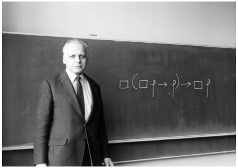
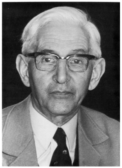
Arithmetical Interpretations
Löb meets Heyting

Albert Visser

Philosophy, Faculty of Humanities, Utrecht University

Lewis meets Brouwer, ESSLLI
Utrecht, July 29, 2021

Overview

An Ultra-brief History

Löb's Logic

Interpretability Logic

Intuitionistic Provability Logic

Universiteit Utrecht

Overview

An Ultra-brief History

Löb's Logic

Interpretability Logic

Intuitionistic Provability Logic

Overview

An Ultra-brief History

Löb's Logic

Interpretability Logic

Intuitionistic Provability Logic

Overview

An Ultra-brief History

Löb's Logic

Interpretability Logic

Intuitionistic Provability Logic

Overview

An Ultra-brief History

Löb's Logic

Interpretability Logic

Intuitionistic Provability Logic

History 1

The idea of interpreting modal logic in arithmetic goes (at least) back to a note of 1933 by Kurt Gödel on interpreting intuitionistic propositional logic in S4. Gödel notes that $\Box(\Box p \rightarrow p)$ does not hold if one interprets $\Box$ as provability in a formal system U.

The development received a forceful impetus when Martin Löb in 1955 formulated the purely modal Löb Conditions as a precondition for being a provability predicate. Leon Henkin who was the referee suggested Löb's Principle as a modal principle implicit in what Löb was doing. The resulting logic of Löb's Conditions with Löb's Principle is Löb's Logic GL.

GL has some beautiful properties like uniqueness and explicit definability of modalised points.

History 2

In 1976, Robert Solovay proves the arithmetical completeness of GL.

Around 1986, Albert Visser formulates axioms for a modal logic of interpretability. In 1988/1990 Volodya Shavrukov and Alessandro Berarducci independently prove the arithmetical completeness of the logic ILM. In 1990, Albert Visser proves the arithmetical completeness of the logic ILP. The dual reading of interpretability gives a Lewis arrow.

From 1980 on Albert Visser starts studying the provability logic of Heyting Arithmetic HA. In the context of this project the role of Σ^0_1-preservativity emerges: a typical Lewis arrow.

In 2018, Mohammad Ardeshir and Mojtaba Mojtahedi give a characterisation of the provability logic of Σ^0_1-substitutions. What the precise provability logic of HA is, is still a great open question.

Overview

The Logic

Löb's Logic is given by:

GL1. If $\vdash \varphi$, then $\vdash \Box\varphi$.

GL2. $\vdash \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$.

GL3. $\vdash \Box\varphi \to \Box\Box\varphi$.

GL4. $\vdash \Box(\Box\varphi \to \varphi) \to \Box\varphi$.

L1-3 are the Löb Conditions.

Löb's Rule is admissible for GL:

LR If $\vdash \Box\varphi \to \varphi$, then $\vdash \varphi$.

We can alternatively axiomatise GL by L1-3 plus LR or by L1,2,4. This last fact is due to Dick de Jongh.

Peano Arithmetic 1

Peano Arithmetic PA is given by the following axioms.

PA1. $\vdash Sx = Sy \to x = y$.

PA2. $\vdash Sx \neq 0$.

PA3. $\vdash x + 0 = 0$.

PA4. $\vdash x + Sy = S(x + y)$.

PA5. $\vdash x \times 0 = 0$.

PA6. $\vdash x \times Sy = x \times y + x$.

PA7. $\vdash (A0 \wedge \forall x\,(Ax \to ASx)) \to \forall x\, Ax$.

Peano Arithmetic 1.5

Every purely arithmetical insight that you can think of spontaneously can be proved in PA. One needs deep insight to find principles that are independent of it.

You are recommended to search on internet for the Hydra Game and for Goodstein sequences.

Excercise:

1. Prove the associativity of addition.
2. Prove the commutativity of addition.

Peano Arithmetic 2

- A formula is Δ^0_0 if it only contains bounded quantifiers, i.e., quantifiers of the form $\forall x < t$, $\forall x \leq t$, $\exists x < t$, $\exists x \leq t$. Here e.g. $\forall x < t\, B$ is short for: $\forall x\,(x < t \rightarrow B)$. The variable x may not occur in t.
- A formula S is Σ^0_1 if it of the form $\exists x_0 \ldots \exists x_{n-1}\, S_0$, where S_0 is Δ_0.
- A formula P is Π^0_1 if it of the form $\forall x_0 \ldots \forall x_{n-1}\, P_0$, where P_0 is Δ_0.

A Σ^0_1-formula is classically provably equivalent to the negation of Π^0_1-formula.

Arithmetisation 1

We can code syntactic objects as numbers. What is more the numerical tracking functions of syntactic operations can be represented in arithmetic and their recursive definitions can be verified.

Suppose we order the binary strings length-first and use the corresponding numbering as Gödel numbering.

0	□	5	BA	10	ABB	15	AAAA
1	A	6	BB	11	BAA	16	AAAB
2	B	7	AAA	12	BAB	17	AABA
3	AA	8	AAB	13	BBA	18	AABB
4	AB	9	ABA	14	BBB	19	ABAA

Tracking function of concatenation:

$$\begin{array}{ccccc} \mathtt{AB} & \star & \mathtt{BA} & = & \mathtt{ABBA} \\ \downarrow & \downarrow & \downarrow & & \downarrow \\ 4 & \circledast & 5 & = & 21 \end{array}$$

Arithmetisation 2

We can define a numerical predicate prov that mimicks PA-provability.

We write:

- $\ulcorner A \urcorner$ for the Gödel number of A.
- $\underline{0} := 0$, $\underline{n+1} := \mathrm{S}\underline{n}$.
  $\underline{n}$ is the numeral of n.

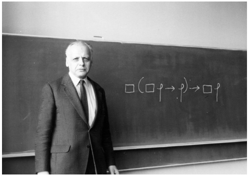
Arithmetical Interpretation

Suppose f maps propositional variables to arithmetical sentences. We define $(\cdot)^f$ from the modal to the arithmetical language as follows.

- $p^f := f(p)$,
- $(\cdot)^f$ commutes with the non-modal connectives.
- $(\Box\varphi)^f := \mathrm{prov}(\underline{\ulcorner \varphi^f \urcorner})$.
- $\Lambda_{\mathsf{PA}} := \{\varphi \mid \text{for all } f\text{, we have } \mathsf{PA} \vdash \varphi^f\}$.

We note that we have a case of *transubstantiation* here: a predicate of numbers is transformed into an operator on formulas.

Löb and Solovay



- Löb 1955: $\mathsf{GL} \subseteq \Lambda_{\mathsf{PA}}$
- Solovay 1976: $\Lambda_{\mathsf{PA}} \subseteq \mathsf{GL}$.

Löb's result holds for extensions of S^1_2. Solovay's result for Σ^0_1-sound extensions of EA. So, the results are remarkably stable.

Excercise: Prove the results of Löb and Solovay for the case where we restrict ourselves to ordinary propositional logic.

Overview

An Ultra-brief History

Löb's Logic

Interpretability Logic

Intuitionistic Provability Logic

Translation 1

We treat only a special case of interpretation that suffices for extensions of PA. We first use the term-elimination algorithm to reduce to relational signature.

- A translation τ of signature Θ to a signature Ξ is given by the following data:
- A Ξ-formula $P_\tau(v_0, \ldots, v_{n-1})$ for each n-ary Θ-predicate P.
- A domain formula $\delta_\tau(v_0)$.

We lift the translation to the full Θ-language as follows.

- $(P(x_0, \ldots, x_{n-1}))^\tau := P_\tau(x_0, \ldots, x_{n-1})$.
- $(\cdot)^\tau$ commutes with the propositional connectives.
- $(\forall x\, A)^\tau := \forall x\, (\delta_\tau(x) \to A^\tau)$, $(\exists x\, A)^\tau := \exists x\, (\delta_\tau(x) \wedge A^\tau)$.

Translation 2

Term Elimination: $x + (y \times z) = y + 0$ translates to:

$$\exists u \, \exists v \exists w \, (\mathrm{M}(y, z, u) \wedge \mathrm{A}(x, u, v) \wedge \mathrm{Z}(w) \wedge \mathrm{A}(y, w, v)).$$

Interpretability

We define:

- $V \rhd U$ iff, for some $\tau : \Theta_U \to \Theta_V$, for all A such that $U \vdash A$, we have $V \vdash A^\tau$.
- $A \rhd_U B$ iff $(U + A) \rhd (U + B)$.

We zoom in on the case where $U := \mathsf{PA}$.

Theorem

Over PA*, interpretability is* Π^0_1*-conservativity:*
$A \rhd_{\mathsf{PA}} B$ *iff, for all* $P \in \Pi^0_1$*, we have, if* $\mathsf{PA} + B \vdash P$*, then* $\mathsf{PA} + A \vdash P$*.*

Theorem (Orey)

There is a sentence O*, such that* $\top \rhd_{\mathsf{PA}} O$ *and* $\top \rhd_{\mathsf{PA}} \neg O$*.*

Interpretability Logic 1

The logic ILM is given by GL plus the following axioms:

J1 $\vdash \Box(\varphi \to \psi) \to (\varphi \rhd \psi)$

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

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

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

J5 $\vdash \Diamond \varphi \rhd \varphi$

M $\vdash (\varphi \rhd \psi) \to ((\varphi \wedge \Box \chi) \rhd (\psi \wedge \Box \chi))$

We note that if we define $\Box \varphi$ as $\neg \varphi \rhd \bot$, then we can drop J4.

Ytilibaterpretni Logic

If we define $\varphi \mathrel{-\!\!\!3} \psi$ as $\neg\psi \rhd \neg\varphi$, we find after renaming:

J1 $\vdash \Box(\varphi \to \psi) \to (\varphi \mathrel{-\!\!\!3} \psi)$

J2 $\vdash ((\varphi \mathrel{-\!\!\!3} \psi) \wedge (\psi \mathrel{-\!\!\!3} \chi)) \to (\varphi \mathrel{-\!\!\!3} \chi)$

J3 $\vdash ((\varphi \mathrel{-\!\!\!3} \psi) \wedge (\varphi \mathrel{-\!\!\!3} \chi)) \to (\varphi \mathrel{-\!\!\!3} (\psi \wedge \chi))$

J4 $\vdash (\varphi \mathrel{-\!\!\!3} \psi) \to (\Box\varphi \to \Box\psi)$

J5 $\vdash \varphi \mathrel{-\!\!\!3} \Box\varphi$

M $\vdash (\varphi \mathrel{-\!\!\!3} \psi) \to ((\Box\chi \to \varphi) \mathrel{-\!\!\!3} (\Box\chi \to \psi))$

We note that $\Box\varphi$ is $\top \mathrel{-\!\!\!3} \varphi$.

Universiteit Utrecht

Interpretability Logic 2

ILM is arithmetically sound for interpretations which translate $\rhd$ as $\rhd_{\mathrm{PA}}$.

Frank Veltman provided his Veltman semantics for theories of Interpretability Logic amongst which ILM. Dick de Jongh and Frank Veltman (1990) provide completeness proofs for various logics.

Volodya Shavrukov (1988) and, independently, Alessandro Berarducci (1990) prove arithmetical completeness for ILM.

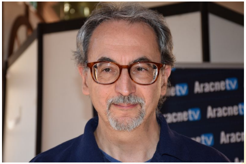
ILM has explicit unique modalised fixed points.

The existence of Orey sentences illustrates the invalidity of:

- $\vdash ((\varphi \rhd \psi) \wedge (\varphi \rhd \chi)) \to (\varphi \rhd (\psi \wedge \chi))$
- $\vdash ((\varphi \dashv\!\!\!\!\,\mathrm{3}\ \chi) \wedge (\psi \dashv\!\!\!\!\,\mathrm{3}\ \chi)) \to ((\varphi \vee \psi) \dashv\!\!\!\!\,\mathrm{3}\ \chi)$

De Jongh, Veltman, Shavrukov, Berarducci



Overview

An Ultra-brief History

Löb's Logic

Interpretability Logic

Intuitionistic Provability Logic

Brouwer and Heyting

Heyting Arithmetic 1

Heyting arithmetic is the same theory as Peano Arithmetic just with the logic changed to intuitionistic logic.

Caveat: Classically we can equivalently axiomatise Peano Arithmetic replacing Induction by the Minimum Principle or by Collection. Constructively, the minimum principle implies Excluded Third and Collection is much weaker than Induction.

To prove Excluded Third from the Minimum Principle consider the minimal element of $(x = 0 \wedge A) \vee x = 1$.

Collection: $\vdash \forall x < a \exists y\, Axy \rightarrow \exists b \forall x < a \exists y < b\, Axy$.
Compare with set theory:
$\vdash \forall x \in a \exists y\, Axy \rightarrow \exists b \forall x \in a \exists y \in b\, Axy$.

Slogan

SAY *NO* TO NEGATIVITY!

Do not say: If a theory is consistent then it does not prove its own consistency.

Do say: If a theory is proves its own consistency, then it is inconsistent.

The right formulation immediately makes clear that we can transform a consistency-proof in a proof of falsum. Inspection shows that this transformation is p-time.

Heyting Arithmetic 2

- For any Δ^0_0-formula D, we have $\mathrm{HA} \vdash D \vee \neg D$.
- For any Π^0_1-formula P, there is a Σ^0_1-formula S, such that $\mathrm{HA} \vdash P \leftrightarrow \neg S$.

  We have $\mathrm{HA} \vdash \neg\neg P \to P$.
- Consider e.g. the Σ^0_1-sentence $\Box_{\mathrm{HA}} \bot$. There is no sentence A such that $\mathrm{HA} \vdash \Box_{\mathrm{HA}} \bot \leftrightarrow \neg A$.

  If we had that, then $\mathrm{HA} \vdash \neg\neg \Box_{\mathrm{HA}} \bot \to \Box_{\mathrm{HA}} \bot$. *Quod non*.
- All known natural sentences independent of PA are connected to strength. We have many principles, like the logical form of Church's Thesis that are independent but are not connected to strength. Compare this to ZF and the Continuum Hypothesis.

The Disjunction Property 1

Constructive theories like HA (often) have the Disjunction Property DP:

- If $\mathrm{HA} \vdash A \vee B$, then $\mathrm{HA} \vdash A$ or $\mathrm{HA} \vdash B$.

Harvey Friedman shows that over HA the Disjunction Property plus Consistency is equivalent to Σ^0_1-reflection. Moreover, the Disjunction Property implies the numerical Existence Property.

- If $\mathrm{HA} \vdash \exists x\, Ax$, then, for some n, we have $\mathrm{HA} \vdash A\underline{n}$.

We have $\mathrm{HA} \nvdash \Box_{\mathrm{HA}}(A \vee B) \to (\Box_{\mathrm{HA}} A \vee \Box_{\mathrm{HA}} B)$. The best approximation to the disjunction property is Leivant's Principle:

- $\mathrm{HA} \vdash \Box_{\mathrm{HA}}(A \vee B) \to \Box_{\mathrm{HA}}(A \vee \Box_{\mathrm{HA}} B)$

The Disjunction Property 2

Note that:

$$\begin{aligned}\mathsf{HA} \vdash \Box_{\mathsf{HA}}(\Box_{\mathsf{HA}} \bot \vee \neg \Box_{\mathsf{HA}} \bot) &\to \Box_{\mathsf{HA}}(\Box_{\mathsf{HA}} \bot \vee \Box_{\mathsf{HA}} \neg \Box_{\mathsf{HA}} \bot) \\ &\to \Box_{\mathsf{HA}}(\Box_{\mathsf{HA}} \bot \vee \Box_{\mathsf{HA}} \bot) \\ &\to \Box_{\mathsf{HA}} \Box_{\mathsf{HA}} \bot\end{aligned}$$

So HA does not prove Excluded Third for $\Box_{\mathsf{HA}} \bot$. Also we see that the provability logic of a theory cannot be monotonic in the theory.

Iemhoff, Ardeshir, Mojtahedi, Zoethout




Intuitionistic Provability Logic 1

Classical provability logic is remarkably stable. We have the same logic, to wit GL for all theories T extending Elementary Arithmetic that are Σ^0_1-sound.

The intuitionistic case stands in stark contrast to that. We *do* have iGL, the intuitionistic version of GL. However, we have a wild array of possible further principles, e.g., we have in Λ_{HA}:

- $\vdash \Box(\varphi \vee \psi) \to \Box(\varphi \vee \Box\psi)$
- $\vdash \Box\neg\neg\Box\varphi \to \Box\Box\varphi$
- $\vdash \Box(\neg\neg\Box\varphi \to \Box\varphi) \to \Box\Box\varphi$

We note that Λ_{HA} trivialises if we add classical logic in the sense that it proves $\Box\Box\bot$.

Intuitionistic Provability Logic 1.5

What happens when we restrict ourselves to the non-modal language?

De Jongh's theorem tells us that we get precisely IPC. This result is not at all trivial.

There is a salient extension of HA, to wit $\mathrm{HA} + \mathrm{MP} + \mathrm{ECT}_0$ of which the propositional logic is unknown. It could even turn out to be complete Π^0_2.

Markov

We define $(\cdot)^A$ as follows:

- $B^A := (B \vee A)$ of B is atomic.
- $(\cdot)^A$ commutes with all connectives except the 0-ary ones.

We have:

- If $\mathsf{HA} \vdash B$, then $\mathsf{HA} \vdash B^A$.
- $\mathsf{HA} \vdash S^A \leftrightarrow (S \vee A)$, for $S \in \Sigma^0_1$.

Suppose $\mathsf{HA} \vdash \neg\neg S$. Then, $\mathsf{PA} \vdash S$. So, S is true and, hence, by Σ^0_1-completeness, $\mathsf{HA} \vdash S$.

Better: Suppose $\mathsf{HA} \vdash \neg\neg S$. Then $\mathsf{HA} \vdash (\neg\neg S)^S$. So, $\mathsf{HA} \vdash ((S \vee S) \to S) \to S$. Ergo, $\mathsf{HA} \vdash S$.
This gives a p-time proof transformation.

We find: $\mathsf{HA} \vdash \Box_{\mathsf{HA}} \neg\neg S \to \Box_{\mathsf{HA}} S$.

Anti-Markov

Suppose HA $\vdash \neg\neg S \to S$. Then, HA $\vdash (\neg\neg S \to S)^{\neg S}$. So,

$$\mathrm{HA} \vdash (((S \vee \neg S) \to \neg S) \to \neg S) \to (S \vee \neg S).$$

Ergo, HA $\vdash S \vee \neg S$.
Again a p-time proof transformation.

We find: HA $\vdash \Box_{\mathrm{HA}}(\neg\neg S \to S) \to \Box_{\mathrm{HA}}(S \vee \neg S)$.

$$\begin{aligned}
\mathrm{HA} \vdash \Box_{\mathrm{HA}}(\neg\neg \Box_{\mathrm{HA}} A \to \Box_{\mathrm{HA}} A) \quad &\to \quad \Box_{\mathrm{HA}}(\Box_{\mathrm{HA}} A \vee \neg \Box_{\mathrm{HA}} A) \\
&\to \quad \Box_{\mathrm{HA}}(\Box_{\mathrm{HA}} A \vee \Box_{\mathrm{HA}} \neg \Box_{\mathrm{HA}} A) \\
&\to \quad \Box_{\mathrm{HA}}(\Box_{\mathrm{HA}} A \vee \Box_{\mathrm{HA}} \bot) \\
&\to \quad \Box_{\mathrm{HA}} \Box_{\mathrm{HA}} A
\end{aligned}$$

Intuitionistic Provability Logic 2

- Let CP be the principle $\varphi \to \Box\varphi$. CP does not trivialise over iGL. We know an extension, say U, of HA such that $\Lambda_U = \text{iGL} + \text{CP}$. (Jetze Zoethout en Albert Visser)
- We know the closed fragments of Λ_{HA}, $\Lambda_{\text{HA+MP}}$, $\Lambda_{\text{HA+CP}_{\text{HA}}}$. The closed fragment of $\Lambda_{\text{HA+ECT}_0}$ is *terra incognita*.
- Mohammad Ardeshir and Mojtaba Mojtahedi recently characterised the provability logic of Σ^0_1-substitutions of HA. Next stage: boolean combinations of Σ^0_1-sentences.
- The (verifiably) admissible rules of HA are the admissible rules of IPC (Albert Visser for the non-verifiable case and Rosalie Iemhoff for the verifiable case).
- We have a current conjecture of what Λ_{HA} could be, but it could even turn out to be Π^0_2-complete.

Σ^0_1-preservativity 1

We define:

- $A \dashv_{\mathsf{HA}} B :\leftrightarrow \forall S \in \Sigma^0_1 \, (\Box_{\mathsf{HA}}(S \to A) \to \Box_{\mathsf{HA}}(S \to B))$.

Classically, Σ^0_1-preservativity can be transformed into Π^0_1-conservativity, but, constructively, it cannot.

Σ^0_1-preservativity turned out to be a useful notion to study the provability logic of HA. Also, perhaps, in the richer language the axiomatisation could be easier.

The most amazing thing of Preservativity Logic is that it satisfies full LHC$^\sharp$.

Σ_1^0-preservativity 2

We have an analogue of the Orey sentence. We have:

$$O \dashv_{\mathsf{HA}} \bot \text{ and } \neg O \dashv_{\mathsf{HA}} \bot.$$

As a consequence, by Di, we have:

$$(O \vee \neg O) \dashv \bot.$$

So while $\vdash (O \vee \neg O) \to \bot$ is impossible on pain of inconsistency, we do have $\vdash (O \vee \neg O) \dashv \bot$. Thus, this interpretation of the Lewis Arrow can illustrate the invalidity of ET in the most direct way.

Thank You

Universiteit Utrecht