Provability Logic

Benjamin Church

May 18, 2021

Contents

1 Introduction

1.1 First-Order Languages

.

Definition: A *vocabulary* or *signature* σ is a set of "non-logical" symbols which may be of three types:

1. Constant symbols (e.g. 0)

2. n-ary function symbols (e.g. $+$)

3. n-ary relation symbols (e.g. $\in$)

Along with the signature, a first-order language has a set of "logical" symbols:

1. A countable list of variable symbols: $x_1, x_2, x_3, \cdots$

2. Logical connectives: $\neg, \vee, \wedge, \rightarrow$

3. Quantifiers: $\forall$ (we get $\exists \iff \neg\forall\neg$ for free)

4. An equality relation: $=$

5. Punctuation: ( ) , etc.

Definition: The set of *terms* of a first-order language L with vocabulary σ is defined inductively as follows:

1. Any variable or constant symbol is a term.

2. If f is an n-ary function symbol and $t_1, \ldots, t_n$ are terms then $f(t_1, \ldots, t_n)$ is a term. For a binary operator (2-ary function), say $\circ$, we will often write $(t_1 \circ t_2)$ to mean $\circ(t_1, t_2)$.

Definition: The set of *formulas* of a first-order language L with vocabulary σ is defined inductively as follows:

1. If s, t are terms then $(s = t)$ is a formula. Furthermore if $R \in \sigma$ is an n-ary relation symbol and $t_1, \ldots, t_n$ are terms then $R(t_1, \ldots, t_n)$ is a formula. For a 2-ary relation we will often write sRt to mean $R(s, t)$.

2. If A and B are formulas then $\neg A$, $(A \vee B)$, $(A \wedge B)$, and $(A \rightarrow B)$ are all formulas.

3. If x is a variable symbol and φ a formula in which x is free (φ contains x but no quantifiers over x) then $\forall x\ \varphi$ and $\exists x\ \varphi$ are formulas.

Definition: A *sentence* of a first-order language is a formula with no free variables.

Definition: A first-order theory is a first-order language L along with a set Γ of first-order L-sentences which are referred to as axioms.

1.2 Proof Theory

There are many possible first-order deduction systems each with its own unique flavor. A deduction system has logical axioms and rules of inference on formulas of L. A formal proof beginning with some assumptions is a sequence of L-formulas each of which is either a logical axiom, an assumption, or the result of a rule of inference applied to previous formulas. Here we work with an example which is a variant of Hilbert's propositional logic formal system H extended to first order logic.

Definition: Hilbert's system H has logical connectives $\{\neg, \to\}$ and the following axiom schemas: for any formulas A, B, C the following are axioms of H,

(H1) $A \to (B \to A)$

(H2) $(A \to (B \to C)) \to ((A \to B) \to (A \to C))$

(H3) $\neg A \to (A \to B)$

(H4) $(\neg A \to A) \to A$

The formal system H has one rule of inference known as modus ponens (MP),

$$\frac{A \quad (A \to B)}{B}$$

We can define the formulas $A \vee B$ to stand for $\neg A \to B$ and $A \wedge B$ to stand for $\neg(A \to \neg B)$ and $A \leftrightarrow B$ to stand for $(A \to B) \wedge (B \to A)$ etc.

Definition: We say that a first-order theory Γ *syntactically entails* or, more simply, *proves* A if there exists a formal proof using axioms of Γ and first-order rules of inference. We write this as $\Gamma \vdash A$.

Example 1.1. We show that $\vdash_{\mathrm{H}} A \to A$ for any formula A.

(1)	$[A \to ((A \to A) \to A)] \to [(A \to (A \to A)) \to (A \to A)]$	axiom (L2)
(2)	$A \to ((A \to A) \to A)$	axiom (L1)
(3)	$(A \to (A \to A)) \to (A \to A)$	MP 1,2
(4)	$A \to (A \to A)$	axiom (L1)
(5)	$A \to A$	MP 3,4

Remark 1.2.1. Clearly, proofs in H are horrible. Luckily the following wonderful theorem means we will rarely need to provide explicit proofs.

Theorem 1.1 (Gödel)**.** *Every propositional tautology is a theorem of H.*

Remark 1.2.2. By tautology here we mean always evaluates to true under the standard semantics for $\neg$ and $\to$. In these semantics all axioms of H are tautologies and modus ponens is locally sound.

Definition: The formal system FO extends H by adding the additional axioms,

(EQ1) $\forall x\,(x = x)$

(EQ2) $\forall x\,[(x = t) \to (A(x) \to A(t))]$

(FO1) $\forall x\,A(x) \to A(t)$ where t is any term whose variables are not bound in A.

(FO2) $\forall x\,(A \to B(x)) \to (A \to \forall x\,B(x))$ where x is free in x and not in A.

and the additional rule of inference called generalization (Gen),

$$\frac{A}{\forall x\,A}$$

Remark 1.2.3. We will use the notation $A(x)$ to denote that x is a free variable in A and then $A(t)$ to denote A with t substituted for x.

Definition: A first-order theory Γ is *consistent* if there does not exist a statement A such that $\Gamma \vdash A$ and $\Gamma \vdash \neg A$.

Definition: A first-order theory Γ is *complete* if for every L-sentence A we have either $\Gamma \vdash A$ or $\Gamma \vdash \neg A$.

Lemma 1.2 (Categorization of Consistency)**.** *Γ is proof-theoretically consistent if and only if there exists a first-order sentence A such that $\Gamma \nvdash A$.*

Proof. If Γ is consistent and $\Gamma \vdash A$ then $\Gamma \nvdash \neg A$. If Γ is not consistent then $\Gamma \vdash A$ and $\Gamma \vdash \neg A$ for some A. Using (H3), $\Gamma \vdash \neg A \to (A \to B)$ so applying MP twice gives $\Gamma \vdash B$ for any B. □

2 A Theory For the Natural Numbers

2.1 The Language of Number Theory

Definition: The first-order language L_{NN} has signature $\sigma = \{\mathbf{0}, s, +, \cdot\}$ where $\mathbf{0}$ is a constant symbol, s is a 1-ary function symbol, and $+$ and $\cdot$ are 2-ary function symbols.

Example 2.1. We be define the abbreviation $x < y$ to mean $\exists z\,(x + s(z) = y)$.

Definition: For each natural number $i \in \mathbb{N}$ we denote the term $s^i(\mathbf{0})$ by the bold-face numeral $\mathbf{i}$.

2.2 Robinson Arithmetic and Peano Arithmetic

Now that we have a first-order language in which to do number theory, we need an actual theory.

Definition: Robinson Arithmetic, denoted as $\mathbf{Q}$, is the first-order theory over L_{NN} with the set of axioms,

(Q1) $\forall x\, \neg(s(x) = \mathbf{0})$

(Q2) $\forall x \forall y\, [(s(x) = s(y)) \to (x = y)]$

(Q3) $\forall x\, (x + \mathbf{0} = x)$

(Q4) $\forall x \forall y\, (x + s(y) = s(x + y))$

(Q5) $\forall x\, (x \cdot \mathbf{0} = \mathbf{0})$

(Q6) $\forall x \forall y\, (x \cdot s(y) = (x \cdot y) + x)$

(Q7) $\forall x\, [(x = \mathbf{0}) \vee \exists y\, (x = s(y))]$

Remark 2.2.1. We see that $\mathbf{Q}$ is arithmetic without induction. You might think that we cannot do very much in $\mathbf{Q}$ since it is a very weak theory. However $\mathbf{Q}$ is sufficiently powerful to cause its own essential incompleteness. In fact, $\mathbf{Q}$ is the minimal theory necessary to prove the representability theorem. For completeness, we will now define the more familiar framework for number theory.

Definition: Peano Arithmetic (**PA**) is the first-order theory over L_{NN} which has axioms (Q1) - (Q6) and additionally the axiom schema of induction,

(PA) $\varphi(\mathbf{0}) \to [\forall x\, (\varphi(x) \to \varphi(s(x))) \to \forall x\, \varphi(x)]$

for each formula φ with x free. Note we have dropped (Q7) since it is a consequence of the induction axiom.

Definition: An extension of $\mathbf{Q}$ is a first-order theory Γ over the language L_{NN} such that $\Gamma \vdash \mathbf{Q}$, in particular if $\Gamma \supset \mathbf{Q}$.

Remark 2.2.2. Clearly **PA** is an extension of $\mathbf{Q}$. In fact, the extension is proper.

2.3 Representing Functions and Relations

Definition: A relation $R \subset \mathbb{N}^n$ is *strongly representable* or simply *representable* in Γ, an extension of $\mathbf{Q}$ if there exists a formula $A(x_1, \ldots, x_n)$ in L_{NN} with n free variables such that for all natural numbers $a_1, \ldots, a_n \in \mathbb{N}$ we have,

$$\begin{aligned} R(a_1, \ldots, a_n) &\implies \Gamma \vdash A(\mathbf{a_1}, \ldots, \mathbf{a}_n) \\ \neg R(a_1, \ldots, a_n) &\implies \Gamma \vdash \neg A(\mathbf{a_1}, \ldots, \mathbf{a}_n) \end{aligned}$$

In this case we say that A *represents* R over Γ.

Definition: An arithmetic function $f : \mathbb{N}^n \to \mathbb{N}$ is representable over Γ iff there exists a formula $A(x_1, \dots, x_n, x_{n+1})$ of L_{NN} with $n+1$ free variables such that for all natural numbers $a_1, \dots, a_n \in \mathbb{N}$ with $b = f(a_1, \dots, a_n)$ we have,

$$\Gamma \vdash \forall x \, [A(\mathbf{a}_1, \dots, \mathbf{a}_n, x) \leftrightarrow (x = \mathbf{b})]$$

Remark 2.3.1. A function being representable is equivalent to its graph G_f being representable.

Definition: A relation $R \subset \mathbb{N}^n$ is *weakly representable* in Γ if there exists a formula $A(x_1, \dots, x_n)$ in L_{NN} with n free variables such that for all natural numbers $a_1, \dots, a_n \in \mathbb{N}$ we have,

$$R(a_1, \dots, a_n) \iff \Gamma \vdash A(\mathbf{a_1}, \dots, \mathbf{a}_n)$$

In this case we say that A *weakly represents* R over Γ.

Lemma 2.1. *If Γ is consistent then weak representability implies representability.*

Proof. It suffices to show that if $\Gamma \vdash A(\mathbf{a_1}, \dots, \mathbf{a}_n)$ then $R(a_1, \dots, a_n)$. Indeed, by consistency, $\Gamma \nvdash \neg A(\mathbf{a_1}, \dots, \mathbf{a}_n)$ so therefore $R(a_1, \dots, a_n)$. □

3 Computability Theory

We would like to construct representable functions. It turns out that there is a deep connection between computability and representability. More generally, the incompleteness theorems rely on arithmetic capturing the power of computable functions.

3.1 μ-Recursive Functions

The notion of *computability* or an *effective procedure* for computing a function is not a well-defined notion. We begin with a concrete definition for a class of clearly computable arithmetic functions. It turns out that in some sense these are *all* the computable functions.

Definition: An arithmetic function $F : \mathbb{N}^n \to \mathbb{N}$ is *recursive* if F is one of,

1. a starting function: addition ($(a, b) \mapsto a + b$), multiplication ($\cdot$), projection ($U_{n,k}(a_1, \dots, a_n) = a_k$), or less-then characteristic ($K_<(a, b) = 1$ if $a < b$ and zero otherwise).

2. a compositions of recursive functions $F = G \circ (H_1, \dots, H_k)$

3. a minimalization of a regular recursive function

$$F(a_1, \dots, a_n) = \mu x[G(a_1, \dots, a_n, x) = 0]$$

where the regularity condition on G means that such a zero is always required to exist for all natural numbers $a_1, \dots, a_n \in \mathbb{N}$.

3.2 Recursive and Recursively Enumerable Sets

Definition: A relation $R \subset \mathbb{N}^n$ is *recursive* (R) if there exists a recursive arithmetic function $f : \mathbb{N}^n \to \mathbb{N}$ such that $R = \{(a_1, \ldots, a_n) \in \mathbb{N}^n \mid f(a_1, \ldots, a_n) = 0\}$.

Definition: A relation $R \subset \mathbb{N}^n$ is *recursively enumerable* (RE) if R can be written as $R(a_1, \ldots, a_n) \iff \exists x\, Q(a_1, \ldots, a_n, x)$ where $Q \subset \mathbb{N}^n$ is a recursive relation.

Proposition. A set $S \subset \mathbb{N}$ is RE iff it is enumerated by a recursive function.

Remark 3.2.1. This proposition explains the terminology *recursively enumerable.*

3.2.1 Church-Turing Thesis

There is no clear universally agreed upon *a priori* definition for what it means for a function to be *effectively computable.* However, logicians Alonzo Church and Alan Turing proved that a wide class of models of computation (μ-recursive functions, Turning machines, λ-calculi) are all equivalently powerful. Therefore, we define *effectively computable* functions to be exactly those computable by any of these equivalent models of computation. Often, we will invoke this thesis to show that a given function is recursive if we can find an informal effective procedure for computing it. It should be stressed that such a use of the Church-Turing thesis is never necessary for proving meta-logical theorems it is simply a time-saving device for lazy logicians who don't want to explicitly construct recursive functions. It is only strictly necessary to invoke the Church-Turning thesis when computability is assumed as a hypothesis since we must develop a formal proof using some explicit model of computation.

3.3 The Representability Theorem

Theorem 3.1. *Let $f : \mathbb{N}^n \to \mathbb{N}$ be recursive function then f is representable over* $\mathbf{Q}$.

Proof. Very technical but conceptually easy. Show that all starting functions are representable and that given representable functions that we can construct representations of their composition and minimization. □

Corollary 3.2. *Let $R \subset \mathbb{N}^n$ be a recursive relation then R is representable over* $\mathbf{Q}$.

Proof. There exists a recursive $f : \mathbb{N}^n \to \mathbb{N}$ such that f vanishes exactly on R. Then f is representable by some L_{NN} formula $A(x_1, \ldots, x_{n+1})$ such that for all natural numbers $a_1, \ldots, a_n \in \mathbb{N}$ and $b = f(a_1, \ldots, a_n)$ then,

$$\mathbf{Q} \vdash \forall x\, [A(\mathbf{a}_1, \ldots, \mathbf{a}_n, x) \leftrightarrow (x = \mathbf{b})]$$

Let $B(x_1, \ldots, x_n) = A(x_1, \ldots, x_n, \mathbf{0})$. Then I claim that,

$$\begin{aligned} R(a_1, \ldots, a_n) &\implies \Gamma \vdash B(\mathbf{a_1}, \ldots, \mathbf{a}_n) \\ \neg R(a_1, \ldots, a_n) &\implies \Gamma \vdash \neg B(\mathbf{a_1}, \ldots, \mathbf{a}_n) \end{aligned}$$

and thus B represents R. □

4 Number Theory Swallows Itself

4.1 Gödel Numbering

We need some way of expressing the metalanguage of formulas and proofs inside of number theory such that we can use number theory to prove statements of its own meta-theory. This is accomplished by encoding formulas as natural numbers.

Theorem 4.1. *There exists an injective function* $\#_g : \mathrm{FOR}_{L_{\mathrm{NN}}} \to \mathbb{N}$ *such that its image* $S = \mathrm{Im}\, \#_g$ *is a recursive set.*

Proof. Consider encoding each symbol as a unique integer and then a sequence of symbols via $p_1^{a_1} \cdots p_n^{a_n}$ where p_i is the i^{th} prime and a_i is the code of the i^{th} symbol. By uniqueness of prime factorization, this function is injective. Checking its image is recursive is highly technical so I will simply invoke the Church-Turing thesis since there exists an effective procedure to factor a number, translate it into a string of symbols, and check if this string can be produced by the rules for forming well-formed formulas. The last step is effectively computable because there are a finite number of formulas of the correct length or less (restricting to only the variables which appear in the target string) so we can simply try each. □

Remark 4.1.1. The function $\#_g$ encodes each formula as a natural number such that the set of codes corresponding to well-formed formulas is computable.

Definition: Let A be a formula and $a = \#_g(A)$ its Gödel number. Then let $\ulcorner A \urcorner$ be the term $\mathbf{a}$.

Remark 4.1.2. This notation is intentionally suggestive of quotation in natural language. In fact, the Gödel sentence is not best described as saying "I am provable" but rather the Quine sentence,

"when preceded by its quotation is unprovable"
when preceded by its quotation is unprovable.

which is self-referential since the object of the sentence ("when preceded by its quotation is unprovable" when preceded by its quotation) is a copy of the entire sentence. This sentence accomplishes self-reference without the self-referential "machinery" of the pronoun "I" and therefore is a much better model for how such self-reference can unintentionally arise in number theory.

4.2 The Provability Predicate

Definition: A theory Γ with language L_{NN} is *recursively axiomatized* if $\#_g(\Gamma)$ is recursive.

Remark 4.2.1. Intuitively, a theory Γ is axiomatized if there exists an algorithm which can decide if a given string is an axiom of the theory.

Theorem 4.2. *Let Γ be recursively axiomatized. We may extend $\#_g^\Gamma : \mathrm{PRF}_\Gamma \to \mathbb{N}$ to encoding valid Γ-proofs as a sequence of formulas which, using the technique used above, we can encode in a single number. Again, we require that g_Γ be injective and have recursive image such that the codes of valid proofs comprise a computable set. Furthermore the relation, $\mathrm{CHKPRF}_\Gamma \subset \mathbb{N}^2$ defined to contain (a, p) iff a is the code of a valid formula and p is the code of a valid proof of the formula encoded by a is a recursive relation.*

Proof. We rely here on the Church-Turning thesis to show that such relations are recursive. They are effectively computable since checking a proof requires only checking each line to see if it is an axiom (which is decidable by hypothesis) or the result of applying one of finitely many rules of inference to the finitely many preceding sentences. This is clearly computable. □

Definition: Since CHKPRF_Γ is recursive it is Γ-representable. Let $\mathscr{Prf}_\Gamma(x, y)$ be a formula of L_{NN} such that,

$$\begin{aligned} \mathrm{CHKPRF}_\Gamma(a, p) &\implies \Gamma \vdash \mathscr{Prf}_\Gamma(\mathbf{a}, \mathbf{p}) \\ \neg\mathrm{CHKPRF}_\Gamma(a, p) &\implies \Gamma \vdash \neg\, \mathscr{Prf}_\Gamma(\mathbf{a}, \mathbf{p}) \end{aligned}$$

Definition: The provability predicate $\mathscr{Bew}_\Gamma(x)$ is the formula $\exists p\ \mathscr{Prf}_\Gamma(x, p)$.

Remark 4.2.2. The notation $\mathscr{Bew}$ derives from the German word *Beweis* for proof.

Lemma 4.3. *If $\Gamma \vdash A$ then $\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner A \urcorner)$.*

Proof. If $\Gamma \vdash A$ then there exists a proof of A which has code p and let A have code a. Therefore, $\mathrm{CHKPRF}_\Gamma(a, p)$ so $\Gamma \vdash \mathscr{Prf}_\Gamma(\mathbf{a}, \mathbf{p})$. Now the axiom (FO1) gives,

$$\Gamma \vdash \forall y \,\neg\, \mathscr{Prf}_\Gamma(\mathbf{a}, y) \to \neg\, \mathscr{Prf}_\Gamma(\mathbf{a}, \mathbf{p})$$

Thus, taking the contrapositive,

$$\Gamma \vdash \mathscr{Prf}_\Gamma(\mathbf{a}, \mathbf{p}) \to \mathscr{Bew}_\Gamma(\ulcorner A \urcorner)$$

so by modus ponens $\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner A \urcorner)$. □

Remark 4.2.3. We will see in the following sections that under the additionaly hypothesis of *ω-consistency* the provability predicate $\mathscr{Bew}_\Gamma(x)$ actually weakly represents theoremhood. However, we will also see that theoremhood is not strongly representable.

4.3 Self-Reference

Lemma 4.4 (Diagonalization). *Let $F(x)$ be an L_{NN} formula with one free variable. Then there exists a 'fixed-point' sentence ψ such that,*

$$\mathbf{Q} \vdash \psi \leftrightarrow F(\ulcorner \psi \urcorner)$$

Proof. There exists a recursive function $d : \mathbb{N} \to \mathbb{N}$ such that when $a = \#_g(A)$ where $A(x)$ is a formula with at least one free variable then $d(a) = \#_g(A(\mathbf{a})) = \#_g(A(\ulcorner A \urcorner))$ (for now we appeal to the Church-Turing thesis). Therefore, D is represented by some formula $D(x, y)$ such that for all $a \in \mathbb{N}$ and $b = d(a)$ we have,

$$\mathbf{Q} \vdash \forall y \, [D(\mathbf{a}, y) \leftrightarrow (y = \mathbf{b})]$$

Now define the formula with one free variable,

$$\varphi := \forall y \, [D(x, y) \to F(y)]$$

Let $a = \#_g(\varphi)$ be its Gödel number and then substitute $\mathbf{a} = \ulcorner \varphi \urcorner$ for x in φ,

$$\psi := \varphi(\ulcorner \varphi \urcorner) := \forall y \, [D(\ulcorner \varphi \urcorner, y) \to F(y)]$$

The Gödel number of ψ is $q = \#_g(\varphi(\ulcorner \varphi \urcorner)) = d(a)$ so we apply the representation of d applied at $d(a) = q$,

$$\mathbf{Q} \vdash \forall y \, [D(\ulcorner \varphi \urcorner, y) \leftrightarrow (y = \ulcorner \varphi(\ulcorner \varphi \urcorner) \urcorner)]$$

Using the tautology,

$$\mathbf{Q} \vdash (A \leftrightarrow B) \to [(A \to C) \leftrightarrow (B \to C)]$$

we find,

$$\mathbf{Q} \vdash \forall y \, [D(\ulcorner \varphi \urcorner, y) \to F(y)] \leftrightarrow \forall y \, [(y = \ulcorner \varphi(\ulcorner \varphi \urcorner) \urcorner) \to F(y)]$$

Which we can write as,

$$\mathbf{Q} \vdash \varphi(\ulcorner \varphi \urcorner) \leftrightarrow F(\ulcorner \varphi(\ulcorner \varphi \urcorner) \urcorner)$$

and using $\psi := \varphi(\ulcorner \varphi \urcorner)$ we have,

$$\mathbf{Q} \vdash \psi \leftrightarrow F(\psi)$$

□

Remark 4.3.1. If we interpret $F(\ulcorner \psi \urcorner)$ to represent "the formula ψ has property F" then the diagonal lemma proves the existence of self-referential fixed points. The sentence $\psi \leftrightarrow F(\ulcorner \psi \urcorner)$ "says" that ψ is true if and only if ψ has property F. In other words, ψ has an interpretation as the sentence: "I have property F." As described earlier, the diagonal sentence is more accurately modeled in natural language as the Quine sentence,

"when preceded by its quotation has property F"
when preceded by its quotation has property F.

In fact, the above proof of the diagonalization lemma closely resembles the construction of a Quine sentence: we take a sentence which refers to its object applied to (preceded by) its own quotation and apply it to (preceding it by) its own quotation. The predicate $\varphi(x)$ encodes "x when applied to its quotation (Gödel number) has property F" and the self-referential statement ψ is exactly φ applied to its quotation.

4.4 Godel Incompleteness I

In this and the following sections, let $\perp$ stand for your favorite contradiction, say $(\mathbf{0} = \mathbf{1})$ or $(x = y) \land \neg(x = y)$ etc. Any choice is as good as any other as long as $\Gamma \vdash \perp$ implies that Γ is inconsistent (which the above certainly do).

Definition: A theory Γ is ω-consistent if for all formulas $A(x)$ with one free variable Γ cannot simultaneously prove $\exists x\, A(x)$ and $\neg A(\mathbf{n})$ for each natural number $n \in \mathbb{N}$.

Lemma 4.5. *ω-consistency implies consistency.*

Proof. For each formula with one free variable $A(x)$ either $\Gamma \nvdash \exists x\, A(x)$ or for some $n \in \mathbb{N}$ we have $\Gamma \nvdash \neg A(\mathbf{n})$. Therefore, there exists some formula that Γ cannot prove which implies that Γ is consistent. □

Lemma 4.6. *If Γ is ω-consistent and $\Gamma \nvdash A$ then $\Gamma \nvdash \mathscr{Bew}_\Gamma(\ulcorner A \urcorner)$. In particular,*

$$\Gamma \vdash A \iff \Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner A \urcorner)$$

meaning that $\mathscr{Bew}_\Gamma(x)$ weakly represents the theoremhood relation on formulas.

Proof. Suppose that $\Gamma \nvdash A$ and $a = g(A)$ is the Gödel number. Then for each $n \in \mathbb{N}$ we have $\neg\text{CHKPRF}_\Gamma(a, n)$ since there exist no valid proofs of A. Therefore we have $\Gamma \vdash \neg\, \mathscr{Prf}_\Gamma(\mathbf{a}, \mathbf{n})$ for each $n \in \mathbb{N}$ so by ω-consistency $\Gamma \nvdash \mathscr{Bew}_\Gamma(\ulcorner A \urcorner)$. □

Corollary 4.7. *If Γ is ω-consistent then Γ is consistent so $\Gamma \nvdash \perp$ and thus, by the previous lemma, $\Gamma \nvdash \mathscr{Bew}_\Gamma(\ulcorner \perp \urcorner)$ and thus $\Gamma \nvdash \mathscr{Bew}_\Gamma(\ulcorner \mathscr{Bew}_\Gamma(\ulcorner \perp \urcorner) \urcorner)$ etc.*

Theorem 4.8 (Gödel)**.** *Any ω-consistent recursively axiomatized extension of $\mathbf{Q}$ is incomplete. In particular, if Γ is a recursively axiomatized extension of $\mathbf{Q}$ then there exists a sentence $\mathscr{G}_\Gamma$ such that,*

1. *if Γ is consistent then $\Gamma \nvdash \mathscr{G}_\Gamma$*
2. *if Γ is ω-consistent then $\Gamma \nvdash \neg\mathscr{G}_\Gamma$.*

Proof. Let Γ be a consistent recursively axiomatized extension of $\mathbf{Q}$. Since Γ is recursively axiomatized $\mathscr{Prf}_\Gamma$ and $\mathscr{Bew}_\Gamma$ exist. The fixed-point theorem proves the existence of a sentence $\mathscr{G}_\Gamma$ such that,

$$\Gamma \vdash \mathscr{G}_\Gamma \leftrightarrow \neg\, \mathscr{Bew}_\Gamma(\ulcorner \mathscr{G}_\Gamma \urcorner)$$

Suppose that $\Gamma \vdash \mathscr{G}_\Gamma$ then $\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner \mathscr{G}_\Gamma \urcorner)$. However, using $\Gamma \vdash \mathscr{G}_\Gamma$ and the self-reference equivalence, $\Gamma \vdash \neg\, \mathscr{Bew}_\Gamma(\ulcorner \mathscr{G}_\Gamma \urcorner)$ contradicting the consistency of Γ.

Suppose that $\Gamma \vdash \neg\mathscr{G}_\Gamma$. By the consistency of Γ we cannot have $\Gamma \vdash \mathscr{G}_\Gamma$. However, $\Gamma \vdash \neg\mathscr{G}_\Gamma$ and self-reference shows that $\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner \mathscr{G}_\Gamma \urcorner)$. Because $\Gamma \nvdash \mathscr{G}_\Gamma$, this contradicts the ω-consistency of Γ. □

Remark 4.4.1. The sentence $\mathscr{G}_\Gamma$ expresses "I am not provable" through Quinian self-reference. This is captured formally through $\Gamma \vdash \mathscr{G}_\Gamma \leftrightarrow \neg\, \mathscr{Bew}_\Gamma(\ulcorner \mathscr{G}_\Gamma \urcorner)$.

4.5 Löb's Theorem

Remark 4.5.1. In this section we assume that Γ is a recursively axiomatized extension of **PA**.

Lemma 4.9 (Hilbert-Bernays-Löb). *The provability predicate satisfies,*

1. $\Gamma \vdash A \implies \Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner A \urcorner)$
2. $\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner A \to B \urcorner) \to (\mathscr{Bew}_\Gamma(\ulcorner A \urcorner) \to \mathscr{Bew}_\Gamma(\ulcorner B \urcorner))$
3. $\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner A \urcorner) \to \mathscr{Bew}_\Gamma(\ulcorner \mathscr{Bew}_\Gamma(\ulcorner A \urcorner) \urcorner)$

Proof. SKETCH THIS PROOF!!!! □

Remark 4.5.2. The first Hilbert-Bernays derivability condition states that $\mathscr{Bew}_\Gamma(x)$ *weakly* represents theoremhood (it cannot strongly represent it however as we shall show). The second condition states that modus ponens is *provably (within* Γ*)* a rule of inference of Γ. Finally, the third Hilbert-Bernays derivability condition is the formalization of the first property *within the system* Γ, saying that Γ can prove that if it can prove A then it can prove that it can prove A.

Theorem 4.10 (Löb). *If* $\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner A \urcorner) \to A$ *then* $\Gamma \vdash A$ *for any sentence* A.

Proof. Via the fixed point theorem applied to $\mathscr{Bew}_\Gamma(x) \to A$, there exists a sentence B such that,

$$\Gamma \vdash B \leftrightarrow (\mathscr{Bew}_\Gamma(\ulcorner B \urcorner) \to A)$$

Applying HB1 to one direction gives,

$$\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner B \to (\mathscr{Bew}_\Gamma(\ulcorner B \urcorner) \to A) \urcorner)$$

and then applying HB2 twice we deduce,

$$\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner B \urcorner) \to (\mathscr{Bew}_\Gamma(\ulcorner \mathscr{Bew}_\Gamma(\ulcorner B \urcorner) \urcorner) \to \mathscr{Bew}_\Gamma(\ulcorner A \urcorner))$$

However, HB3 gives,

$$\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner B \urcorner) \to \mathscr{Bew}_\Gamma(\ulcorner \mathscr{Bew}_\Gamma(\ulcorner B \urcorner) \urcorner)$$

and thus putting the previous two together,

$$\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner B \urcorner) \to \mathscr{Bew}_\Gamma(\ulcorner A \urcorner)$$

Now we use the hypothesis $\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner A \urcorner) \to A$ to get a proof,

$$\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner B \urcorner) \to A$$

but since $\mathscr{Bew}_\Gamma(\ulcorner B \urcorner) \to A$ is provably equivalent to B we find $\Gamma \vdash B$ so by HB1 $\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner B \urcorner)$ and thus $\Gamma \vdash A$ by modus ponens. □

Remark 4.5.3. This theorem is truly remarkable because it says that **Q** and all extensions are "maximally modest" in the sense that the do not "believe" in their own validity (i.e. a proof of A entails A) except for statements they already know to be true. Furthermore, it answers the fascinating question posed by Henkin.

Remark 4.5.4. After seeing Gödel's proof of the first incompleteness theorem Henkin asked about a subtle modification. What if we apply the fixed-point lemma not to $\neg\,\mathscr{Bew}_\Gamma(x)$ but to simply $\mathscr{Bew}_\Gamma(x)$? Then there would exist a sentence $\mathscr{H}$,

$$\Gamma \vdash \mathscr{H} \leftrightarrow \mathscr{Bew}_\Gamma(\ulcorner \mathscr{H} \urcorner)$$

This sentence has the interpretation "I am provable" which seems to convey no information at all! However, clearly for such a sentence we have,

$$\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner \mathscr{H} \urcorner) \to \mathscr{H}$$

and thus by Löb's theorem we get $\Gamma \vdash \mathscr{H}$. So in fact, such a Henkin sentence which asserts its own provability must actually be provable.

4.6 Godel Incompleteness II

Finally, Löb's theme gives us enough machinery to give an elegant proof of the second incompleteness theorem.

Definition: The sentence $\mathscr{Con}_\Gamma$ is given by $\neg\,\mathscr{Bew}_\Gamma(\ulcorner \bot \urcorner)$ which expresses the consistency of the theory Γ.

Remark 4.6.1. We have shown that if $\Gamma \vdash \neg\,\mathscr{Con}_\Gamma$ then Γ is not ω-consistent. However, we are about to show a much more interesting result.

Theorem 4.11 (Gödel)**.** *Let Γ be a consistent recursively axiomatized extension of* **Q** *then Γ cannot prove $\mathscr{Con}_\Gamma$.*

Proof. By Löb's theorem if $\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner \bot \urcorner) \to\bot$ then $\Gamma \vdash\bot$. However, $\mathscr{Bew}_\Gamma(\ulcorner \bot \urcorner) \to\bot$ is equivalent to $\mathscr{Con}_\Gamma$. Thus if $\Gamma \vdash \mathscr{Con}_\Gamma$ then $\Gamma \vdash\bot$ contradicting the consistency of Γ. Taking the contrapositive, $\Gamma \nvdash\bot \implies \Gamma \nvdash \mathscr{Con}_\Gamma$ i.e. the consistency of Γ implies that Γ cannot prove $\mathscr{Con}_\Gamma$. □

Remark 4.6.2. Gödel's second incompleteness theorem is often stated provocatively as: a theory's proof of its own consistency establishes its inconsistency. This makes sense because an inconsistent theory can prove anything including its own consistency.

4.7 Löb's Theorem Formalized inside Number Theory

Wonderfully, we can formalize the proof of Löb's theorem inside the system Γ so that we may apply Löb *inside* formal proofs.

Theorem 4.12 (Löb)**.** *For any sentence A of L_{NN},*

$$\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner \mathscr{Bew}_\Gamma(\ulcorner A \urcorner) \to A \urcorner) \to \mathscr{Bew}_\Gamma(\ulcorner A \urcorner)$$

Proof. Let $B := \mathscr{Bew}_\Gamma(\ulcorner \mathscr{Bew}_\Gamma(\ulcorner A \urcorner) \to A \urcorner)$ and $C := \mathscr{Bew}_\Gamma(\ulcorner A \urcorner)$. Then HB2 gives,

$$\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner B \to C \urcorner) \to (\mathscr{Bew}_\Gamma(\ulcorner B \urcorner) \to \mathscr{Bew}_\Gamma(\ulcorner C \urcorner))$$

Furthermore, since $B := \mathscr{Bew}_\Gamma(\ulcorner C \to A \urcorner)$,

$$\Gamma \vdash B \to (\mathscr{Bew}_\Gamma(\ulcorner C \urcorner) \to \mathscr{Bew}_\Gamma(\ulcorner A \urcorner))$$

and by HB3 (since B begins with $\mathscr{Bew}$),

$$\Gamma \vdash B \to \mathscr{Bew}_\Gamma(\ulcorner B \urcorner)$$

Given $\mathscr{Bew}_\Gamma(\ulcorner B \to C \urcorner)$ we get $\mathscr{Bew}_\Gamma(\ulcorner B \urcorner) \to \mathscr{Bew}_\Gamma(\ulcorner C \urcorner)$. Additionally, given B we get $\mathscr{Bew}_\Gamma(\ulcorner B \urcorner)$ so we get $\mathscr{Bew}_\Gamma(\ulcorner C \urcorner)$ but B also gives $\mathscr{Bew}_\Gamma(\ulcorner C \urcorner) \to \mathscr{Bew}_\Gamma(\ulcorner A \urcorner)$ so we get $C := \mathscr{Bew}_\Gamma(\ulcorner A \urcorner)$. Thus by propositional logic,

$$\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner B \to C \urcorner) \to (B \to C)$$

Therefore, applying Löb's theorem,

$$\Gamma \vdash B \to C$$

which, expanded out is,

$$\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner \mathscr{Bew}_\Gamma(\ulcorner A \urcorner) \to A \urcorner) \to \mathscr{Bew}_\Gamma(\ulcorner A \urcorner)$$

□

4.8 Gödel Incompleteness Formalized inside Number Theory

Much in the way that Löb theorem can be formalized inside number theory, we can formalize the proofs of the incompleteness theorems inside the formal system itself. In fact, we can further formalize the notion that consistency implies the unprovability of the Gödel sentence and thus its truth to give an alternative proof of the second incompleteness theorem and furthermore a demonstration of the provable logical equivalence of all Gödel sentences.

Theorem 4.13. *Let $\mathscr{G}_\Gamma$ be a Gödel sentence for Γ then,*

$$\Gamma \vdash \mathscr{Con}_\Gamma \leftrightarrow \mathscr{G}_\Gamma$$

In particular, all Gödel sentences are provably logically equivalent.

Proof. Since $\mathscr{G}_\Gamma$ is a Gödel sentence,

$$\Gamma \vdash \mathscr{G}_\Gamma \leftrightarrow \neg \mathscr{Bew}_\Gamma(\ulcorner \mathscr{G}_\Gamma \urcorner)$$

Therefore, applying HB1 and HB2,

$$\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner \mathscr{G}_\Gamma \urcorner) \leftrightarrow \mathscr{Bew}_\Gamma(\ulcorner \neg \mathscr{Bew}_\Gamma(\ulcorner \mathscr{G}_\Gamma \urcorner) \urcorner)$$

However, by HB3,

$$\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner \mathscr{G}_\Gamma \urcorner) \to \mathscr{Bew}_\Gamma(\ulcorner \mathscr{Bew}_\Gamma(\ulcorner \mathscr{G}_\Gamma \urcorner)\urcorner)$$

Furthermore, since $\Gamma \vdash \neg A \to (A \to \bot)$ by HB1 and HB2 twice we get,

$$\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner \neg A \urcorner) \to (\mathscr{Bew}_\Gamma(\ulcorner A \urcorner) \to \mathscr{Bew}_\Gamma(\ulcorner \bot \urcorner))$$

Applying this to $A := \mathscr{Bew}_\Gamma(\ulcorner \mathscr{G}_\Gamma \urcorner)$ we find,

$$\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner \mathscr{G}_\Gamma \urcorner) \to \mathscr{Bew}_\Gamma(\ulcorner \bot \urcorner)$$

However, $\Gamma \vdash \bot \to \mathscr{G}_\Gamma$ and thus applying HB1 and HB2 we find,

$$\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner \bot \urcorner) \to \mathscr{Bew}_\Gamma(\ulcorner \mathscr{G}_\Gamma \urcorner)$$

In summary,

$$\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner \mathscr{G}_\Gamma \urcorner) \leftrightarrow \mathscr{Bew}_\Gamma(\ulcorner \bot \urcorner)$$

However, $\Gamma \vdash \mathscr{G}_\Gamma \leftrightarrow \neg\, \mathscr{Bew}_\Gamma(\ulcorner \mathscr{G}_\Gamma \urcorner)$ and $\mathscr{Con}_\Gamma := \neg\, \mathscr{Bew}_\Gamma(\ulcorner \bot \urcorner)$ which implies that,

$$\Gamma \vdash \mathscr{G}_\Gamma \leftrightarrow \mathscr{Con}_\Gamma$$

□

Corollary 4.14. *If Γ is consistent then by Gödel incompleteness I we know $\Gamma \nvdash \mathscr{G}_\Gamma$ and thus $\Gamma \nvdash \mathscr{Con}_\Gamma$ giving an alternative proof of incompleteness II.*

Theorem 4.15 (Formalized Gödel I)**.**

$$\Gamma \vdash \omega\text{-}\mathscr{Con}_\Gamma \to (\neg\, \mathscr{Bew}_\Gamma(\ulcorner \mathscr{G}_\Gamma \urcorner) \land \neg\, \mathscr{Bew}_\Gamma(\ulcorner \neg \mathscr{G}_\Gamma \urcorner))$$

Where ω-$\mathscr{Con}_\Gamma$ is the sentence $\neg\, \mathscr{Bew}_\Gamma(\ulcorner \mathscr{Bew}_\Gamma(\ulcorner \bot \urcorner)\urcorner)$ expressing weak ω-consistency.

Proof. First, by HB3,

$$\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner \bot \urcorner) \to \mathscr{Bew}_\Gamma(\ulcorner \mathscr{Bew}_\Gamma(\ulcorner \bot \urcorner)\urcorner)$$

and therefore,

$$\Gamma \vdash \omega\text{-}\mathscr{Con}_\Gamma \to \mathscr{Con}_\Gamma$$

We have already proven above that,

$$\Gamma \vdash \mathscr{Con}_\Gamma \to \neg\, \mathscr{Bew}_\Gamma(\ulcorner \mathscr{G}_\Gamma \urcorner)$$

and thus by transitivity of implication,

$$\Gamma \vdash \omega\text{-}\mathscr{Con}_\Gamma \to \neg\, \mathscr{Bew}_\Gamma(\ulcorner \mathscr{G}_\Gamma \urcorner)$$

The negation of the Gödel property gives,

$$\Gamma \vdash \neg \mathscr{G}_\Gamma \leftrightarrow \mathscr{Bew}_\Gamma(\ulcorner \mathscr{G}_\Gamma \urcorner)$$

and thus by HB1 and HB2 we have,

$$\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner \neg \mathscr{G}_\Gamma \urcorner) \leftrightarrow \mathscr{Bew}_\Gamma(\ulcorner \mathscr{Bew}_\Gamma(\ulcorner \mathscr{G}_\Gamma \urcorner) \urcorner)$$

However, by HB3,

$$\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner \neg \mathscr{G}_\Gamma \urcorner) \rightarrow \mathscr{Bew}_\Gamma(\ulcorner \mathscr{Bew}_\Gamma(\ulcorner \neg \mathscr{G}_\Gamma \urcorner) \urcorner)$$

Furthermore via $\Gamma \vdash \neg \mathscr{G}_\Gamma \rightarrow (\mathscr{G}_\Gamma \rightarrow \bot)$ and HB1 and HB2 repeatedly we find,

$$\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner \mathscr{Bew}_\Gamma(\ulcorner \mathscr{G}_\Gamma \urcorner) \urcorner) \rightarrow (\mathscr{Bew}_\Gamma(\ulcorner \mathscr{Bew}_\Gamma(\ulcorner \neg \mathscr{G}_\Gamma \urcorner) \urcorner) \rightarrow \mathscr{Bew}_\Gamma(\ulcorner \mathscr{Bew}_\Gamma(\ulcorner \bot \urcorner) \urcorner))$$

and thus by transitivity of implications,

$$\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner \neg \mathscr{G}_\Gamma \urcorner) \rightarrow \mathscr{Bew}_\Gamma(\ulcorner \mathscr{Bew}_\Gamma(\ulcorner \bot \urcorner) \urcorner)$$

contradicting ω-consistency. That is, taking the contrapositive,

$$\Gamma \vdash \omega\text{-}\mathscr{Con}_\Gamma \rightarrow \neg \mathscr{Bew}_\Gamma(\ulcorner \neg \mathscr{G}_\Gamma \urcorner)$$

giving both implications which together show that,

$$\Gamma \vdash \omega\text{-}\mathscr{Con}_\Gamma \rightarrow (\neg \mathscr{Bew}_\Gamma(\ulcorner \mathscr{G}_\Gamma \urcorner) \wedge \neg \mathscr{Bew}_\Gamma(\ulcorner \neg \mathscr{G}_\Gamma \urcorner))$$

□

Remark 4.8.1. Just as we needed ω-consistency in the standard proof of Gödel Incompleteness I, in the formalized version we require a stronger hypothesis than $\mathscr{Con}_\Gamma$, we need $\neg \mathscr{Bew}_\Gamma(\ulcorner \mathscr{Bew}_\Gamma(\ulcorner \mathscr{G}_\Gamma \urcorner) \urcorner)$ which expresses the idea that ω-consistency requires Γ to be unable to prove that it can prove a contradiction. In fact this hypothesis is somewhat weaker than full ω-consistency so this is an abuse of notation.

Theorem 4.16 (Formalized Gödel II)**.**

$$\Gamma \vdash \mathscr{Con}_\Gamma \rightarrow \neg \mathscr{Bew}_\Gamma(\ulcorner \mathscr{Con}_\Gamma \urcorner)$$

Proof. Apply formalized Löb with $A = \bot$ to give,

$$\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner \mathscr{Bew}_\Gamma(\ulcorner \bot \urcorner) \rightarrow \bot \urcorner) \rightarrow \mathscr{Bew}_\Gamma(\ulcorner \bot \urcorner)$$

However, $\mathscr{Bew}_\Gamma(\ulcorner \bot \urcorner) \rightarrow \bot$ is logically equivalent to $\neg \mathscr{Bew}_\Gamma(\ulcorner \bot \urcorner)$ which is $\mathscr{Con}_\Gamma$. Furthermore if $\Gamma \vdash A \leftrightarrow B$ then $\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner A \urcorner) \leftrightarrow \mathscr{Bew}_\Gamma(\ulcorner B \urcorner)$ by HB1 and HB2 so we have,

$$\Gamma \vdash \mathscr{Bew}_\Gamma(\ulcorner \mathscr{Con}_\Gamma \urcorner) \rightarrow \mathscr{Bew}_\Gamma(\ulcorner \bot \urcorner)$$

which is exactly the contrapositive of formalized Gödel incompleteness II. □

5 GL Provability Logic

5.1 Modal Logic

Modal logics are formal systems given by standard predicate calculus with a modal predicate $\Box$ which expresses some form of "necessity." We first define the simplest so called "normal" modal logic **K** named for Saul Kripke.

Definition: The formal system **K** has logical connectives $\{\rightarrow, \neg, \Box\}$ and sentences are built from an infinite list of propositional variables $p, q, \dots$. It is defined by having as axioms,

1. tautologies of propositional calculus (say take Hilbert's system H for concrete axiomatization)
2. the modal distribution axiom (K),

$$\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$$

and as rules of inference has,

1. modus ponens (MP),

$$\frac{A \quad (A \rightarrow B)}{B}$$

2. the necessitation rule,

$$\frac{A}{\Box A}$$

Remark 5.1.1. **K** is the basis for most modal logics, however it is too weak to capture most modal notions. For proability logic we need a stronger extension which is named **GL** for Gödel and Löb.

Definition: The formal system **GL** is simply **K** with the added two axioms,

(4) $\Box A \rightarrow \Box\Box A$

(L) $\Box(\Box A \rightarrow A) \rightarrow \Box A$

5.2 Modal Semantics

5.3 Arithmetic Soundness

Definition: Let Γ be We define an *arithmetical realization* to be a logical map ϕ : **GL** $\rightarrow$ **PA** (that is a map on sentences which preserves logical connectives i.e. a morphism of the Boolean algebra of sentences) which satisfies the property that for any sentence A of **GL**,

$$\phi(\Box A) = \mathscr{Bew}_\Gamma(\ulcorner \phi(A) \urcorner)$$

Remark 5.3.1. When the realization ϕ is unambiguous we will often write A^* for $\phi(A)$ the realization of A.

Theorem 5.1 (Arithmetical Soundness)**.** *If* **GL** $\vdash A$ *then* **PA** $\vdash A^*$ *for any arithmetical realization* $*$.

Proof. This proceeds by induction of proofs. It suffices to show that the axioms of **GL** are realized by provably statments of **PA** and that rules of inference in **GL** are sound in **PA**. Since **PA** contains axiom schema for all propositional tautologies and the rule of inference MP we simply need to check the necessitation deduction,

$$\mathbf{PA} \vdash A^* \implies \mathbf{PA} \vdash (\Box A)^*$$

and modal axioms,

$$\begin{aligned} &\mathbf{PA} \vdash (\Box(A \to B) \to (\Box A \to \Box B))^* \\ &\mathbf{PA} \vdash (\Box A \to \Box\Box A)^* \\ &\mathbf{PA} \vdash (\Box(\Box A \to A) \to \Box A)^* \end{aligned}$$

Using the properties of $*$ we see this is equivalent to asking that,

$$\begin{aligned} &\mathbf{PA} \vdash A^* \implies \mathbf{PA} \vdash \mathscr{Bew}_\Gamma(\ulcorner A^* \urcorner) \\ &\mathbf{PA} \vdash \mathscr{Bew}_\Gamma(\ulcorner A^* \to B^* \urcorner) \to (\mathscr{Bew}_\Gamma(\ulcorner A^* \urcorner) \to \mathscr{Bew}_\Gamma(\ulcorner B^* \urcorner) \\ &\mathbf{PA} \vdash \mathscr{Bew}_\Gamma(\ulcorner A^* \urcorner) \to \mathscr{Bew}_\Gamma(\ulcorner \mathscr{Bew}_\Gamma(\ulcorner A^* \urcorner) \urcorner) \\ &\mathbf{PA} \vdash \mathscr{Bew}_\Gamma(\ulcorner \mathscr{Bew}_\Gamma(\ulcorner A^* \urcorner) \to A^* \urcorner) \to \mathscr{Bew}_\Gamma(\ulcorner A^* \urcorner) \end{aligned}$$

which are exactly the three Hilbert-Bernays derivability conditions and formalized Löb's theorem which have shown to be probable in **PA**. □

5.4 The Existence of Modal Fixed Points

5.5 Arithmetic Completeness

Theorem 5.2 (Arithmetical Completeness, Solovay, 1976)**.** *If* **PA** $\vdash A^*$ *for any arithmetical realization* $*$ *then* **GL** $\vdash A$.

Remark 5.5.1. This theorem is remarkable because it captures the overarching logic of **PA** in a modal logic based of propositional calculus without quantifiers. This result is made even more remarkable by the following decidability theorem for **GL**.

Theorem 5.3. *The theoremhood relation for* **GL** *is decidable i.e. the decision problem for* **GL** *is solvable.*

Remark 5.5.2. We know by Church and Turring that the decision problem for **PA** is unsolvable that there does not exist an algorithm which can decide theoremhood in **PA**. Therefore, it is suprising and powerful that we can capture probability logic inside **PA** with the *decidable* theory **GL**.

Remark 5.5.3. We will end with a application of arithmetic completeness to generating undecidable arithmetical sentences. It is not difficult to show that $\mathbf{GL} \nvdash \Box p \vee \neg \Box p$. Then Solovay's proof alows us to construct a realization such that $\mathbf{PA} \nvdash (\Box p \vee \neg \Box p)^*$. However,

$$(\Box p \vee \neg \Box p)^* = \mathscr{B}\mathit{ew}_\Gamma(\ulcorner p^* \urcorner) \vee \neg\, \mathscr{B}\mathit{ew}_\Gamma(\ulcorner p^* \urcorner)$$

so if $\mathbf{PA} \vdash p^*$ or $\mathbf{PA} \vdash \neg p^*$ then by HB1 we would have $\mathbf{PA} \vdash \mathscr{B}\mathit{ew}_\Gamma(\ulcorner p^* \urcorner)$ or $\mathbf{PA} \vdash \mathscr{B}\mathit{ew}_\Gamma(\ulcorner \neg p^* \urcorner)$ contradicting Solovay's construction. Thus p^* is an undecidable arithmetic sentence giving us further examples of oddities besides Gödel sentences.