

On tame semantics for interpretability logic

Vicent Navarro Arroyo

joint work with

Joost J. Joosten

31st Workshop on Logic, Language, Information and Computation

14 - 17 July 2025

Porto, Portugal

July 15

Department of Philosophy,

University of Barcelona

Motivation

Semantics and intersections

Pencil frame

Dealing with confluence

A new semantics: Minimal Veltman Semantics

Completeness of Minimal Veltman Semantics

Confluence revisited

Motivation

Motivation

In Provability Logic, for a fixed theory T (PL) $\Box A$ reads as

“ A ” is provable in T .

Interpretability Logic (IL) extends PL adding $A \triangleright B$ which means

$T + A$ interprets $T + B$

We say that S interprets T – $S \triangleright T$ – if there exists a mapping

$$j: \text{Form}_T \rightarrow \text{Form}_S$$

that *preserves structure*, for example, if \circ is a binary logical connective, then $(\varphi \circ \psi)^j = \varphi^j \circ \psi^j$ such that moreover

$$\forall \varphi (\Box_T \varphi \rightarrow \Box_S \varphi^j).$$

Example

Natural numbers can be interpreted as sets.

We can define the interpretability logic of a theory T .

$$\text{IL}(T) := \{A \mid \forall * \ T \vdash A^*\},$$

where A is a formula in the language $L_{\Box, \triangleright}$

$$F := \perp \mid \text{Prop} \mid F \rightarrow F \mid \Box F \mid F \triangleright F,$$

and $*$ is a translation sending propositional variables to arithmetical sentences.

Motivation

The axioms of the basic interpretability **IL** are

$$\text{L1 } \Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$$

$$\text{J2 } (A \triangleright B) \wedge (B \triangleright C) \rightarrow A \triangleright C$$

$$\text{L2 } \Box A \rightarrow \Box \Box A$$

$$\text{J3 } A \triangleright C \wedge B \triangleright C \rightarrow A \vee B \triangleright C$$

$$\text{L3 } \Box(\Box A \rightarrow A) \rightarrow \Box A$$

$$\text{J4 } A \triangleright B \rightarrow (\Diamond A \rightarrow \Diamond B)$$

$$\text{J1 } \boxed{\Box(A \rightarrow B) \rightarrow A \triangleright B}$$

$$\text{J5 } \boxed{\Diamond A \triangleright A}$$

Remark

- J1 tells us that the identity translation yields an interpretation.
- J5 represents Henkin's completeness theorem formalised.

Motivation

There are some interesting principles of interpretability. Namely,

$$M := A \triangleright B \rightarrow A \wedge \Box C \triangleright B \wedge \Box C \quad (\text{Montagna})$$

$$P := A \triangleright B \rightarrow \Box(A \triangleright B) \quad (\text{Persistence})$$

It is known that

$$IL(PA) := ILM \quad (\text{Full induction})$$

and

$$IL(I\Sigma_1) := ILP \quad (\text{Finitely Axiomatized}).$$

ILM and ILP motivate the characterisation of $IL(All)$.

$$IL(All) := \{A \mid \forall T \supseteq I\Delta_0 + Exp \ \forall * \ T \vdash A^*\},$$

the interpretability logic of all “reasonable” arithmetical theories.

Remark

$$IL(All) \subsetneq ILM \cap ILP$$

We present some advances on its modal characterization.

Semantics and intersections

Semantics and intersections

In interpretability logic, models are 4-tuples

$$\mathcal{M} := \langle W, R, \{S_x\}_{x \in W}, V \rangle$$

where

- $W \neq \emptyset$
- $R \subseteq W \times W$
- $S_x \subseteq x \upharpoonright \times x \upharpoonright$
- $V: \text{Prop} \rightarrow \mathcal{P}(W)$

$$x \upharpoonright := \{y \mid xRy\}.$$

R transitive and conversely well-founded;

S_x is reflexive, transitive and contains R on $x \upharpoonright$.

$\mathcal{F} = \langle W, R, \{S_x\}_{x \in W} \rangle$ denotes a frame.

Sometimes we denote models as $\mathcal{M} = \langle \mathcal{F}, V \rangle$.

Propositions, implications and *falsum* (\perp) are forced as usual.

The forcing of formulas $\Box A$ is

$$\mathcal{M}, x \Vdash \Box A: \iff \forall y (xRy \rightarrow \mathcal{M}, y \Vdash A).$$

The forcing of formulas $A \triangleright B$ is

$$\mathcal{M}, x \Vdash A \triangleright B: \iff \forall y (xRy \wedge \mathcal{M}, y \Vdash A \rightarrow \exists z: yS_x z \wedge \mathcal{M}, z \Vdash B).$$

Semantics and intersections

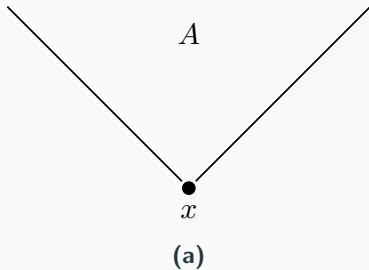
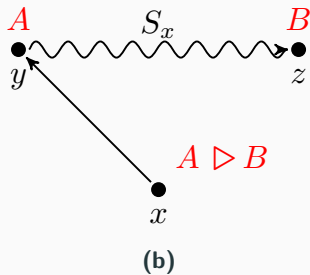


Figure 1: (a) $\Box A$ is forced at x



(b) $A \triangleright B$ is forced at x

Validity on models and frames is defined as follows.

Validity

- **Validity of a formula on a model:**

$$\mathcal{M} \models \varphi \text{ iff } \mathcal{M}, w \Vdash \varphi, \text{ for all } w \in W.$$

- **Validity of a formula on a frame:**

$$\mathcal{F} \models \varphi \text{ iff } \forall V \langle \mathcal{F}, V \rangle \models \varphi.$$

- **Validity of a scheme:** A model or a frame validates a scheme X ($\mathcal{M} \models X$ and $\mathcal{F} \models X$, respectively) iff it validates all X 's instances.

Semantics and intersections

The **frame condition** of a scheme X is a first (or higher) order predicate formula \mathcal{C} such that

$$\forall \mathcal{F} (\mathcal{F} \models \mathcal{C} \iff \mathcal{F} \models X).$$

Example

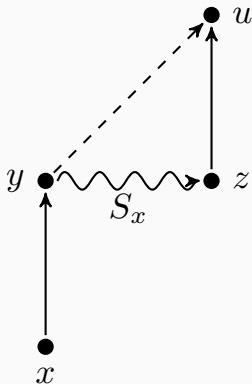
$$\mathcal{F} \models \Box A \rightarrow \Box \Box A \iff \mathcal{F} \models \forall x, y, z \left(xRy \wedge yRz \rightarrow xRz \right)$$

Frame conditions of ILM and ILP.

$$\mathcal{F} \models M \iff \mathcal{F} \models xRyS_xzRu \rightarrow yRu.$$

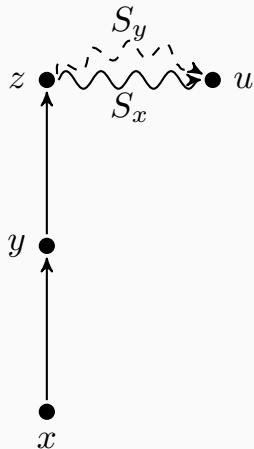
$$\mathcal{F} \models P \iff \mathcal{F} \models xRyRzS_xu \rightarrow zS_yu.$$

Semantics and intersections



(a)

Figure 2: Frame condition of M (a)



(b)

Frame condition of P (b)

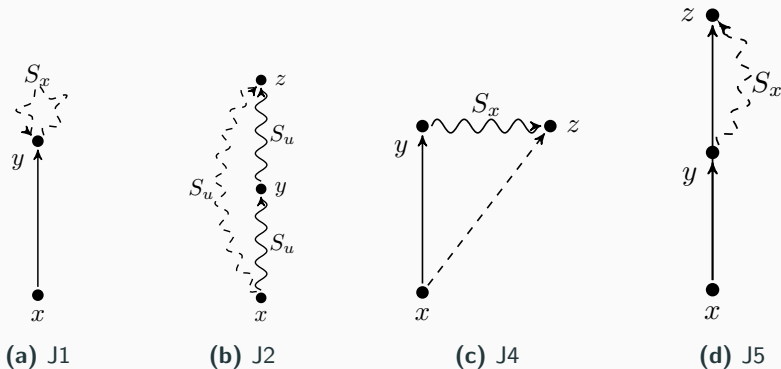


Figure 3: Frame definition reflecting axioms $\Box(A \rightarrow B) \rightarrow A \triangleright B$ (J1), $A \triangleright B \wedge B \triangleright C \rightarrow A \triangleright C$ (J2), $A \triangleright B \rightarrow (\Diamond A \rightarrow \Diamond B)$ (J4) and $\Diamond A \triangleright A$ (J5)

Semantics and intersections

Sometimes we need to close on the frame properties.

Closure

The closure of a (proto-) frame $\mathcal{F} := \langle W, R, \{S_x\}_{x \in W} \rangle$ under some principle X is the smallest structure

$\overline{\mathcal{F}}^X := \langle W, \overline{R}^X, \{\overline{S}_x^X\}_{x \in W} \rangle$ satisfying X such that $R \subseteq \overline{R}^X$ and $S_x \subseteq \overline{S}_x^X$, for every $x \in W$.

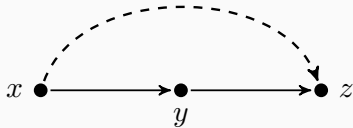


Figure 4: Transitive closure

Semantics and intersection

Frame operator

If $L = \{\phi_i\}_i$ is a set of atomic predicates (like xRy or yS_xz , etc.), we define the **IL-frame induced by L** , $\overline{\mathcal{F}(\bigwedge_i \phi_i)}^{\text{IL}}$, as the universal closure of the smallest proto-frame that satisfies all atomic predicates.

For brevity, we will write $\mathcal{F}(\bigwedge_i \phi_i)$.

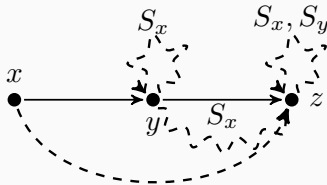


Figure 5: Closure of $\{xRy, yRz\}$ under **IL** frame requirements.

Semantics and intersection

Let \mathfrak{F} be a class of **IL**-frames. We define the interpretability logic corresponding to \mathfrak{F} .

$$\mathbf{IL}[\mathfrak{F}] := \{A : \text{for all } \mathcal{F} \in \mathfrak{F}, \mathcal{F} \models A\}.$$

Let F_{xyz} denote any first or higher order formula where the only free variables are x, y, z .

We now define the following class of conditions.

$$\begin{aligned} \mathcal{C}_{\mathbf{ILP} \cap_S \mathbf{ILM}} := \\ \{F_{xyz} \rightarrow xS_yz : \mathbf{ILP} \models F_{xyz} \rightarrow xS_yz \wedge \mathbf{ILM} \models F_{xyz} \rightarrow xS_yz\}. \end{aligned}$$

Also, we define the class

$$\mathfrak{M} := \{\mathcal{F} \models \mathbf{ILW} : \forall C \in \mathcal{C}_{\mathbf{ILP} \cap_S \mathbf{ILM}}, \mathcal{F} \models C\}.$$

The principle W is

$$W := A \triangleright B \rightarrow A \triangleright (B \wedge \Box \neg A)$$

and its frame condition is that there are no $S_x; R$ infinite chains.

Conjecture 1 (Goris, Joosten 2020)

$$\mathbf{IL}(\text{All}) = \mathbf{IL}[\mathfrak{All}].$$

Recall

$$\mathbf{IL}(\text{All}) := \{A \mid \forall T \supseteq I\Delta_0 + \text{Exp} \ \forall * T \vdash A^*\}.$$

$M \cap P$ -closure

Given a proto-frame $\mathcal{F} = \langle W, R, S \rangle$, its $M \cap P$ -closure is $\overline{\mathcal{F}}^{M \cap P} := \overline{\mathcal{F}}^M \cap \overline{\mathcal{F}}^P = \langle W, \overline{R}^M \cap \overline{R}^P, \overline{S}^M \cap \overline{S}^P \rangle$.

As an example, consider the principle M_0

$$M_0 := A \triangleright B \rightarrow \Diamond A \wedge \Box C \triangleright B \wedge \Box C,$$

whose frame condition is

$$\forall x, y, z, u, v \left(xRyRzS_xuRv \rightarrow yRv \right).$$

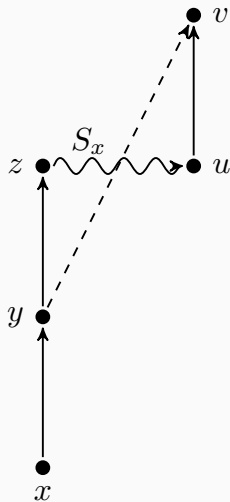
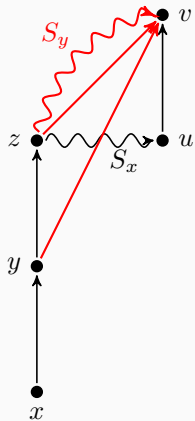


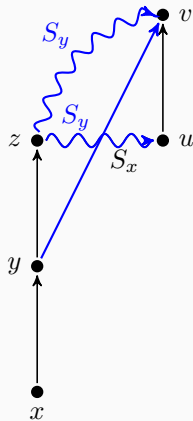
Figure 6: M_0

Semantics and intersection



(a)

Figure 7: (a) M closure and



(b)

(b) P closure

$M \cap_{\mathcal{F}} P$ -clause set

We define the $M \cap_{\mathcal{F}} P$ -clause set as

$$\bigwedge_i \phi_i \rightarrow \varphi : \in M \cap_{\mathcal{F}} P \text{ iff } \overline{\mathcal{F}(\bigwedge_i \phi_i)}^{M \cap P} \models \varphi$$

whenever $\{\phi_i\}_i \cup \{\varphi\}$ is a set of atomic predicates so that $\mathcal{F}(\bigwedge_i \phi_i)$ defines a proto-frame.

Remark

$\bigwedge_i \phi_i \rightarrow \varphi$ is a Horn clause.

Non-empty since the M_0 frame condition belongs to it.

It is known the *Broad* series and the *Slim* hierarchy belong to it.

Semantics and intersection

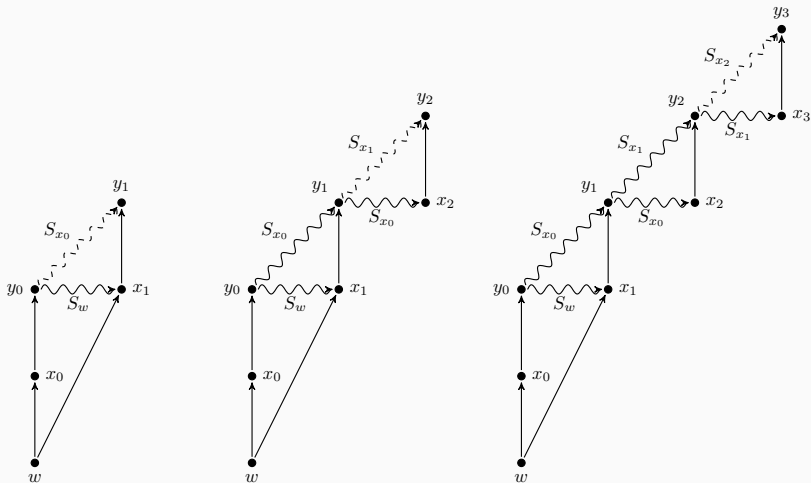


Figure 8: Slim (or Staircase) hierarchy

Semantics and intersection

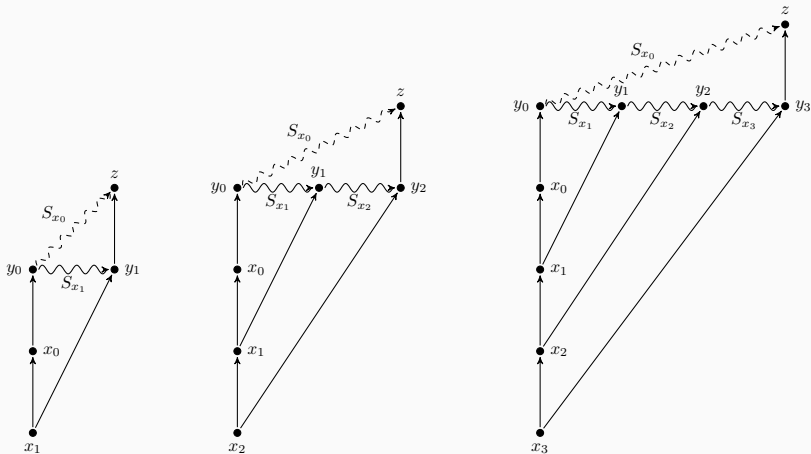


Figure 9: Broad series

Semantics and intersection

$M \cap_{\mathcal{F}} P$ defines a fragment of $\mathbf{IL}[\mathfrak{All}]$.

Let us define the lower-case class of \mathbf{IL} -frames

$$\mathfrak{all} := \{\mathcal{F} \models \mathbf{IL}W : \forall C \in M \cap_{\mathcal{F}} P, \mathcal{F} \models C\}.$$

Theorem

$$\mathbf{IL}[\mathfrak{all}] \subseteq \mathbf{IL}[\mathfrak{All}].$$

Remark

- It is unknown if $\mathbf{IL}[\mathfrak{all}] \subset \mathbf{IL}[\mathfrak{All}]$.
- $\mathbf{IL}[\mathfrak{all}]$ entails the frame conditions of *Broad* and *Slim*.

It is natural to conjecture that

Conjecture 2

$$\mathbf{IL}[\mathfrak{all}] = \mathbf{IL}(\mathbf{All}).$$

This new conjecture strengthens the old conjecture.

Conjecture 1 (Goris, Joosten 2020)

$$\mathbf{IL}(\mathbf{All}) = \mathbf{IL}[\mathfrak{All}].$$

How can we get a grip on $M \bigcap_{\mathcal{F}} P$?

One may try to focus on the clauses that imply an R -pair and conjecture that

Conjecture 3

Consider an **IL**-frame $\mathcal{F} = \langle W, R, S \rangle$. Then, for any $x, y \in W$, we have that $x\bar{R}^M y \wedge x\bar{R}^P y \wedge \neg(xRy) \rightarrow x\bar{R}^{M_0} y$.

Nonetheless, this is disproven by the...

Pencil frame

Pencil frame

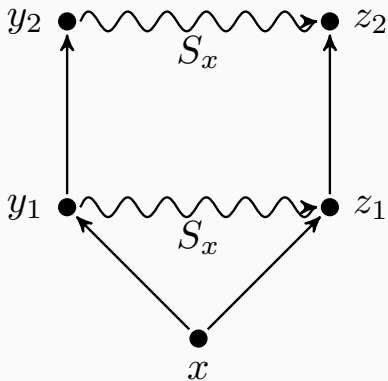


Figure 10: Pencil frame.

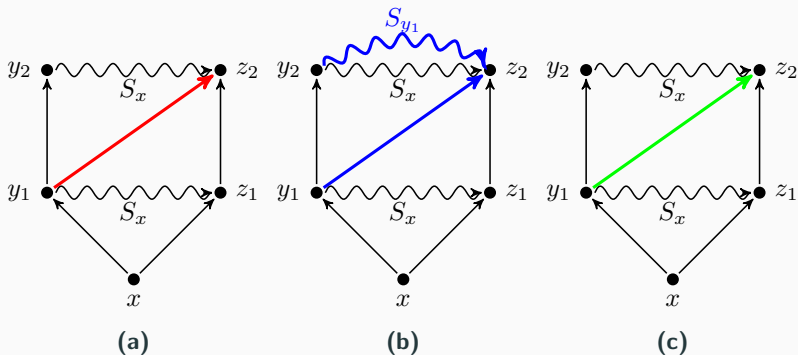


Figure 11: (a) M-closure (b) P-closure (c) Intersection.

Remark

Observe the **green arrow** is not in the M_0 -closure.

Dealing with confluence

On confluence

The unnecessary confluence of Pencil frames hint at their modal undefinability.

Confluence is inherent in interpretability logics (e.g., $xRyS_xz$ implies xRz), but we can unravel **IL**-models into bisimilar *tree-like* models w.r.t. R relations.

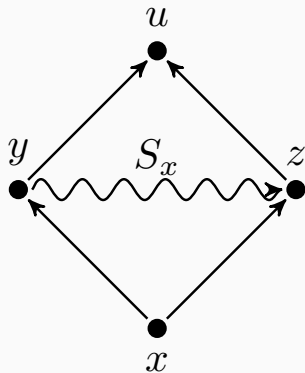
Tree-like IL-model

An **IL**-model $\mathcal{M} = \langle W, R, S, V \rangle$ is **tree-like** if

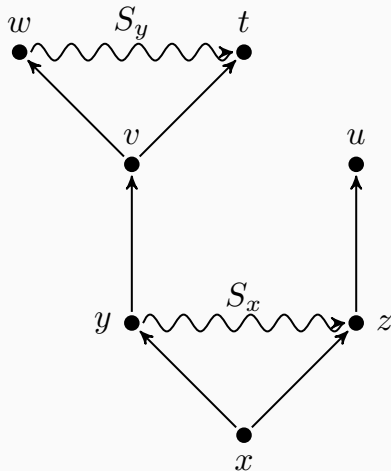
- (TL1) there exists a unique root regarding R and,
- (TL2) for every world except for the root, there is a immediate unique predecessor regarding R_0

$$xR_0y \text{ iff } xRy \text{ and } \neg\exists z: xRzRy$$

On confluence



(a)



(b)

Figure 12: (a) Frame **not** satisfying TL2

(b) Frame satisfying TL2.

IL-bisimulation

Two **IL**-models $\mathcal{M} = \langle W, R, S, V \rangle$ and $\mathcal{M}' = \langle W', R', S', V' \rangle$ are **bisimilar**, $\mathcal{M} \underline{\leftrightarrow} \mathcal{M}'$, if there is some $\emptyset \neq Z \subseteq W \times W'$ s.t.:

1. **In:** If wZw' , then $w \in V(p)$ iff $w' \in V'(p)$, $\forall p \in \text{Prop}$.
2. **Back:** If wZw' and there is $u \in W$ such that wRu , then there is $u' \in W'$ such that $w'R'u'$ and uZu' . Also, if $u'S'_w v'$, for some $v' \in W'$, then there is $v \in W$ such that $uS_w v$ and vZv' .
3. **Forth:** If wZw' and there is $u' \in W'$ such that $w'R'u'$, then there is $u \in W$ such that wRu and uZu' . Also, if $uS_w v$, for some $v \in W$, then there is $v' \in W'$ such that $u'S'_w v'$ and vZv' .

Use $\mathcal{M}, x \underline{\leftrightarrow} \mathcal{M}', x'$ to indicate that there is a bisimulation connecting x and x' .

On confluence

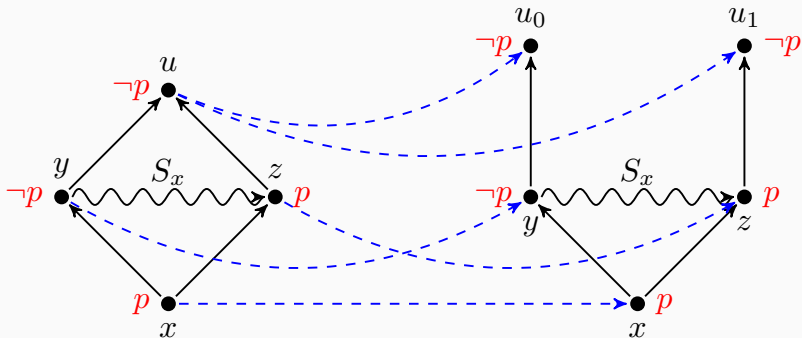


Figure 13: IL-bisimulation between two IL-models represented in dashed blue arrows.

Key idea

These models *mimic* each other.

In fact, bisimilar **IL**-models prove the same modal formulas.

Invariant for bisimulation

A modal formula φ is **invariant under bisimulation** if whenever $x \leftrightarrow x'$, then $x \models \varphi$ iff $x' \models \varphi$.

Of course, we have the following theorem.

Theorem 1

Modal formulas are invariant under bisimulation for **IL**-models.

For *restricted* **IL**-models we can find bisimilar tree-like **IL**-models.

Restriction of an **IL**-model

If $\mathcal{M} = \langle W, R, S, V \rangle$ is an **IL**-model, then its **restriction** to $w \in W$ is a model $\mathcal{M} \upharpoonright w = \langle W_{|w}, R_{|w}, S_{|w}, V_{|w} \rangle$ where $W_{|w} = w \upharpoonright \cup \{w\}$, $R_{|w} = \{uRv : u, v \in W_{|w}\}$, $S_{|w} = \{S_u\}_{u \in W_{|w}}$ and $V_{|w} : \text{Prop} \rightarrow \mathcal{P}(W_{|w})$.

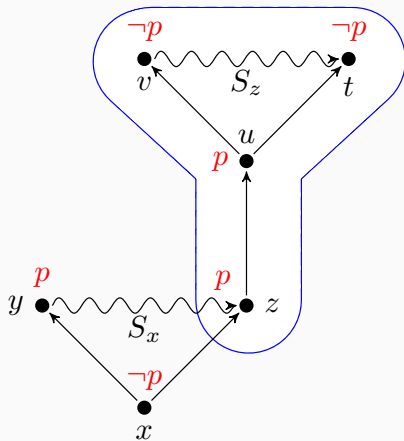


Figure 14: Inclosed in blue: a restricted **IL**-model.

Indeed,

Theorem 2

For each **IL**-model $\mathcal{M} = \langle W, R, S, V \rangle$ and world $w_0 \in W$, $\mathcal{M}|_{w_0}$ is bisimilar to a tree-like **IL**-model $\mathcal{M}' = \langle W', R', S', V' \rangle$ that is R -wise, in other words, according to the relation R .

Key idea

Use paths.

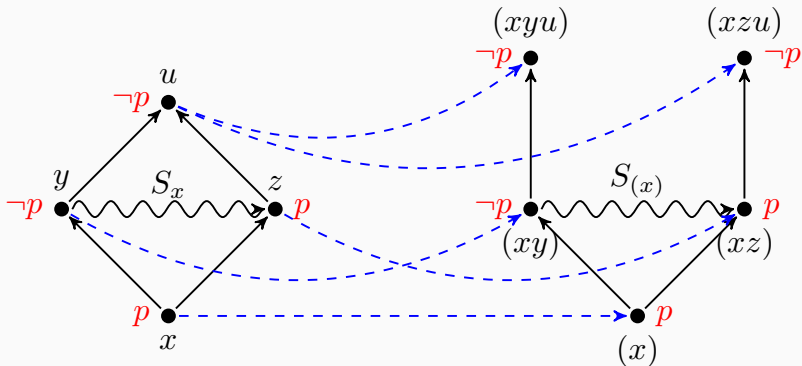


Figure 15: Example of a bisimulation using paths.

A new semantics: Minimal Veltman Semantics

Minimal Veltman Semantics

Let us focus on frames and propose an alternative semantics that avoids certain confluences.

Key idea

To remove many S relations for easier control, with the missing relations compensated by the truth definition of \triangleright .

We propose the following frames.

Minimal Veltman frames

Consider a non-empty countable set of worlds W , $R \subseteq W \times W$ and, for each $x \in W$, $S_x \subseteq \downarrow x \times \downarrow x$ with R transitive and Noetherian; for every $x \in W$, S_x is **irreflexive and antitransitive**; for every $x, y, z \in W$, $yS_x z \rightarrow \neg yRz$. Then, $\mathcal{F} = \langle W, R, \{S_x\}_{x \in W} \rangle$ is a **Minimal Veltman frame** or MV-frame and \mathcal{C}^{MVF} is the **class of MV-frames**.

Minimal Veltman Semantics

Minimal Veltman models

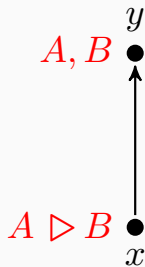
A **Minimal Veltman model** (MV-model) is a tuple $\langle W, R, \{S_x\}_{x \in W}, V \rangle$, where $\langle W, R, \{S_x\}_{x \in W} \rangle$ is an MV-frame, and $V : \text{Prop} \rightarrow \mathcal{P}(W)$ is a valuation. The forcing relation \Vdash_{MV} follows the standard definition for **IL**-models, except for the \triangleright -modality: ¹

$$\mathcal{M}, x \Vdash_{\text{MV}} A \triangleright B \iff \forall y (xRy \Vdash_{\text{MV}} A \rightarrow \exists z \textcolor{red}{y(R \cup S_x)^* z} \Vdash_{\text{MV}} B).$$

$$\mathcal{M}, x \Vdash A \triangleright B \iff \forall y (xRy \Vdash A \rightarrow \exists z \textcolor{red}{yS_x z} \Vdash_{\text{MV}} B).$$

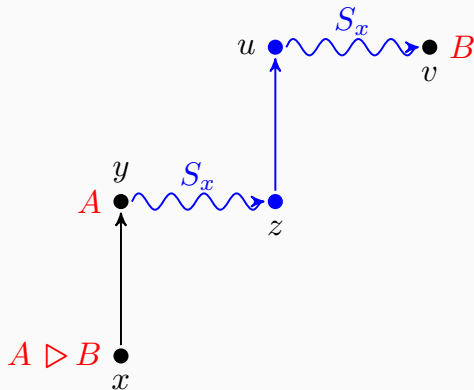
¹Given a binary relation Z , we denote Z^* as the composition of 0 or more copies of Z . We set Z^+ as the composition of 1 or more copies of Z . In fact, $Z^+ = Z; Z^*$.

Minimal Veltman Semantics



(a)

Figure 16: (a) 0 copies of $R \cup S_x$



(b)

(b) 3 copies of $R \cup S_x$

Remark

- We denote the **class of MV-models** as \mathcal{C}^{MVM} .
- Validity for formulas and schemes in MV-models and MV-frames is defined as usual, using \Vdash_{MV} and \models_{MV} for forcing and consequence, respectively.

IL is sound wrt the Minimal Veltman Semantics (MVS).

Theorem 3

$$\mathbf{IL} \vdash \varphi \Rightarrow \forall \mathcal{F} \in \mathcal{C}^{\text{MVF}} \mathcal{F} \models_{\text{MV}} \varphi.$$

Completeness of Minimal Veltman Semantics

We can define bisimulations between **IL**-models and MV-models.

Bisimulation between IL-models and MV-models

An **IL**-model $\mathcal{M} = \langle W, R, S, V \rangle$ and an MV-model $\mathcal{M}' = \langle W', R', S', V' \rangle$ are **bisimilar**, $\mathcal{M} \underline{\leftrightarrow} \mathcal{M}'$, if there is some $\emptyset \neq Z \subseteq W \times W'$ such that:

1. **In:** If wZw' , then $w \in V(p)$ iff $w' \in V'(p)$, $\forall p \in \text{Prop}$.
2. **Back:** If wZw' and there is $u \in W$ such that wRu , then there is $u' \in W'$ with $w'R'u'$ and uZu' . Also, if $u'(R' \cup S'_{w'})^*v'$, for some $v' \in W'$, then there is $v \in W$ such that uS_wv and vZv' .
3. **Forth:** If wZw' and there is $u' \in W'$ such that $w'R'u'$, then there is $u \in W$ such that wRu and uZu' . Also, if uS_wv , for some $v \in W$, then there is $v' \in W'$ such that $u'(R' \cup S'_{w'})^*v'$ and vZv' .

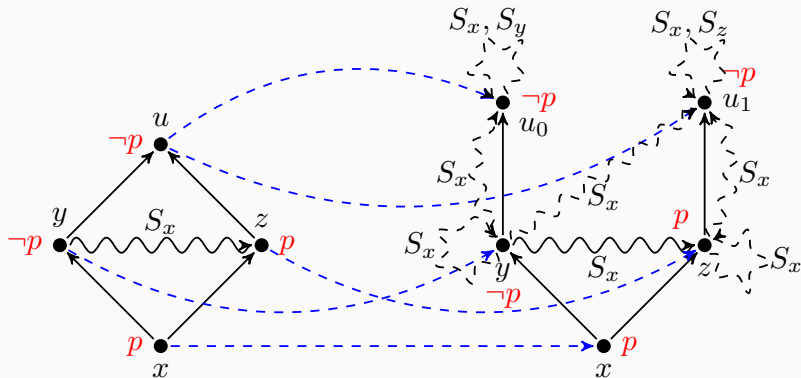


Figure 17: The model on the right is an MV-model. The dashed arrows are the missing arrows.

As always,

Theorem 5

Modal formulas are invariant under bisimulation between **IL**- and MV-models.

and we can prove that

Theorem 6

Each **IL**-model is bisimilar to a MV-model.

Key idea

Remove S -relations incompatible with the MV-frame definition.

Completeness

$\forall \mathcal{F} \in \mathcal{C}^{\text{MVM}} \mathcal{F} \models_{\text{MV}} \varphi \Rightarrow \mathbf{IL} \vdash \varphi.$

Overview of the proof.

- Assume $\mathbf{IL} \not\vdash \varphi$.
- By \mathbf{IL} -completeness, there is an \mathbf{IL} -model such that $\mathcal{M}, w \not\models \varphi$.
- $\mathcal{M} \Leftrightarrow \mathcal{M}'$ for some MV-model \mathcal{M}' . (Thm. 6)
- \mathcal{M} and \mathcal{M}' prove the same modal formulas. (Thm. 5)
- Thus, $\mathcal{M}' \not\models_{\text{MV}} \varphi$.

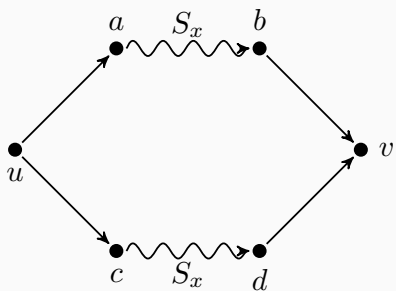


Confluence revisited

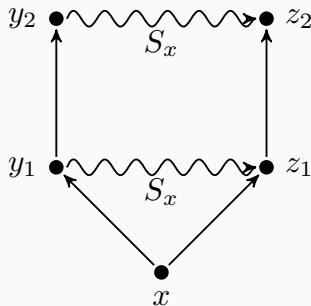
Confluence revisited

Avoiding confluence in frames is desirable. We present the (non-trivial) *Unique Path Condition*.

$$(uR^*aS_xbR^*v \wedge uR^*cS_xdR^*v) \rightarrow \langle a, b \rangle = \langle c, d \rangle. \quad (\text{UPath})$$



(a)



(b)

Figure 18: (a,b) Examples of frames that **do not** satisfy UPath

That motivates the definition of a new type of models that avoids confluences.

Tree-some (not threesome) models

A rooted MV-model is **tree-some** if it satisfies (UPath).

Also, we can define bisimulations between MV-models.

MV-bisimulation

Two MV-models $\mathcal{M} = \langle W, R, S, V \rangle$ and $\mathcal{M}' = \langle W', R', S', V' \rangle$ are **MV-bisimilar** ($\mathcal{M} \Leftrightarrow_{\text{MV}} \mathcal{M}'$) if there is a $\emptyset \neq Z \subseteq W \times W'$:

1. **In:** If wZw' , then $w \in V(p)$ iff $w' \in V'(p)$, $\forall p \in \text{Prop}$.
2. **Back:** If wZw' and there exists $u \in W$ such that wRu , then there exists $u' \in W'$ such that $w'R'u'$ and uZu' . Also, if $u'(R' \cup S'_{w'})^*v'$, for some $v' \in W'$, then there exists $v \in W$ such that $u(R' \cup S_w)^*v$ and vZv' .
3. **Forth:** If wZw' and there exists $u' \in W'$ such that $w'R'u'$, then there exists $u \in W$ such that wRu and uZu' . Also, if $u(R' \cup S_w)^*v$, for some $v \in W$, then there exists $v' \in W'$ such that $u'(R' \cup S'_{w'})^*v'$ and vZv' .

Of course, modal formulas are invariant under MV-bisimulation.

Theorem 7

Modal formulas are invariant under MV-bisimulation.

Theorem 8

For each MV-model $\mathcal{M} = \langle W, R, S, V \rangle$ and each world $w_0 \in W$ we have that $\mathcal{M} \upharpoonright w_0$ is bisimilar to a tree-some MV-model $\mathcal{M}' = \langle W', R', S', V' \rangle$.

Key idea

Use (more complicated) labeled paths.

This fact allows us to avoid the Pencil model, but Slim and Broad series work with tree-some frames.

Confluence revisited

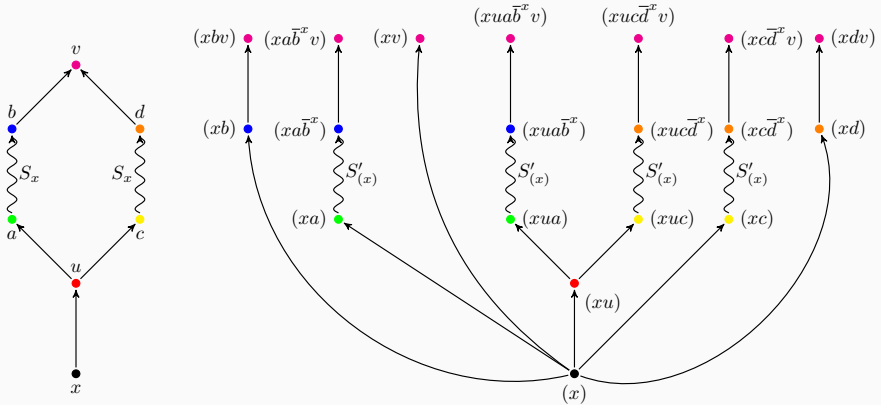


Figure 19: Full unraveling. Bisimilarity indicated by colours.

Summary (Summer-e) and Further work

All in all,

- We strengthen the old conjecture by focusing on Horn clauses;
- We show that all known principles fall in this class;
- We addressed unnecessary confluence via unraveling techniques over tree-like structures;
- We define a new semantics to mitigate confluences and proved its soundness and completeness using bisimulation.

There is still some work to do.

- Explore if every (restricted) **ILM**- or **ILP**-model admits a bisimilar (tree-some) MV-model satisfying the corresponding M- or P-frame condition.

Thank you! Obrigado!