Modal Undefinability and Ultrafilter Extensions for IL
(Advances in Modal Logic 2026)

Vicent Navarro Arroyo
Co-authored with Joost J. Joosten,
Cosimo Perini Brogi and Fèlix Frigola González

University of Barcelona

29/6/2026

Outline

Main Goal of the Talk

This talk has two interconnected parts.

Part I

- Introduce a first-order frame condition for interpretability logic.
- Show that the frame condition is *not modally definable*.

Part II

- Develop ultrafilter extensions for Veltman semantics.
- Provide the algebraic machinery needed toward a future Goldblatt–Thomason theorem for interpretability logic.

Interpretability Logic

Interpretability logics were introduced by Visser as an extension of provability logic.

Language of IL**:**

$$\mathbf{Form}_{\mathsf{IL}} ::= p \mid \bot \mid (\varphi \to \psi) \mid \Box\varphi \mid (\varphi \rhd \psi)$$

We keep in mind the **informal reading** (T is a theory):

- $\Box\varphi$: 'φ is provable in T."
- $\varphi \rhd \psi$: '$T + \varphi$ interprets $T + \psi$."

Interpretability logic captures structural properties of formal theories far beyond ordinary provability logic.

Why Interpretability Logic?

Unlike provability logic, different arithmetical theories may induce different interpretability logics.

Examples:

- IL $\longrightarrow$ basic interpretability logic.
- ILM $\longrightarrow$ logic of theories with full induction.
- ILP $\longrightarrow$ logic of finitely axiomatizable theories proving the totality of super-exponentiation.

There is a long-standing **open problem**:

Open Question

What is the interpretability logic of all reasonable arithmetical theories? (IL(All))

Frame conditions are one of the main tools for attacking this problem.

Outline

The Logic IL

IL extends classical propositional logic by:

- **L1** $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$
- **L2** $\Box\varphi \to \Box\Box\varphi$
- **L3** $\Box(\Box\varphi \to \varphi) \to \Box\varphi$

- **J1** $\Box(\alpha \to \beta) \to (\alpha \rhd \beta)$
- **J2** $(\alpha \rhd \beta) \wedge (\beta \rhd \gamma) \to (\alpha \rhd \gamma)$
- **J3** $(\alpha \rhd \gamma) \wedge (\beta \rhd \gamma) \to (\alpha \vee \beta) \rhd \gamma$
- **J4** $(\alpha \rhd \beta) \to (\Diamond\alpha \to \Diamond\beta)$
- **J5** $\Diamond\alpha \rhd \alpha$

together with Modus Ponens and Necessitation Rule:

$$\text{If } \vdash_{\mathsf{IL}} \varphi, \text{ then } \vdash_{\mathsf{IL}} \Box\varphi. \qquad \text{(Nec.)}$$

A **Veltman frame** is a structure

$$\mathcal{F} = \langle W, R, \{S_w\}_{w \in W} \rangle$$

such that

- R is transitive;
- R is conversely well-founded;
- every S_w is reflexive;
- every S_w is transitive;
- $R \cap R[w] \subseteq S_w$.

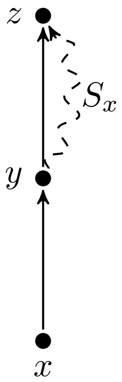


Figure: $R \cap R[w] \subseteq S_w$.

Interpretability requires a family of relations $\{S_w\}_{w \in W}$ rather than a single accessibility relation.

The **interpretability forcing** clause:

$$w \Vdash \varphi \rhd \psi :\iff \forall u\Big(wRu \wedge u \Vdash \varphi \Rightarrow \exists v\,(uS_w v \wedge v \Vdash \psi)\Big).$$

$$\underset{x}{\bullet}\, {\color{red} A \rhd B}$$

This additional structure makes correspondence theory considerably harder than in ordinary modal logic.

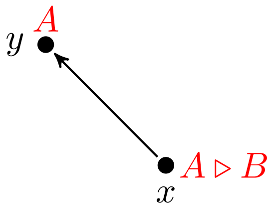
The **forcing on the whole Veltman frame** $\mathcal{F} := \langle W, R, \{S_w\}_{w\in W}\rangle$ is defined as:

$$\mathcal{F} \vDash \varphi :\iff \forall w \in W\, w \Vdash \varphi.$$

Semantics of $\rhd$

The **interpretability forcing** clause:

$$w \Vdash \varphi \rhd \psi :\iff \forall u\Big(wRu \wedge u \Vdash \varphi \Rightarrow \exists v\,(uS_w v \wedge v \Vdash \psi)\Big).$$



This additional structure makes correspondence theory considerably harder than in ordinary modal logic.

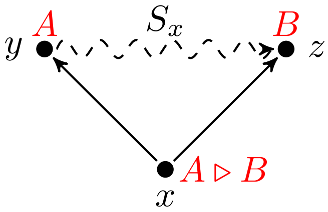
The **forcing on the whole Veltman frame** $\mathcal{F} := \langle W, R, \{S_w\}_{w \in W}\rangle$ is defined as:

$$\mathcal{F} \vDash \varphi :\iff \forall w \in W\, w \Vdash \varphi.$$

Semantics of $\rhd$

The **interpretability forcing** clause:

$$w \Vdash \varphi \rhd \psi :\iff \forall u \Big(wRu \land u \Vdash \varphi \Rightarrow \exists v \left(uS_w v \land v \Vdash \psi \right) \Big).$$



This additional structure makes correspondence theory considerably harder than in ordinary modal logic.

The **forcing on the whole Veltman frame** $\mathcal{F} := \langle W, R, \{S_w\}_{w \in W} \rangle$ is defined as:

$$\mathcal{F} \vDash \varphi :\iff \forall w \in W\, w \Vdash \varphi.$$

Important Extensions of IL

There are two well-known interpretability logics:

ILM: (Montagna's principle)

$$\varphi \rhd \psi \to (\varphi \wedge \Box\phi) \rhd (\psi \wedge \Box\phi)$$

whose **frame condition** is:

$$xRyS_xzRu \to yRu.$$

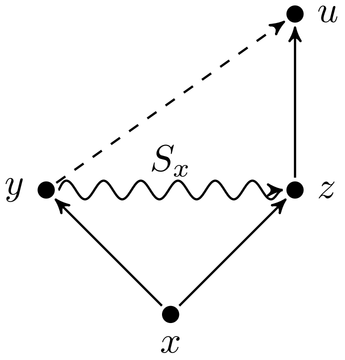


Figure: ILM frame condition.

Important Extensions of IL

ILP: (Persistence principle)

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

whose **frame condition** is:

$$xRyRzS_xu \to zS_yu.$$

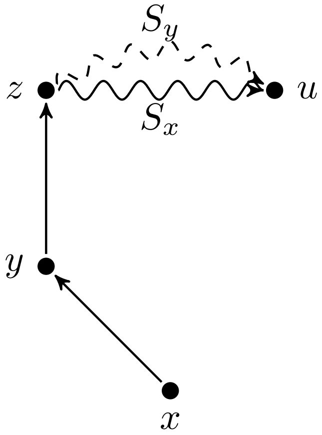


Figure: ILP frame condition.

Outline

The Pencil Frame Condition

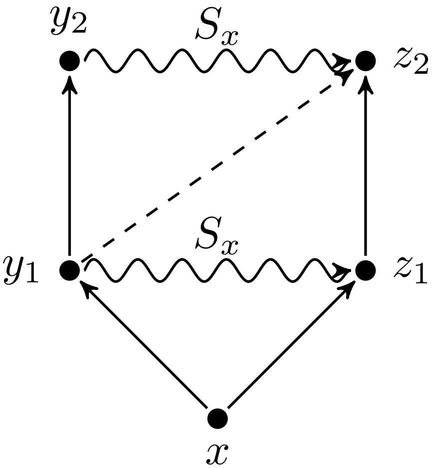


Figure: The frame condition $\mathcal{C}_{\mathrm{Pencil}}$

The Pencil Frame Condition

We studied the *Pencil* frame property

$$xRyS_xzRu \wedge yRvS_xu \to yRu.$$

Let $\mathcal{C}_{\mathrm{Pencil}}$ be the class of Veltman frames satisfying this condition.

Key observations:

- Every ILM frame satisfies it.
- Every ILP frame satisfies it.
- It is a natural candidate for $\mathsf{IL}(\mathsf{All})$.

Main Result I

Theorem

The Pencil Frame Condition is not modally definable.

Equivalently,

$$\neg\exists\varphi \in \mathbf{Form}_{\mathsf{IL}} : \forall\mathcal{F}\Big(\mathcal{F} \models \varphi \iff \mathcal{F} \in \mathcal{C}_{\mathrm{Pencil}}\Big).$$

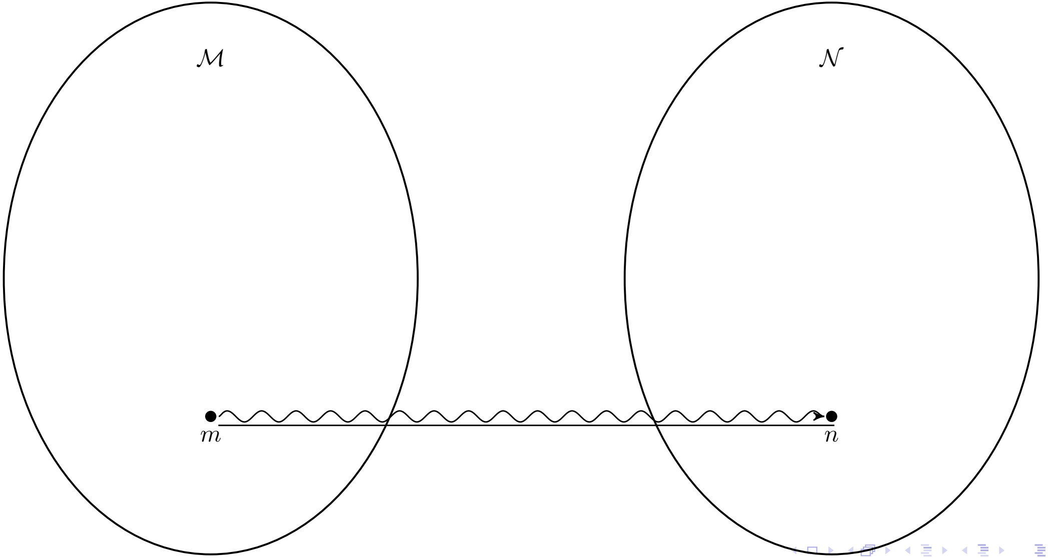
Question: Can Goldblatt-Thomason methods be adapted to interpretability logic?

The proof uses **Veltman bisimulations** (Visser 1990).

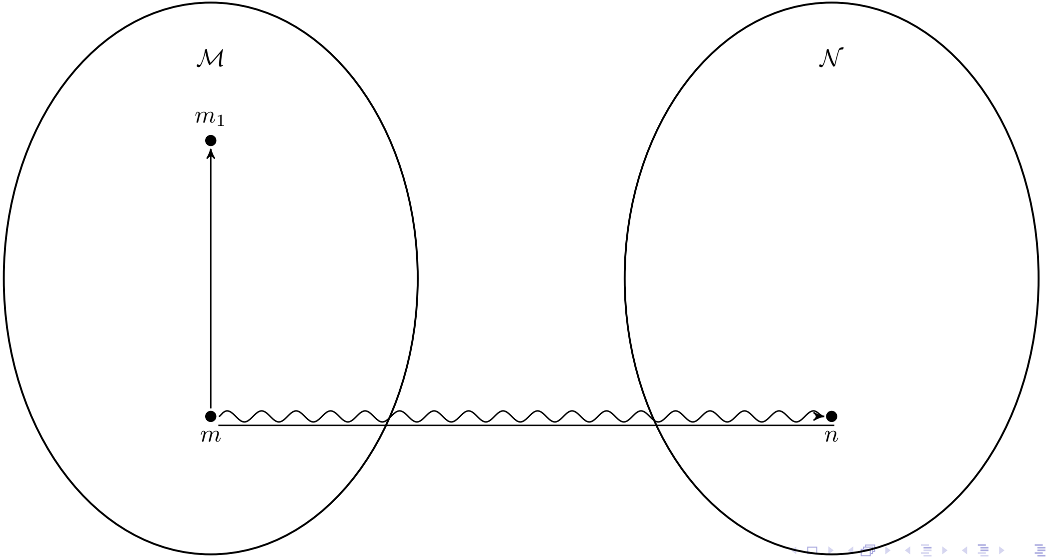
A Veltman bisimulation relation Z between the worlds of two models $\langle W_{\mathcal{M}}, R^{\mathcal{M}}, \{S_w^{\mathcal{M}}\}_{w \in W_{\mathcal{M}}} \rangle$ and $\langle W_{\mathcal{N}}, R^{\mathcal{N}}, \{S_w^{\mathcal{N}}\}_{w \in W_{\mathcal{N}}} \rangle$ satisfies:

1. **Preservation of atomic valuations**: $\forall p \in \mathbf{Atom}\ mZn \Rightarrow (\mathcal{M}, m \Vdash p \iff \mathcal{N}, n \Vdash p)$.
2. (**Zig**) If $mR^{\mathcal{M}}m_1$ and mZn, then there is some n_1 such that $nR^{\mathcal{N}}n_1$ and m_1Zn_1 and for all n_2 in $\mathcal{N}$ such that $n_1S_n^{\mathcal{N}}n_2$ there is some m_2 such that m_2Zn_2 and $m_1S_m^{\mathcal{M}}m_2$.
3. (**Zag**) If $nR^{\mathcal{N}}n_1$ and mZn then there is some m_1 such that mRm_1 and m_1Zn_1 and for all m_2 such that $m_1S_m^{\mathcal{M}}m_2$ there is some n_2 such that m_2Zn_2 and $n_1S_nn_2$.

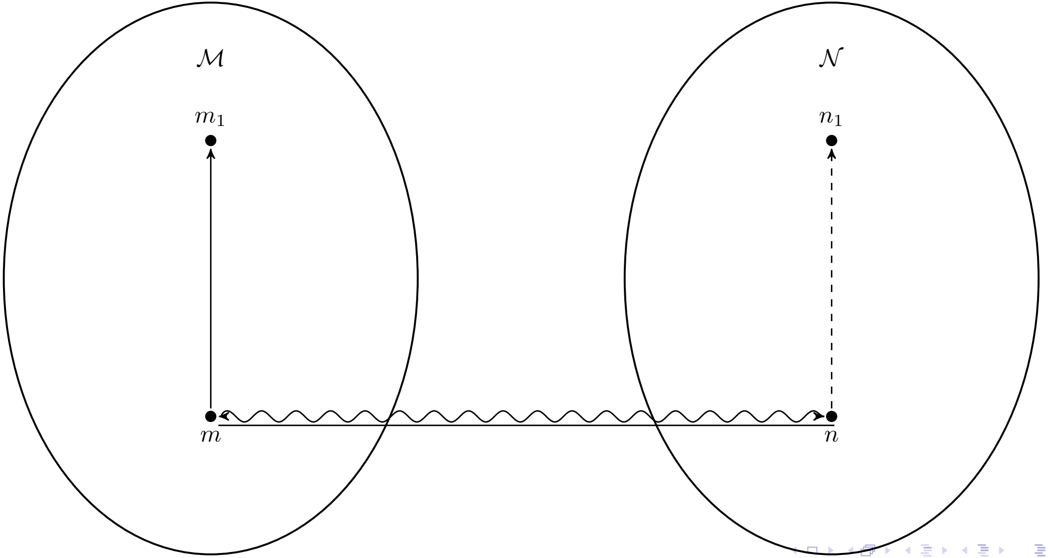
Bisimulations



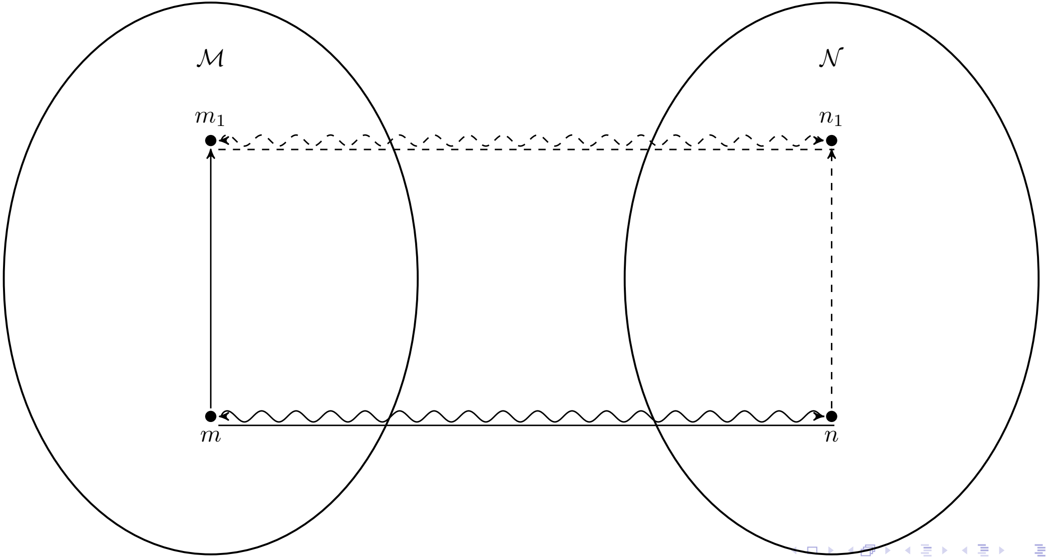
Bisimulations



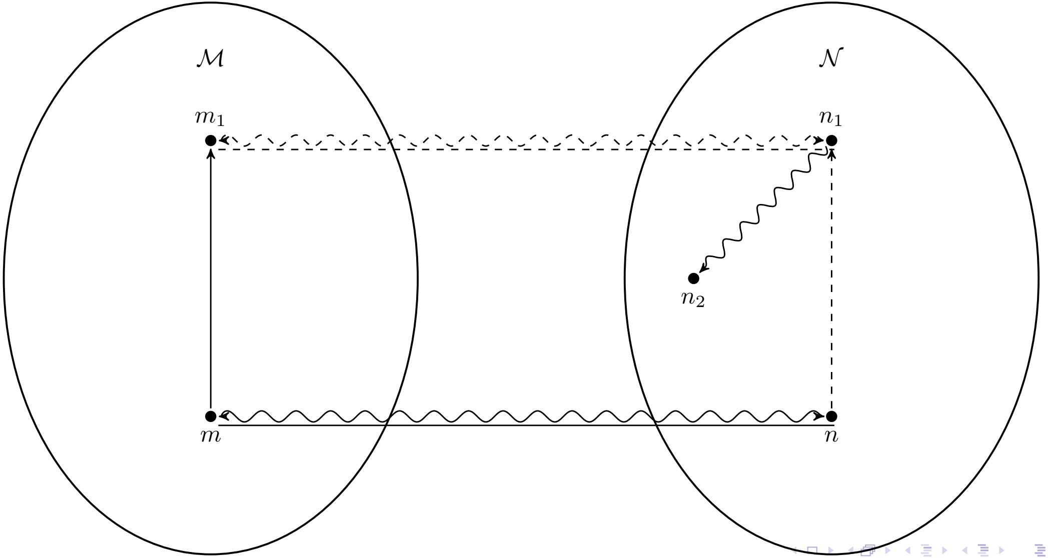
Bisimulations



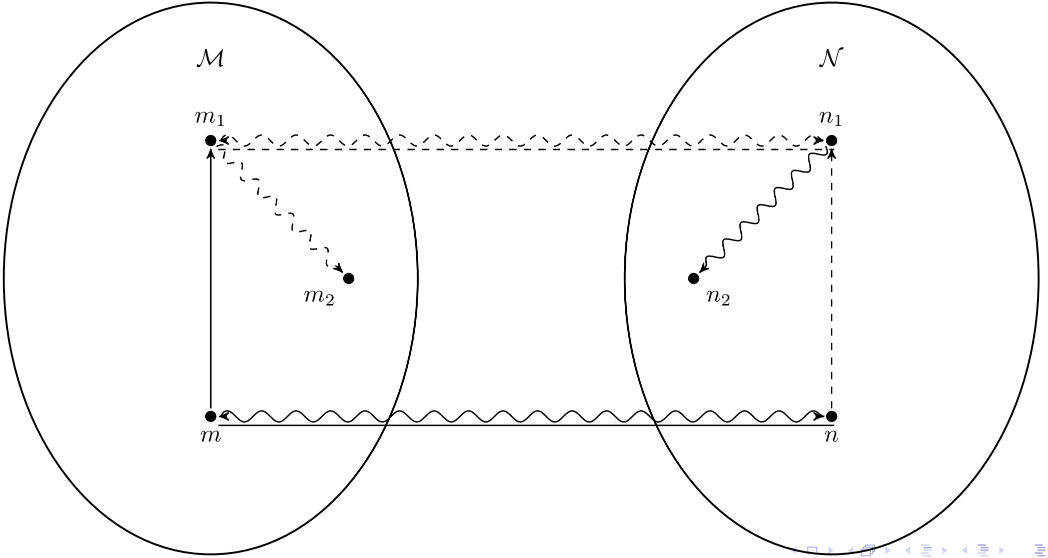
Bisimulations



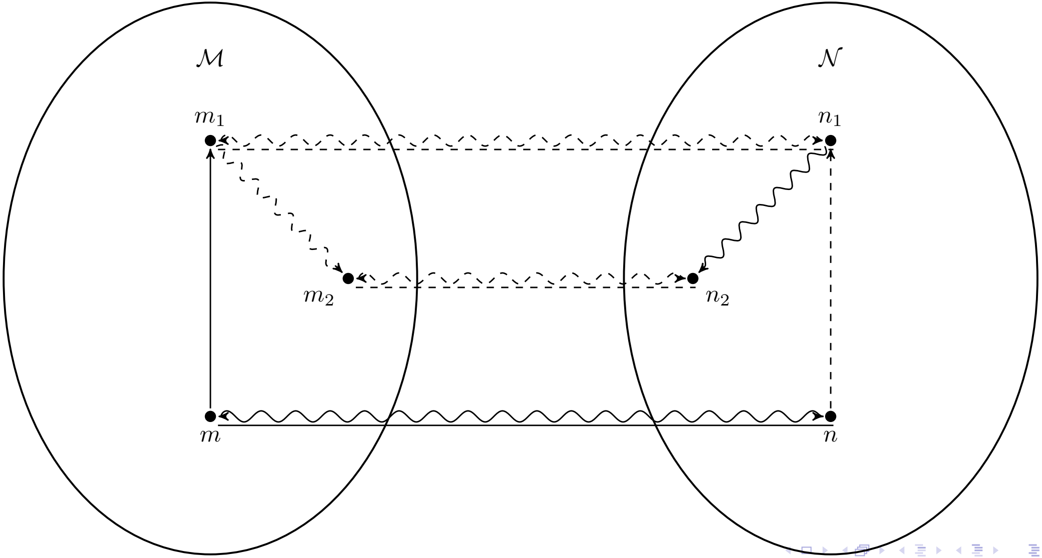
Bisimulations



Bisimulations



Bisimulations



Idea of the Counterexample

Fundamental fact that will be useful:

Modal Invariance

Bisimilar worlds satisfy exactly the same interpretability formulas.

Sketch of the proof...

Idea of the Counterexample

Fundamental fact that will be useful:

Modal Invariance

Bisimilar worlds satisfy exactly the same interpretability formulas.

Sketch of the proof...

Construct two frames that differ by a single S-edge:

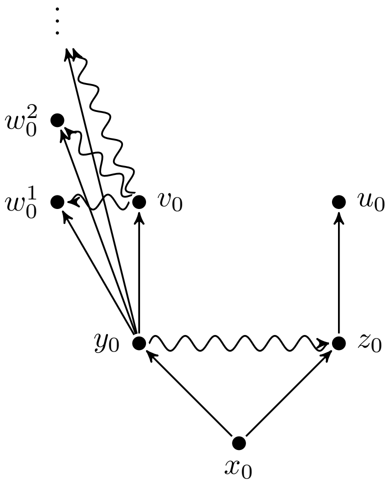
$$\mathcal{F}_0 \qquad \mathcal{F}_1$$

Some facts that we ought make sure:

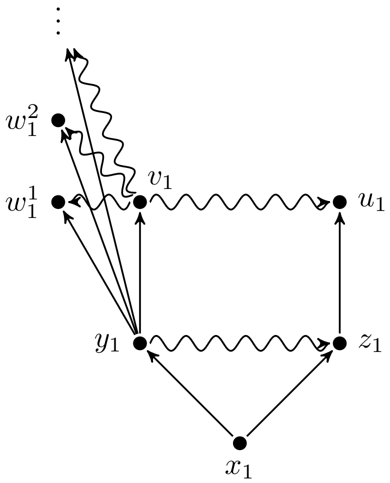
- $\mathcal{F}_0$ satisfies the Pencil Frame Condition.
- $\mathcal{F}_1$ violates the Pencil Frame Condition.
- Appropriate valuations make the resulting models bisimilar.

Therefore no modal formula can distinguish the two classes.

Bisimulation Argument



(a) $\mathcal{F}_0$. Only S_{x_0} is displayed.



(b) $\mathcal{F}_1$. Only S_{x_1} is displayed.

Proof Strategy

Assume by *Reductio ad Absurdum* that

$$\exists \varphi \in \mathbf{Form}_{\mathsf{IL}} : \mathcal{F} \models \varphi \iff \mathcal{F} \in \mathcal{C}_{\text{Pencil}}.$$

Observe that $\mathcal{F}_0 \vDash \varphi$ and $\mathcal{F}_1 \nvDash \varphi$.

We proved the following lemma:

Lemma

Given a valuation ev_1 *on* $\mathcal{F}_1$*, there is another valuation* ev_0 *on* $\mathcal{F}_0$ *such that*

$$\langle \mathcal{F}_0, \mathrm{ev}_0 \rangle \textit{ is bisimilar with } \langle \mathcal{F}_1, \mathrm{ev}_1 \rangle,$$

By **Modal Invariance** we reach a contradiction!

Hence the Pencil Frame Condition is not modally definable.

Outline

Motivation

In ordinary modal logic we have:

Theorem (Goldblatt–Thomason)

*A first-order definable class $\mathcal{C}$ of frames is **definable by a set of modal formulas** if and only if it is closed under taking bounded morphic images, generated subframes, disjoint unions and **reflects ultrafilter extensions**.*

The Pencil Frame Condition is **first-order** (over **second-order frame class**) and **not modaly definable**.

There is a natural question:

Goal

Develop ultrafilter extensions for Veltman semantics as a first step toward a Goldblatt–Thomason theorem.

Ultrafilters

A **filter** f on W is a set $f \subseteq \mathcal{P}(W)$ that satisfies:

1. **properness**: $\varnothing \notin f$.
2. **upward closure**:
$$\text{If } A \in f \text{ and } A \subseteq B, \text{ then } B \in f.$$
3. **finite intersection closure**:
$$\text{If } A, B \in f, \text{ then } A \cap B \in f.$$

A filter f over W is an **ultrafilter** ($f \in \mathrm{U}(W)$) if it satisfies:

1. **totality**:
$$A \in f \quad \text{or} \quad W \setminus A \in f.$$

Principal ultrafilter: $\Pi_w := \{A \subseteq W : w \in A\}$.

(Ultra)filters are the algebraic counterpart of **(complete) theories**.

Algebraic Semantics

We have that

$$R^{-1}(\llbracket A \rrbracket_{\mathcal{M}}) = \llbracket \Diamond A \rrbracket_{\mathcal{M}}$$

For a relation R over some set $W \neq \varnothing$ and $Y \subseteq W$ define

$$\widehat{R}^{-1}(Y) := \{x : \forall y (xRy \longrightarrow y \in Y)\}.$$

This is the algebraic counterpart of ordinary modal accessibility:

$$\widehat{R}^{-1}(\llbracket A \rrbracket_{\mathcal{M}}) = \llbracket \Box A \rrbracket_{\mathcal{M}}$$

The Main New Operator

Interpretability requires an analogue of $\rhd$:

$$S^{-1}(\llbracket A \rrbracket_{\mathcal{M}}, \llbracket B \rrbracket_{\mathcal{M}}) = \llbracket A \rhd B \rrbracket_{\mathcal{M}}.$$

Given a Veltman frame $\mathcal{F} = \langle W, R, \{S_w\}_{w \in W} \rangle$, we define $S^{-1} : \wp(W) \times \wp(W) \longrightarrow \wp(W)$ by

$$S^{-1}(X, Y) := \{ w \in W \mid \forall x \in X . (wRx \Rightarrow \exists y \in Y . xS_w y) \}$$

This is the algebraic counterpart of $\varphi \rhd \psi$.

Assuring Successors

Canonical completeness proofs use *assuring labels*. We introduced a filter version.

Let $I \subseteq \wp(W) \setminus \{\varnothing\}$ have the **FIP** (Finite Intersection Property), and let $f, g \in \mathrm{U}(W)$. We define $f \prec_I g$ as follows:

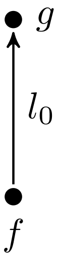
$$f \prec_I g \quad \Longleftrightarrow \quad \left[\forall A \subseteq W \forall^{\text{fin}} \{S_i \in I\} \left(S^{-1}(\overline{A}, \bigcup_i \overline{S_i}) \in f \Rightarrow A, \widehat{R^{-1}} A \in g \right) \right].$$

Intuition:

- I stores label information. ($I \subseteq g$).
- g behaves as a labelled successor of f.
- The construction mirrors assuring successors in canonical models.

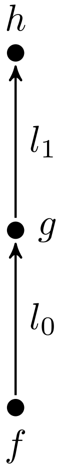
This is the mechanism that allows labels to survive the ultrafilter construction.

Assuring successors



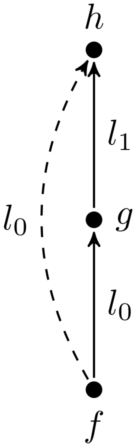
(a) h is l_0-critical

Assuring successors



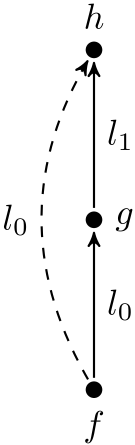
(a) h is l_0-critical

Assuring successors

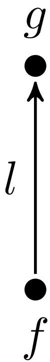


(a) h is l_0-critical

Assuring successors

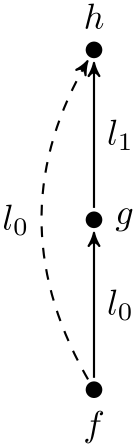


(a) h is l_0-critical

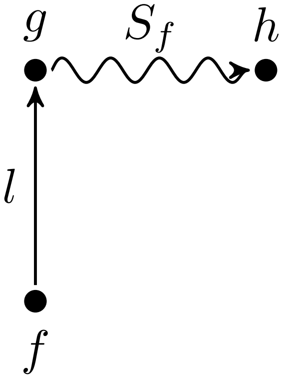


(b) Assuring successors are successors

Assuring successors

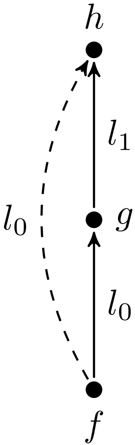


(a) h is l_0-critical

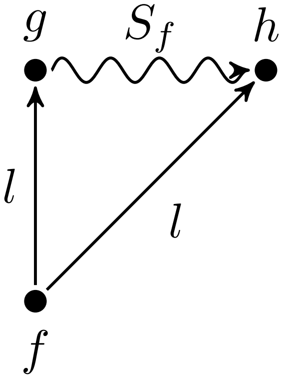


(b) Assuring successors are successors

Assuring successors



(a) h is l_0-critical



(b) Assuring successors are successors

Ultrafilter Extension of a Veltman Frame

The ultrafilter extension of the Veltman frame $\mathcal{F}^{\mathrm{ue}}$ is defined recursively:

1. W^{ue} is recursively defined:
    1. $(f, \langle\rangle) \in W^{\mathrm{ue}}$, for each ultrafilter f over W.
    2. $(f, \sigma) \in W^{\mathrm{ue}} \wedge f \prec_I g \Rightarrow (g, \sigma^\frown\langle I\rangle) \in W^{\mathrm{ue}}$.
2. R^{ue} is defined as the transitive closure of $R^{\mathrm{ue}}_{\mathrm{one}}$ which is defined as follows:

$$\langle f, \sigma\rangle R^{\mathrm{ue}}_{\mathrm{one}} \langle g, \sigma^\frown\langle I\rangle\rangle \iff f \prec_I g$$

Ultrafilter Extension of a Veltman Frame

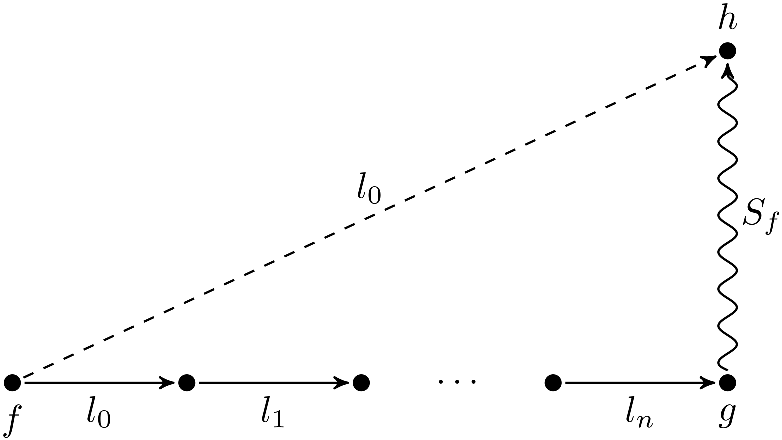


Figure: We need this property.

Ultrafilter Extension of a Veltman Frame

3. $S^{\mathrm{ue}}_{\langle f,\sigma\rangle}$ is the transitive reflexive closure of $R^{\mathrm{ue}}[\langle f,\sigma\rangle]^2 \cap R^{\mathrm{ue}}$ and $S^{\mathrm{ue}}_{\langle f,\sigma\rangle,\mathrm{one}}$:

$$\langle g,\tau\rangle S^{\mathrm{ue}}_{\langle f,\sigma\rangle,\mathrm{one}}\langle h,\tau'\rangle \iff \begin{cases} \langle f,\sigma\rangle R^{\mathrm{ue}}\langle g,\tau\rangle \text{ and} \\ \langle f,\sigma\rangle R^{\mathrm{ue}}\langle h,\tau'\rangle \text{ and} \\ (\tau)_{|\sigma|} = (\tau')_{|\sigma|} \end{cases}$$

The sequence σ keeps track of the label structure needed to define S.

Ultrafilter extension of a Veltman model:
Given a Veltman model $\mathcal{M} = \langle \mathcal{F}, ev\rangle$, we define its ultrafilter extension $\mathcal{M}^{\mathrm{ue}} := \langle \mathcal{F}^{\mathrm{ue}}, ev^{\mathrm{ue}}\rangle$ where ev^{ue} is naturally characterized by:

$$\langle f,\sigma\rangle \in ev^{\mathrm{ue}}(p) \iff ev(p) \in f.$$

Translation of Formulas

We need to define a translation $\varphi \mapsto \varphi^t$ into algebraic operations on subsets of W.

Important examples:

$$p^t := A_p \text{ where } p \in \texttt{Atm} \text{ and } A_p \text{ a set-variable.}$$

$$(\Box\varphi)^t := \widehat{R^{-1}}(\varphi^t).$$

$$(\varphi \rhd \psi)^t := S^{-1}(\varphi^t, \psi^t).$$

This allows modal reasoning to be carried out inside the algebra of subsets of W.

Algebraic Soundness

For a Veltman frame $\mathcal{F} = \langle W, R, \{S_x\}_{x \in W} \rangle$, take a fixed valuation e that assigns subsets of W to the set-variables A_p. This fixes in the obvious way for each formula φ a set that we denote by $[\![\varphi^t]\!]^{\mathsf{e}}_{\mathcal{F}}$.

Theorem

If $\mathsf{IL} \vdash \varphi$*, then for every frame* $\mathcal{F}$ *and for every valuation e,* $[\![\varphi^t]\!]^{\mathsf{e}}_{\mathcal{F}} = W$*.*

Proof idea:

- Verify all IL axioms algebraically.
- Show closure under Modus Ponens.
- Show closure under Necessitation.

This generalizes the familiar algebraic approach from modal logic.

Main Results

Theorem (Elementary equivalence between a frame and its ultrafilter extension.)

Given $\mathcal{M} = \langle \mathcal{F}, ev \rangle$ a Veltman model. For any formula φ and any $\langle \Pi_x, \sigma \rangle$ in the ultrafilter extension we have

$$\mathcal{M}, x \Vdash \varphi \iff \mathcal{M}^{ue}, \langle \Pi_x, \sigma \rangle \Vdash \varphi.$$

- Prove the stronger truth lemma:

$$[\![\psi]\!]_{\mathcal{M}} \in f \iff \mathcal{M}^{\mathrm{ue}}, \langle f, \sigma \rangle \Vdash \psi.$$

- **Key step:** If $f \prec_l g$, $S^{-1}(A, B) \in f$ and $A \in g$, then there is an ultrafilter h such that $f \prec_l h$ and $B \in h$ (same label!).
- Apply the truth lemma to the principal ultrafilter Π_x (notice $A \in \Pi_x \iff x \in A$):

$$[\![\varphi]\!]_{\mathcal{M}} \in \Pi_x \iff x \in [\![\varphi]\!]_{\mathcal{M}} \iff \mathcal{M}, x \Vdash \varphi.$$

Main Results

Definition (Modal saturation)

Let $\mathcal{M}$ be a modal model and Σ a set of modal sentences.

- Σ is **locally possible** iff $\forall x \in \mathcal{M}\ \forall \Sigma' \subseteq_{\text{fin}} \Sigma\ \ x \Vdash \Diamond \Sigma'$.
- Σ is **globally possible** iff $\forall x \in \mathcal{M}\ \exists y \in \mathcal{M}\ xRy \Vdash \Sigma$.
- $\mathcal{M}$ is **modally saturated** iff Σ is locally possible $\Rightarrow$ Σ is globally possible.

Theorem (Modal saturation of ultrafilter extensions)

For any Veltman model $\mathcal{M} = \langle \mathcal{F}, ev \rangle$, the ultrafilter extension $\mathcal{M}^{ue}$ is modally saturated.

Consequences:

- strong model-theoretic behaviour;
- parallels with ordinary modal logic;
- foundations for Goldblatt–Thomason style results.

Outline

Conclusions

Main contributions:

1. Identification of the Pencil Frame Condition.
2. Proof via bisimulation that it is first-order but not modally definable.
3. Development of ultrafilter extensions for Veltman semantics.
4. Algebraic counterparts for labels and assuring successors.
5. Elementary equivalence and modal saturation.

This provides the first major step toward a Goldblatt–Thomason theorem for interpretability logic.

谢谢

Thank you!