibility preorder $\preceq \subseteq S \times S$.

Knowledge and Belief

$$\begin{array}{lll} \mathbb{B}_{\mathbb{S}} \models Kp & \text{iff} & S \subseteq p \\ \mathbb{B}_{\mathbb{S}} \models Bp & \text{iff} & min_{\preceq} S \subseteq p. \end{array}$$

For non-well-founded spaces we generalise to:

$$\mathbb{B}_{\mathbb{S}} \models Bp \quad \text{iff} \quad \exists w \, \forall u \preceq w \;\; u \in p.$$

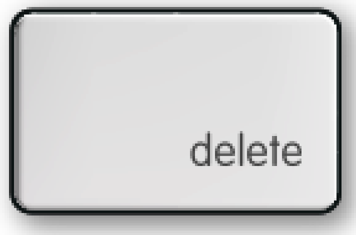
Definition

A **belief-revision method** is a function R that, for any plausibility space $\mathbb{B}_{\mathbb{S}} = (S, \mathcal{O}, \preceq)$ and any observation O outputs a new plausibility space:

$$R(\mathbb{B}_{\mathbb{S}}, O) := (S', \mathcal{O}, \preceq').$$

A belief revision R can be iterated in the following way:
$R(\mathbb{B}_{\mathbb{S}}, \sigma * O) := R(R(\mathbb{B}_{\mathbb{S}}, \sigma), O)$, where σ is a finite sequence of observations.

- **Conditioning** eliminates all worlds of S that do not satisfy the observation.



Lexicographic upgrade

- **Lexicographic upgrade** rearranges the preorder by putting all worlds satisfying the observation to be more plausible than others.

Minimal upgrade

- **Minimal upgrade** rearranges the preorder by making only the most plausible states satisfying the observation more plausible than all others, leaving the rest of the preorder the same.

Belief revision methods

A *one-step revision method* is a function R_1, which on the input of a plausibility space and an observation outputs a new, revised, plausibility space.

Definition

- $\mathrm{Cond}_1((S, \mathcal{O}, \preceq), O) = (S', \mathcal{O}^{S'}, \preceq^{S'})$, where $S' = S \cap O$, $\mathcal{O}^{S'} = \{O \cap S' \mid O \in \mathcal{O}\}$ and $\preceq^{S'} = \preceq \cap (S' \times S')$;
- $\mathrm{Lex}_1((S, \mathcal{O}, \preceq), O) = (S, \mathcal{O}, \preceq')$, where $(s, t) \in \preceq'$ if one of the following conditions holds: $(s, t) \in \preceq^{O}$, $(s, t) \in \preceq^{\overline{O}}$ or $(s \in O$ and $t \notin O)$;
- $\mathrm{Mini}_1((S, \mathcal{O}, \preceq), O) = (S, \mathcal{O}, \preceq')$, where $\preceq'$ is s.t.: if $s \in min(\preceq)$and $t \notin min(\preceq)$ then $(s, t) \in \preceq'$, and $(s, t) \in \preceq'$ iff $(s, t) \in \preceq$, otherwise.

We iterate revision methods: $R(\mathbb{B}, \sigma * O) := R_1(R(\mathbb{B}, \sigma), O)$, for $R \in \{\mathrm{Cond}, \mathrm{Lex}, \mathrm{Mini}\}$.

Learning via Belief Revision

Definition

Every belief-revision method R, together with a prior plausibility $\preceq$ generates in a canonical way a learning method $L_R^{\preceq}$ called a **belief-revision-based learning method**, and given by:

$$L_R^{\preceq}((S, \mathcal{O}), \sigma) := \min_{\preceq} R((S, \mathcal{O}, \preceq), \sigma).$$

Definition

An epistemic space $\mathbb{S}$ is **learnable by a belief-revision method** R if there exists a prior plausibility assignment $\preceq$ such that $L_R^{\preceq}$ learns $\mathbb{S}$.

A. Baltag, N. Gierasimczuk, S. Smets. Truth tracking by belief revision. Studia Logica 2018.

Universality Results

Definition

L is **universal** if it can learn every epistemic space that is learnable.

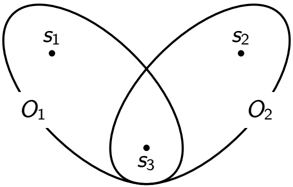
	Conditioning	Lexicographic	Minimal
Positive Streams	YES	YES	NO

Universality Results

Definition
L is **universal** if it can learn every epistemic space that is learnable.

	Conditioning	Lexicographic	Minimal
Positive Streams	YES	YES	NO

Theorem
Minimal revision is not universal.

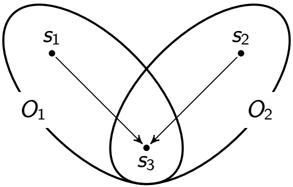


Universality Results

Definition
L is **universal** if it can learn every epistemic space that is learnable.

	Conditioning	Lexicographic	Minimal
Positive Streams	YES	YES	NO

Theorem
Minimal revision is not universal.

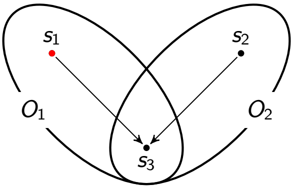


Universality Results

Definition
L is **universal** if it can learn every epistemic space that is learnable.

	Conditioning	Lexicographic	Minimal
Positive Streams	YES	YES	NO

Theorem
Minimal revision is not universal.

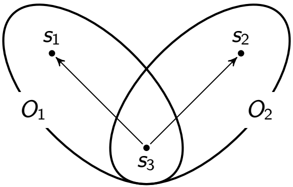


Universality Results

Definition
L is **universal** if it can learn every epistemic space that is learnable.

	Conditioning	Lexicographic	Minimal
Positive Streams	YES	YES	NO

Theorem
Minimal revision is not universal.



Universality Results

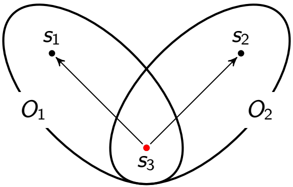
Definition

L is **universal** if it can learn every epistemic space that is learnable.

	Conditioning	Lexicographic	Minimal
Positive Streams	YES	YES	NO

Theorem

Minimal revision is not universal.



Universality Results

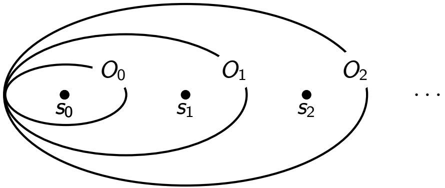
Definition

L is **universal** if it can learn every epistemic space that is learnable.

	Conditioning	Lexicographic	Minimal
Positive Streams	YES	YES	NO

Theorem

There is no universal belief-revision method under well-foundedness.



Universality Results

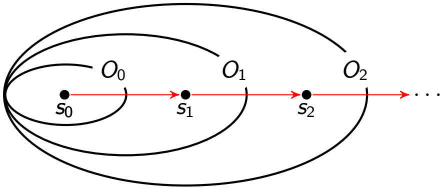
Definition

L is **universal** if it can learn every epistemic space that is learnable.

	Conditioning	Lexicographic	Minimal
Positive Streams	YES	YES	NO

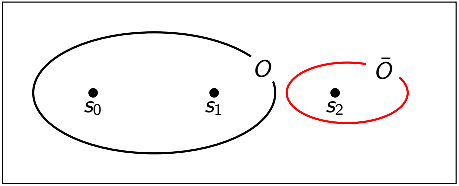
Theorem

There is no universal belief-revision method under well-foundedness.



Is $\neg O$ observable?

An epistemic space $\mathbb{S} = (S, \mathcal{O})$ is **negation-closed** iff if $O \in \mathcal{O}$, then $\bar{O} \in \mathcal{O}$.

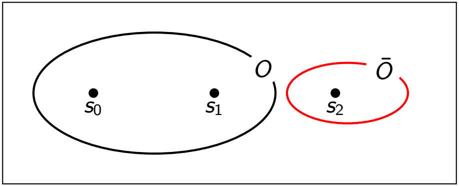


Definition

Let $\mathbb{S} = (S, \mathcal{O})$ be a negation-closed epistemic space. A stream $\vec{O}$ is **fair** with respect to the world s if $\vec{O}$ is complete wrt to s, and contains only finitely many observations O, s.t. $s \notin O$ and every such error is eventually corrected in $\vec{O}$.

Is $\neg O$ observable?

An epistemic space $\mathbb{S} = (S, \mathcal{O})$ is **negation-closed** iff if $O \in \mathcal{O}$, then $\bar{O} \in \mathcal{O}$.



Definition

Let $\mathbb{S} = (S, \mathcal{O})$ be a negation-closed epistemic space. A stream $\vec{O}$ is **fair** with respect to the world s if $\vec{O}$ is complete wrt to s, and contains only finitely many observations O, s.t. $s \notin O$ and every such error is eventually corrected in $\vec{O}$.

	Conditioning	Lexicographic	Minimal
Positive and Negative	YES	YES	NO
Fair Streams	NO	YES	NO

Outline

Types of bias

We combine revision methods with three kinds of ‘cognitive’ limitations:

- issues with accepting the input;
- issues with perceiving the input;
- issues with access to belief state.

We investigate (theoretically and practically) their impact on truth-tracking.

The procedure

After the random generation of an epistemic space, one of the states, s, is randomly designated to be the actual world, and a sound and complete data sequence σ for s is generated. Then a plausibility preorder is randomly generated and each of the (biased) revision methods is called to identify s on σ.

The simulation

Each series of tests consisted of 200 runs, the plausibility spaces consisted of 5 possible worlds and 12 observables, and the incoming data sequence was always longer than the number of observables.

ISSUES WITH ACCEPTING THE INPUT (CONFIRMATION BIAS)

DEFINITION
Given an $(S, \mathcal{O})$, the *stubbornness function* is $D : \mathcal{P}(S) \to \mathbb{N}$.

DEFINITION
R_{CB} is defined in the following way:

$$R_{CB}(\mathbb{B}, \lambda) = \mathbb{B},$$

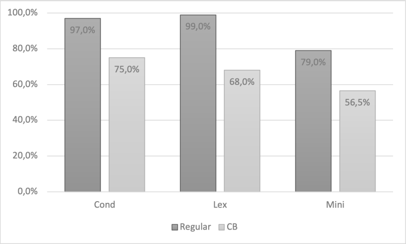
$$R_{CB}(\mathbb{B}, \sigma \cdot p) = \begin{cases} R_1(R_{CB}(\mathbb{B}, \sigma), p) & \text{if} \quad \#p(\sigma) \geq D(\overline{p}), \\ R_{CB}(\mathbb{B}, \sigma) & \text{otherwise.} \end{cases}$$

We obtain $Cond_{CB}$, Lex_{CB}, $Mini_{CB}$.

Results

Proposition
Cond, Lex, Mini are strictly more powerful than $Cond_{CB}$, Lex_{CB}, $Mini_{CB}$.

Corollary
$Cond_{CB}$ and Lex_{CB} are not universal.



Issues with perceiving the input (framing bias)

Definition

Given $(S, \mathcal{O})$, the *framing function* is $FR : \mathcal{O} \to \mathcal{P}(S)$.

Definition (Framing-bias methods)

We define a framing-biased method in the following way:

$$R_{FR}(\mathbb{B}, \lambda) = \mathbb{B},$$

$$R_{FR}(\mathbb{B}, \sigma \cdot p) = R_1(R_{FR}(\mathbb{B}, \sigma), x), \text{ such that } x \in FR(p).$$

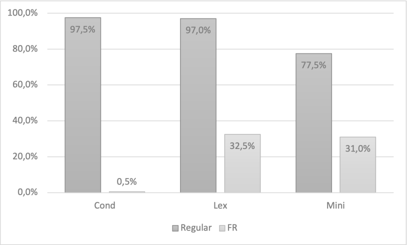
We obtain $Cond_{FR}$, Lex_{FR}, $Mini_{FR}$.

Results

Proposition
$Cond_{FR}$ and Lex_{FR} are not universal.

Proposition
Mini is strictly more powerful than $Mini_{FR}$.

Proposition
Lex_{FR} is universal on fairly framed streams.



Issues with access to belief state (anchoring bias)

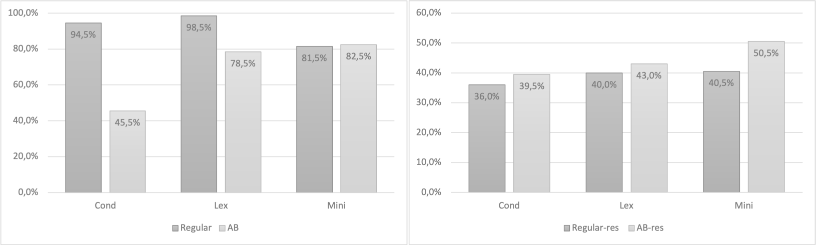
Definition (Anchoring-bias methods)

We define anchoring-biased method in the following way: after receiving p they 'choose' the best world satisfying p. If there are minimal several worlds that are equi-plausible, the method picks one at random.

Additionally, a **resource parameter** (a real number between 0 and 100) halves each time a revision takes place, and the process terminates when the resource is depleted (< 1).

Results

Proposition
$Cond_{AB}$, Lex_{AB} are not universal.



Anchoring ability to select a random world as the candidate for the actual world improves the truth-tracking capability, especially in the case of minimal revision.