Algebras and general frames for interpretability logics

Teo Šestak

University of Zagreb
teo.sestak@fsb.unizg.hr

TACL 2026

Table of Contents

Table of Contents

Interpretability logics

- extend provability logics **GL**,
- “base” system **IL** and extensions,
- are modeled by Veltman frames, which extend Kripke frames for **GL**.

Veltman Semantics

- soundness and weak completeness theorems for **IL**,
- strong completeness does not hold,
- weak completeness proofs are not modular,
- there exist incomplete extensions.

Approach to these problems in modal logic:

- general frames – Kripke frames with additional structure
  - limitations on valuations,
  - include Kripke frames as special cases.
- Boolean algebras with operators – operators $f_\Box$ and $f_\Diamond$ which interpret modal operators.
  - “designed to fit the logic”,
  - include Kripke frames (in a way).
- general frames and Boolean algebras with operators are closely related!
  - a general frame generates a Boolean algebra with operators and vice-versa, with validity being preserved.

Table of Contents

The language $\mathcal{L}(\mathbf{IL})$ of interpretability logics is given via the following grammar:

$$\varphi ::= p \mid \bot \mid \varphi \to \varphi \mid \varphi \rhd \varphi,$$

where $p \in \mathrm{Prop}$ is a propositional variable.

Connectives and operators $\neg$, $\wedge$, $\vee$, $\leftrightarrow$, $\top$, $\Box$ and $\Diamond$ are defined:

- $\neg\varphi := \varphi \to \bot$,
- $\varphi \wedge \psi := \neg(\varphi \to \neg\psi)$,
- $\varphi \vee \psi := \neg\varphi \to \psi$,
- $\varphi \leftrightarrow \psi := (\varphi \to \psi) \wedge (\psi \to \varphi)$,
- $\top := \neg\bot$,
- $\Box\varphi := (\neg\varphi) \rhd \bot$ and
- $\Diamond\varphi := \neg\Box\neg\varphi$.

The system **IL** contains all tautologies and all instantiations of the following axiom schemata:

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

J2 $((\varphi \rhd \psi) \wedge (\psi \rhd \chi)) \to (\varphi \rhd \chi)$,

J3 $((\varphi \rhd \chi) \wedge (\psi \rhd \chi)) \to ((\varphi \vee \psi) \rhd \chi)$,

J4 $(\varphi \rhd \psi) \to (\Diamond\varphi \to \Diamond\psi)$,

J5 $(\Diamond\varphi \rhd \varphi)$.

Rules of inference are:

- modus ponens: from $\varphi \to \psi$ and φ conclude ψ,
- necessitation: from φ conclude $\Box\varphi$.

We use $\mathbf{IL}^+$ to denote an extension of **IL** obtained by adding a set of axiom schemata to the system.

A *proof* of a formula φ in $\mathbf{IL}^+$ is a finite list of formulae such that the last formula is φ and each formula in the list is

- a tautology,
- instantiation of one of the axioms of the system $\mathbf{IL}^+$,
- obtained by one of the rules of inference from preceding formulae.

We write $\vdash \varphi$.

A *derivation* of a formula φ from a set Γ in $\mathbf{IL}^+$ is a finite list of formulae such that the last formula is φ and each formula in the list is

- a theorem of $\mathbf{IL}^+$,
- an element of Γ,
- obtained by modus ponens from preceding formulae.

We write $\Gamma \vdash_{\mathbf{IL}^+} \varphi$.

Table of Contents

Definition

A *Veltman frame* $\mathfrak{F}$ is a triple $(W, R, \{S_w : w \in W\})$, where W is a non-empty set, R is a transitive and conversely well-founded binary relation on W and $\{S_w : w \in W\}$ a collection of binary relations on $R[w]$, where each S_w is reflexive and transitive and the restriction of R onto $R[w]$ is contained in S_w.

Definition

A *Veltman model* is a pair $\mathfrak{M} = (\mathfrak{F}, V)$ of a Veltman frame $\mathfrak{F}$ and a *valuation* function $V : \mathrm{Prop} \to \mathcal{P}(W)$.

A valuation induces a *forcing* relation $\Vdash \subseteq W \times \mathcal{L}(\mathbf{IL})$:

$$
\begin{aligned}
w \Vdash p &\iff w \in V(p), \\
w \Vdash \bot &\qquad \text{for no } w \in W, \\
w \Vdash \varphi \to \psi &\iff w \nVdash \varphi \text{ or } w \Vdash \psi, \\
w \Vdash \Box\varphi &\iff \forall v \left(wRv \Rightarrow v \Vdash \varphi\right), \\
w \Vdash \varphi \rhd \psi &\iff \forall u \left(wRu \mathrel{\&} u \Vdash \varphi \Rightarrow \exists v (uS_w v \mathrel{\&} v \Vdash \psi)\right).
\end{aligned}
$$

This allows us to extend any valuation to the set of all formulae $\varphi \in \mathcal{L}(\mathbf{IL})$:

$$V(\varphi) = \{w \in W : w \Vdash \varphi\}.$$

F. Veltman, D. de Jongh. *Provability Logics for Relative Interpretability*, Mathematical Logic, Springer, Boston, MA, 1990.

G. Japaridze, D. de Jongh. *The Logic of Provability*, Handbook of Proof Theory, Elsevier, Amsterdam, 1998.

Theorem

If $\nvdash_{\mathbf{IL}} \varphi$, then there exists a finite Veltman model $\mathfrak{M}$ such that $\mathfrak{M} \nVdash \varphi$.

Similar theorems hold for some extensions and their corresponding class of frames (Veltman frames with additional constraints), **IL**M and **IL**P.

The strong completeness theorem does not hold. Consider

$$S := \{\Diamond p_0\} \cup \{\Box(p_n \to \Diamond p_{n+1}) : n \in \mathbb{N}\}.$$

Problem: S requires an infinite R-chain, which contradicts the converse well-foundedness of R.

Solution: get rid of converse well-foundedness!

Definition

A *quasi-Veltman frame* $\mathfrak{F}$ is a triple $(W, R, \{S_w : w \in W\})$, where W is non-empty, R is a transitive and irreflexive binary relation on W and $\{S_w : w \in W\}$ a collection of binary relations on $R[w]$, where each S_w is reflexive and transitive and the restriction of R onto $R[w]$ is contained in S_w.

Definition

A *quasi-Veltman model* is a pair $\mathfrak{M} = (\mathfrak{F}, V)$ of a quasi-Veltman frame $\mathfrak{F}$ and a valuation function $V : \mathrm{Prop} \to \mathcal{P}(W)$.

Problem: There exist quasi-Veltman models where $\Box(\Box p \to p) \to \Box p$ does not hold. Consider $(\mathbb{N}, <, \{\leq|_{\{k+1,k+2,\dots\}} : k \in \mathbb{N}\}, V)$, where $V(p) = 2\mathbb{N}$. Then

- for no $n \in \mathbb{N}$ is $n \Vdash \Box p$,
- therefore, for all $n \in \mathbb{N}$, trivially, $n \Vdash \Box p \to p$,
- then we have $n \Vdash \Box(\Box p \to p)$,
- conclusion: for no $n \in \mathbb{N}$ does $n \Vdash \Box(\Box p \to p) \to \Box p$ hold.

Solution: do not allow valuations which do that! General frames limit valuations!

Table of Contents

Definition

A *general* **IL**-*frame* is a pair $(\mathfrak{F}, A)$ of a quasi-Veltman frame and a non-empty set $A \subseteq \mathcal{P}(W)$ of *admissible* subsets of W, closed under the following:

(i) *union:* if $X, Y \in A$, then $X \cup Y \in A$,

(ii) *complement:* if $X \in A$, then $W \setminus X \in A$,

(iii) $m_{\rhd}$*:* if $X, Y \in A$, then $m_{\rhd}(X, Y) \in A$, where

$$m_{\rhd}(X, Y) = \{w \in W : \forall u \in X \ (wRu \to \exists v \in Y \ (uS_w v))\},$$

such that the following holds:

(iv) if $X \in A$, then $m_{\Box}((W \setminus m_{\Box}(X)) \cup X) = m_{\Box}(X)$.

We use the shorthand $m_{\Box}(X) := m_{\rhd}(W \setminus X, \emptyset)$.

Definition

A *model based on a general* **IL***-frame* is a triple $(\mathfrak{F}, A, V)$, where $(\mathfrak{F}, A)$ is a general **IL**-frame, and $V : \mathrm{Prop} \to A$ is an *admissible valuation*.

Truth, validity and satisfiability on general **IL**-frames are defined analogously to the quasi-Veltman case, with the additional constraint of quantifying over admissible valuations, as opposed to all valuations.

- Every Veltman frame is a general **IL**-frame ($A = \mathcal{P}(W)$),
- The system **IL** is sound and strongly complete with respect to the class of all general **IL**-frames,
- Every extension $\mathbf{IL}^+$ of the system **IL** is sound and strongly complete with respect to some class of general $\mathbf{IL}^+$-frames (these are general **IL**-frames which validate all the axioms of $\mathbf{IL}^+$) – the proof uses the construction of canonical $\mathbf{IL}^+$-model, modular for all extensions.

Table of Contents

Definition

An *interpretability algebra* is a tuple $\mathfrak{A} = (A, +, -, 0, f_{\rhd})$, where $(A, +, -, 0)$ is a Boolean algebra and $f_{\rhd} : A \times A \to A$ is an operator which satisfies the following (where we use the shorthands $f_{\Box}(x) := f_{\rhd}(-x, 0)$ and $f_{\Diamond}(x) := -f_{\Box}(-x)$):

1. $f_{\Diamond}(0) = 0$,
2. $f_{\Box}(-f_{\Box}(x) + x) = f_{\Box}(x)$,
3. $f_{\rhd}(x, z) \cdot f_{\rhd}(y, z) = f_{\rhd}(x + y, z)$,
4. $-f_{\Box}(-x + y) + f_{\rhd}(x, y) = 1$
5. $-f_{\rhd}(x, y) + (-f_{\rhd}(y, z)) + f_{\rhd}(x, z) = 1$,
6. $-f_{\rhd}(x, y) + f_{\Box}(-x) + f_{\Diamond} y = 1$,
7. $f_{\rhd}(f_{\Diamond}(x), x) = 1$.

Definition

Let $\mathfrak{A} = (A, +, -, 0, f_{\rhd})$ be an interpretability algebra. A *valuation* is a function $\theta : \mathrm{Prop} \to A$.

A valuation extends to a set of all formulae:

$$\begin{aligned} \theta(\bot) &= 0, \\ \theta(\varphi \to \psi) &= -\theta(\varphi) + \theta(\psi), \\ \theta(\varphi \rhd \psi) &= f_{\rhd}(\theta(\varphi), \theta(\psi)). \end{aligned}$$

Definition

Let φ and ψ be formulae. An expression of the form $\varphi \approx \psi$ is an **IL**-*equality*.
An **IL**-equality $\varphi \approx \psi$ is *true* in an interpretability algebra $\mathfrak{A}$, which we denote as $\mathfrak{A} \models \varphi \approx \psi$, if for every valuation θ, we have $\theta(\varphi) = \theta(\psi)$.

T.Š. *Algebraic Semantics for Interpretability Logics*, Journal of Logic, Language and Information, 2026.

- Every Veltman frame gives rise to an interpretability algebra (on $\mathcal{P}(W)$),
- The system **IL** is sound and complete with respect to the class of all interpretability algebras,
- Every extension $\mathbf{IL}^+$ of the system **IL** is sound and complete with respect to the class $V_{\mathbf{IL}^+}$ of interpretability algebras – proof using Lindenbaum-Tarski algebra, modular for all extensions.

Table of Contents

Let $\mathfrak{F} = (W, R, \{S_w : w \in W\}, A)$ be a general **IL**-frame. Then

$$\mathfrak{F}^* := (A, \cup, {}^c, \emptyset, m_\rhd)$$

is an interpretability algebra.

For admissible valuations θ, the following holds:

$$
\begin{aligned}
(\mathfrak{F}, \theta), w \Vdash \varphi &\quad \text{if and only if} \quad w \in \theta(\varphi), \\
\mathfrak{F} \Vdash \varphi &\quad \text{if and only if} \quad \mathfrak{F}^* \vDash \varphi \approx \top, \\
\mathfrak{F} \Vdash \varphi \leftrightarrow \psi &\quad \text{if and only if} \quad \mathfrak{F}^* \vDash \varphi \approx \psi.
\end{aligned}
$$

For an interpretability algebra $\mathfrak{A}$, we can construct a general **IL**-frame $\mathfrak{A}_*$, using ultrafilters of an algebra $\mathfrak{A}$. – The proof is similar to the construction of the canonical $\mathbf{IL}^+$-model: W consists of pairs (u, τ), where u is an ultrafilter of $\mathfrak{A}$ and τ finite sequence of elements from $\mathfrak{A}$.

Main result:

Theorem

Let $\mathfrak{A}$ be an interpretability algebra. Then

$$(\mathfrak{A}_*)^* \simeq \mathfrak{A}.$$

In other words, every interpretability algebra is (up to an isomorphism) arises from a general **IL**-frame.

Thank you!