Universal frames for *GL* and *IL*

Vedran Čačić

PMF-MO

September 21, 2020
LAP 2020 @Zoom

Provability logic
motivation

What are the minimal properties of provability that make Gödel's incompleteness theorems' proof work?

Main idea: provability as a modal operator!

$$\Box\varphi :\Longleftrightarrow \varphi \text{ is provable}$$

Hilbert–Bernays' conditions $\rightsquigarrow$ modal axioms:

- K $\Box(\varphi \to \psi) \land \Box\varphi \to \Box\psi$
- 4 $\Box\varphi \to \Box\Box\varphi$
- Löb $\Box(\Box\varphi \to \varphi) \to \Box\varphi$

Formulas in closed fragment GL_0:

$$\varphi ::= \bot \mid (\varphi \to \varphi) \mid \Box\varphi$$

GL-frames

semantics for provability logic

Definition

A GL-frame is a structure (W, R),
where W is a nonempty set of worlds,
and R is a transitive accessibility relation on W such that R^{-1} is wellfounded: there is no sequence $w_0 \; R \; w_1 \; R \; w_2 \; R \cdots$.
Notation: $R[w] := \{u \in W \ : \ w \, R \, u\}$ — R-successors of w

Theorem (Soundness, completeness, finite model property)

- *If $\varphi \in GL_0$ is a theorem, it is true in (every world of) every GL-frame.*
- *If $\varphi \in GL_0$ is not a theorem, there exists a finite GL-frame and a world in it, on which φ doesn't hold.*

World and frame depth

Depth is a function from worlds to ordinals, defined inductively (recursively) over R^{-1}:

$$\rho(w) := \sup_{wRv} \rho(v)^{+} = \rho\big[R[w]\big]$$

Lemma (depth lemma)

If $\rho(w) > \alpha$, then there exists v such that $w\,R\,v$ and $\rho(v) = \alpha$.

Corollary

Depth's image $\operatorname{rng}\rho$ *is also an ordinal;*
we call it "frame depth" and denote it by $\rho(\mathfrak{M})$.

Modal equivalence

Definition

We say that worlds w and v (possibly in different frames) are modally equivalent if the same modal formulas hold on them. (In frames, we only look at *closed* modal formulas.)

Modal equivalence is weaker than the isomorphism:
every modal formula "sees" only down to a fixed finite depth (number of nested modal operators in the formula).

So, the worlds at depths $\omega + 1$ and $\omega + 2$ are modally equivalent, but not isomorphic.

(Or, the frames of depths $\omega + 2$ and $\omega + 3$ are not isomorphic.)

Remark

For every ordinal α there is a frame $(\alpha, \ni)$ of depth exactly α.

Universal frame for GL
existence

$\bar{\mathfrak{N}} := (\omega + 1, \ni)$ is universal: for every GL-frame $\mathfrak{N}$ and for every world w within it, the world $\bar{w} := \min\{\rho(w), \omega\}$ within $\bar{\mathfrak{N}}$ is modally equivalent to w.

Moreover, $\bar{w}$ is the unique such world in $\bar{\mathfrak{N}}$: it's obviously the only one at its depth, and the worlds of different finite depths aren't modally equivalent, as well as a world of finite and a world of infinite (ω) depth.

Why? Because for every $n \in \omega$ there is a formula characterizing the worlds of depth exactly n:

$$\chi_0 := \Box\bot, \qquad \chi_{n+1} := \left(\Box^{n+2}\bot \land \Diamond\chi_n\right)$$

(of course, $\Diamond\varphi$ means $\neg\Box\neg\varphi$).
The world ω is characterized by elimination: no χ_n holds on it.

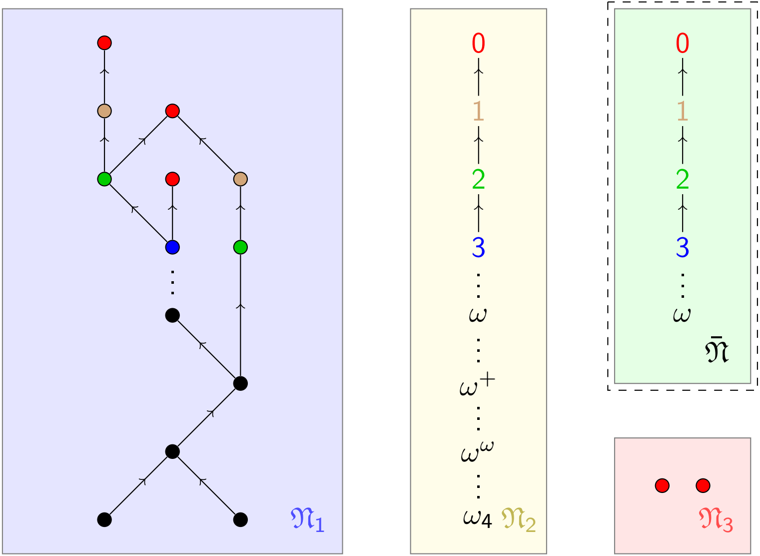
Universal frame for *GL*
uniqueness

Moreover, the $\bar{\mathfrak{N}}$ itself is unique with such property (that for every world in every frame there is a unique modally equivalent world within it) up to isomorphism.

Indeed, it must have a world at every finite depth, and a world at an infinite depth, and in each "bin" it must have exactly one world (since all the worlds at the same depth are modally equivalent, otherwise the world uniqueness property wouldn't hold).
Also, the only world at infinite depth must be at depth exactly ω, since otherwise it would have an R-successor at depth ω by depth lemma, so there'd be two worlds of infinite depth and they'd be modally equivalent.
Similarly, R is uniquely pinpointed, again by depth lemma and world uniqueness.

Universal frame for *GL*

a picture



Interpretability logic
motivation

GL nicely models provability in many reasonable theories.
1976. Solovay proves: *GL* really is the provability logic of *PA*.

What about comparing the (extensions of) theories with respect to relative strength?
Which extension (of some basic theory) can interpret another?

Modal approach: over a basic theory T,

$$\Box\varphi \quad \leftrightsquigarrow \quad T \text{ proves } \varphi$$
$$(\varphi \rhd \psi) \quad \leftrightsquigarrow \quad T + \varphi \text{ interprets } T + \psi$$

One provability, many interpretabilities

$IL := GL +$ "interpretability principles"

- different principles for different basic theories T

Extensions are interpretability logics of various famous theories:

when	who	T	principle	extension
1988	Visser	NBG	permanence	ILP
1990	Berarducci	PA	Montagna	ILM

GIL (common to all reasonable theories) is still sought for.

Interpretability logic
syntax (closed fragment)

- logical constant $\bot$, connective $\to$, modal operator $\rhd$
- (closed) formula: $\varphi ::= \bot \mid (\varphi \to \varphi) \mid (\varphi \rhd \varphi)$
- modal depth:
  - $\delta(\bot) := 0$
  - $\delta(\varphi \to \psi) := \max\{\delta(\varphi), \delta(\psi)\}$
  - $\delta(\varphi \rhd \psi) := 1 + \max\{\delta(\varphi), \delta(\psi)\}$
- everything else ($\neg$, $\top$, $\wedge$, $\vee$, $\leftrightarrow$, $\Box$, $\Diamond$) is derived syntax:

$$\neg A :\Leftrightarrow A \to \bot \qquad A \wedge B :\Leftrightarrow \neg(A \to \neg B)$$

$$\top :\Leftrightarrow \neg\bot \qquad A \vee B :\Leftrightarrow \neg A \to B$$

$$\Box A :\Leftrightarrow \neg A \rhd \bot \qquad A \leftrightarrow B :\Leftrightarrow (A \to B) \wedge (B \to A)$$

$$\Diamond A :\Leftrightarrow \neg\Box\neg A$$

Interpretability logic

derivation system

Axioms:

PT	all propositional tautologies
Löb	$\Box(\Box A \to A) \to \Box A$
J1	$\Box(A \to B) \to (A \rhd B)$
J2	$\big((A \rhd B) \wedge (B \rhd C)\big) \to (A \rhd C)$
J3	$\big((A \rhd C) \wedge (B \rhd C)\big) \to \big((A \vee B) \rhd C\big)$
J4	$(A \rhd B) \to (\Diamond A \to \Diamond B)$
J5	$\Diamond A \rhd A$

Inference rules: MP: $\dfrac{A \qquad A \to B}{B}$ Gen: $\dfrac{A}{\Box A}$

GL is a subtheory of *IL*: K follows from J3

4 follows from J4 and J5

Veltman frames
semantics

$\mathfrak{M} = \big(W, R, (S_w)_{w \in W}\big)$, where (W, R) is a *GL*-frame and each S_w is a transitive reflexive relation on $R[w]$, extending R there

$\Vdash$ is a relation between worlds and formulas:

- $w \nVdash \bot$ for every world w
- $w \Vdash F \to G$ means $w \nVdash F$ or $w \Vdash G$
- $w \Vdash F \rhd G$ means: for every v such that $w \, R \, v \Vdash F$, there is u such that $v \, S_w \, u \Vdash G$.

Universal frame for IL
nonexistence

Theorem

There is no universal frame for IL.

Proof: Assume the contrary, that
there is a Veltman frame $\bar{\mathfrak{M}}$ such that
for every Veltman frame $\mathfrak{M}$,
for every world w within $\mathfrak{M}$,
there is a unique world $\bar{w}$ within $\bar{\mathfrak{M}}$
such that for every closed IL-formula φ,
$w \Vdash \varphi$ if and only if $\bar{w} \Vdash \varphi$.

We'll carefully construct $\mathfrak{M}$ and w such that no $\bar{w}$ can exist.

But first we have to generalize characteristic formulas.

Finitely many types of formulas
of a given depth

For every finite set S of IL_0-formulas, we define the set of all perfect disjunctive normal forms over formulas from S:

$$D_0(S) := \left\{ \bigvee_{\psi \in K} \psi \ : \ K \subseteq \left\{ \bigwedge_{\varphi \in T} \varphi \wedge \bigwedge_{\varphi \in S \setminus T} \neg\varphi \ : \ T \subseteq S \right\} \right\}$$

(with conventions $\bigwedge \emptyset = \top$, $\bigvee \emptyset = \bot$), and proceed recursively:

$$D_{n+1}(S) := D_0\Big(D_n(S) \cup \big\{\varphi \rhd \psi \ : \ \varphi, \psi \in D_n(S)\big\}\Big);$$

then every $\varphi \in IL_0$ has an equivalent formula $\bar{\varphi} \in D_{\delta(\varphi)}(\emptyset)$.

Finitely many types of worlds
of a given depth

So, for every $n \in \mathbb{N}$ there are only finitely many closed *IL*-formulas of modal depth n, up to logical equivalence.

Direct consequence: for every n there are only finitely many *worlds* of depth n, up to modal equivalence.

Why? Because any two worlds of depth n
which aren't modally equivalent
must disagree on some formula φ of modal depth $\leq n$,
so they must disagee on $\bar{\varphi}$;
and there are finitely many ($m := \sum_{i=0}^{n} |D_i(\emptyset)|$) of them.
So there cannot be more than 2^m different "types" of worlds.

Universal frame for IL
a counterexample

Consider the Veltman model $\mathfrak{M}_{\lightning} := \big(W, R, (S_w)_{w \in W}\big)$, where

$$V := \{a, b, c, d, e, f, g\} \quad W := V \cup \{h\} \quad Q := \{cb, ca, ba\}$$

$$P := \{d, e, f, g\} \times \{a, b, c\} \quad R := \{h\} \times V \cup P \cup Q$$

$$S_a := \emptyset \quad S_b := \{aa\} \quad S_c := \{a, b\}^2 \quad S_e := id_{\{a,b,c\}} \cup Q$$

$$S_d := S_e \cup \{ab\} \quad S_f := S_e \quad S_g := S_e \cup \{bc\}$$

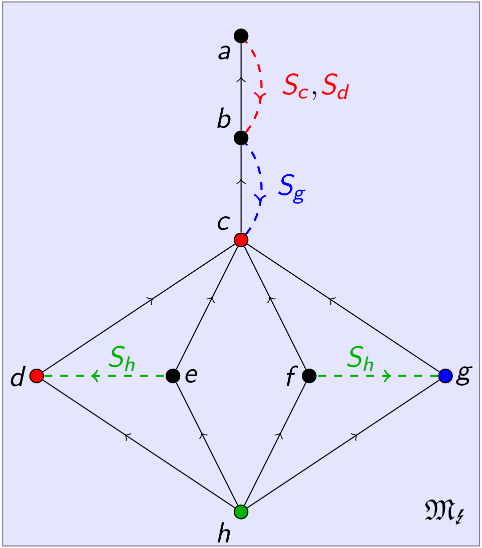
$$S_h := id_V \cup P \cup Q \cup \{ed, fg\}$$

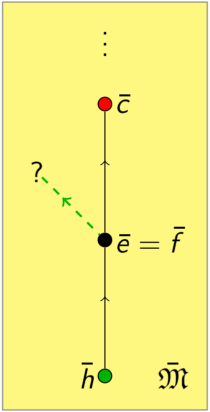
It seems complicated, and it *is* ☺, but not so much:
many of those edges are forced by conditions on R and S_w.

The following slide gives only the non-mandatory arrows.

Universal frame for *IL*

a counterexample — a picture





For every $w \in W$ we have a formula

$$\chi_w := \chi_{\rho(w)} \wedge \bigwedge_{\substack{\varphi \in S \\ w \Vdash \varphi}} \varphi \wedge \bigwedge_{\substack{\varphi \in S \\ w \nVdash \varphi}} \neg\varphi, \quad \text{where } S := D_{\rho(w)}(\emptyset),$$

which characterizes it completely.

Indeed, the first conjunct selects only worlds of depth $\rho(w)$, and then other conjuncts pinpoint exactly which "type" of world (of finitely many at that depth) w is.

Here we need only χ_a to χ_h, that is, only those pertaining to the frame $\mathfrak{M}_{\not{z}}$ — but the whole characteristic formulas theory can be applied to any world of finite depth in any frame.

Modal (non)equivalences

It's easy to see that the worlds d and g aren't modally equivalent: $d \Vdash \chi_a \rhd \chi_b$, while $g \nVdash \chi_a \rhd \chi_b$.

Also easy is to see that e and f are modally equivalent: they have the same R-successors, and the same S-relations.

- First consequence (world uniqueness in universal model): $\bar{e}$ and $\bar{f}$ is the same world in $\bar{\mathfrak{M}}$; denote it by t.
- Second consequence: χ_e and χ_f are equivalent formulas; denote their common $D_3(\emptyset)$-equivalent, $\bar{\chi_e} = \bar{\chi_f}$, by ψ.

ψ doesn't hold on any other world within $\mathfrak{M}_{\not{t}}$ (besides e and f): the only others of the same depth (3) are d and g, and they are eliminated by $\neg(\chi_a \rhd \chi_b)$ and $\neg(\chi_b \rhd \chi_c)$ respectively.

Contradictions
from the h's side

So, at the world h we have both

- χ_4 (formula that says h is at depth 4);
- $\psi \rhd (\chi_d \vee \chi_g)$: since every ψ-validating successor of h is either e or f, and each of those has a S_h-neighbor (respectively d and g) which validates $\chi_d \vee \chi_g$;

but we don't have either of

- $\psi \rhd \chi_d$: since $h\,R\,f \Vdash \psi$, but d (the only world in W validating χ_d) is not an S_h-neighbor of f;
- $\psi \rhd \chi_g$ (analogously, since $h\,R\,e \not S_h\, g$);

We claim that such a world is impossible in $\bar{\mathfrak{M}}$: $\bar{h}$ cannot exist.

Contradictions
from the $\bar{h}$'s side

The fact that $\bar{h} \nVdash \psi \rhd \chi_d$ means that there's a R-successor of h validating ψ, but such that no $S_{\bar{h}}$-neighbor of it validates χ_d. However, since $\psi \in D_3(\emptyset)$ is a characteristic formula (of worlds e and f), that successor must be $\bar{e} = t$ (by world uniqueness).

Recapitulating, we have: $t \; S_{\bar{h}} \; u$ implies $u \nVdash \chi_d$.
Completely analogously, we have: $t \; S_{\bar{h}} \; u$ implies $u \nVdash \chi_g$.

But $\bar{h} \Vdash \psi \rhd (\chi_d \vee \chi_g)$ means that for every R-successor of $\bar{h}$ validating ψ (so particularly for t), there exists its $S_{\bar{h}}$-neighbor u validating $\chi_d \vee \chi_g$. By previous conclusions, such u cannot exist.

Further research

- The key problem is world uniqueness in the universal model. Can we weaken that condition and still obtain something useful? (Probably yes, but not much.)
- There are various extensions of *IL*. Some of those, such as *ILF*, are coinciding with *GL* in their closed fragment — so they do have universal frame. Where's the boundary?
- There are other semantics for *IL*, most notably *generalized Veltman semantics*. Is there a universal generalized Veltman frame?
- Characteristic formulas are horribly long (iterated-exponentials long). Are there polynomial characteristic formulas? (Probably yes.)