Provability Logic and the Completeness Principle

Jetze Zoethout, 4014294

Master Thesis

January 9, 2018

Supervisors: dr. Rosalie Iemhoff and prof. dr. Albert Visser
Third reader: prof. dr. Lev D. Beklemishev
University: Utrecht University
Faculty: Humanities
Department: Philosophy and Religious Studies
Master programme: Philosophy
ECTS: 30

Contents

Contents **ii**

1 Introduction **1**

2 Prerequisites **3**

2.1 Arithmetic and Provability 3

2.2 The Completeness Principle 6

2.3 Fast and Slow Provability 12

2.4 Intuitionistic Modal Logic 16

3 An Abstract Arithmetical Completeness Theorem **21**

3.1 Definition of the Solovay Function 21

3.2 Preservation of the Logical Structure 24

3.3 The Completeness Theorem 28

4 Applications of the Completeness Theorem **31**

4.1 The Sets NNIL and TNNIL 31

4.2 Some Fast (Σ_1-)Provability Logics 35

4.3 The Σ_1-Provability Logic of HA 36

5 Conclusion **38**

Index of symbols **40**

Index of terms **41**

Bibliography **42**

Chapter 1

Introduction

Around 1930, Kurt Gödel proved his celebrated incompleteness theorems. While these results can be seen as the culmination of one era of logical research, they also cleared the way for several new fields within mathematical logic. An example of such a field is *provability logic*, a topic that still occupies logicians today. Provability logic takes one of the main ingredients of Gödel's theorems as its starting point. This ingredient is the formalization of the notion 'formally provable in a certain arithmetical theory T' inside the language of arithmetic itself. Once this step has been taken, one may wonder what a theory T is able to prove about its own notion of provability. This object, i.e. what a theory T can prove about its own notion of provability, is called the provability logic of T. Let us write, as we will below, '$\vdash_T A$' for 'A is formally provable in T', and '$\Box_T A$' for the arithmetical formula expressing that A is formally provable in T. Then under some reasonable assumptions, the following turn out to hold:

(i) if $\vdash_T A$, then $\vdash_T \Box_T A$;

(ii) $\vdash_T \Box_T(A \to B) \to (\Box_T A \to \Box_T B)$;

(iii) $\vdash_T \Box_T A \to \Box_T \Box_T A$.

These are known nowadays as the *Hilbert-Bernays-Löb derivability conditions*. Using the Diagonalization Lemma, another key idea from Gödel's theorems, one can derive from these that $\vdash_T \Box_T(\Box_T A \to A) \to \Box_T A$, a result known as Löb's Theorem. In 1976, Robert Solovay proved that for the theory *Peano Arithmetic*, the schemes (i)-(iii) and the formalized Löb Theorem completely describe the provability logic of Peano Arithmetic.

Provability logics are not monotone in their corresponding theories. That is, if T is a theory extending another theory U, then it is not in general true that the provability logic of T extends the provability logic of U. In light of this, it is all the more surprising that, in the classical case, provability logics are immensely stable. Solovay's proof can be modified to show that any Σ_1-sound theory interpreting Peano Arithmetic has the same provability logic as Peano Arithmetic. We will not explain exactly what this means, but it includes theories as strong as Zermelo-Fraenkel Set Theory (with or without the Axiom of Choice).

Peano Arithmetic is a classical theory, which is why we made the *caveat* 'in the classical case' below. In the intuitionistic case, the situation is completely different. Solovay's proof simply does not work for intuitionistic theories. This shows itself in the fact that the provability logic of *Heyting Arithmetic*, the intuitionistic counterpart of Peano Arithmetic, contains principles that the provability logic of Peano Arithmetic does not share. These principles are somewhat exotic, and it is unknown what the provability logic of Heyting Arithmetic exactly is. In fact, as far as we are aware, there is presently only one intuitionistic theory for which a nontrivial provability logic is known, a result due to A. Visser (see Remark 4.2.1).

In Solovay's proof, the semantics of (classical) modal logic play a major role. The larger part of the proof consists of embedding models for modal logic in a certain way into the theory T. These models are equipped with an accessibility relation. Solovay uses the predicate $\Box_T$ to represent this relation inside the theory T. One may try to give a Solovay-style proof by replacing the models for classical modal logic by models for intuitionistic modal logic. The difficulty about these models, however, is that they also possess an intuitionistic relation, in addition to the accessibility relation. The main question then becomes how we can deal with these two relations.

This question sets the main goal for this thesis: to find an interesting situation where we can give a Solovay-style embedding of a model for intuitionistic modal logic. A. Visser suggested to consider theories that prove their own *completeness principle*, a principle for which the modal semantics is not too complicated. The availability of the completeness principle is an advantage that is specific to the intuitionistic context, since in the classical context, the completeness principle trivializes questions about provability logic. Furthermore, Visser suggested to use a nonstandard notion of provability to interpret one of the two relations on our models. This approach turned out to be successful, and our Solovay-style embedding is presented in detail below. Our Solovay-style embedding can be used to obtain a variety of results in provability logic. Among these is the determination of the Σ_1-*provability logic of Heyting Arithmetic*, an object related to the ordinary provability logic of Heyting Arithmetic. This is not a new result. It was already obtained in 2014 by M. Ardeshir and S. Mojtaba Mojtahedi [1], but the present work arrives at it in a different way. We stress, however, that our proof could not have been devised without the work from the paper [1]. First of all, it is of course easier to determine a provability logic if one already knows what it should be. Moreover, even though our proof is different, we do use some key ingredients from the paper [1], most notably the TNNIL-algorithm.

Let us briefly outline the structure of the thesis. First of all, in Chapter 2, we discuss all the necessary prerequisite knowledge, and fix our notation. This chapter contains no essentially new results, but we do prove some results from the paper [10] under weaker assumptions. For reasons of space (and energy), we will not spell out any specific Gödel numberings or give an explicit definition of the predicate $\Box_T$. Therefore, it will be useful to have some experience with Gödel's incompleteness theorems and with provability logic (in the classical case) when reading this thesis. A reader that is already familiar with (some of) the concepts discussed in Chapter 2 may want to read (a portion of) this chapter only superficially, and refer back to it if necessary. In Chapter 3, we present our Solovay-style embedding, and formulate a completeness theorem. This theorem will be stated in an abstract way that does not yet mention any specific theories or provability predicates. In Chapter 4, we will present several applications of our completeness theorem, among which the determination of the Σ_1-provability logic of Heyting Arithmetic.

Chapter 2

Prerequisites

In this chapter, we develop some notation and theory that will be used in the later chapters. First, in Section 2.1, we fix some basic notions about arithmetical theories and provability predicates. Then, in Section 2.2, we discuss the T-translation, which will lead to theories that prove their own completeness. In Section 2.3, we turn our attention to two nonstandard notions of provability, called fast and slow provability. Finally, in Section 2.4, we develop some intuitionistic modal (propositional) logic.

2.1 Arithmetic and Provability

All the theories we shall consider will be theories for intuitionistic predicate logic with equality. As our proof system, we pick natural deduction with equality. An *axiom* will be viewed as a special case of an inference rule, namely as an inference rule whose premiss set is empty. For equality, we have the axiom $x = x$, and an inference rule for substitution. The language in which our theories will be fomulated will be the *language of arithmetic* $\mathcal{L} = \{0, \mathsf{S}, +, \times, \mathsf{exp}\}$. Here 0 is a constant symbol, S is a unary function symbol and $+$, $\times$ and exp are binary function symbols. For each $n \in \mathbb{N}$, we can define the $\mathcal{L}$-term $\mathsf{S} \ldots \mathsf{S}0$, where the S occurs exactly n times. This term is called the *numeral* of n, and we denote it by $\overline{n}$. Usually, however, we omit the bar and just write n for the numeral of n. For terms s and t, we define $s \leq t$ as $\exists x\,(s + x = t)$ and $s < t$ as $\exists x\,(s + x + 1 = t)$. Here x should not occur in s or t, of course. We notice that the language $\mathcal{L}$ has a straightforward interpretation in the natural numbers, yielding the *standard model* $\mathbb{N}$. We introduce two special classes of formulae.

Definition 2.1.1. (i) The set of Δ_0-*formulae* is defined by recursion, as follows:

(a) all atomic $\mathcal{L}$-formulae are Δ_0-formulae;

(b) the set of Δ_0-formulae is closed under conjunction, disjunction and implication;

(c) if A is a Δ_0-formula, and t is an $\mathcal{L}$-term not containing the variable x, then the formulae $\exists x\,(x < t \wedge A)$ and $\forall x\,(x < t \to A)$ are also Δ_0-formulae.

We write $A \in \Delta_0$ if A is a Δ_0-formula.

(ii) The set of Σ_1-*formulae* consists of all $\mathcal{L}$-formulae of the form $\exists x\, A$, where $A \in \Delta_0$. We write $S \in \Sigma_1$ if S is a Σ_1-formula.

To each $\mathcal{L}$-expression α (which can be a term, a formula or a sequence of formulae), we assign a *Gödel number* $\ulcorner\alpha\urcorner$ in some reasonable way. More precisely, we require that elementary syntactic operations concerning $\mathcal{L}$ are primitive recursive in their Gödel numbers.

Definition 2.1.2. A *theory* T will be a pair $(\mathrm{Th}(T), \mathsf{Ax}_T)$, where Ax_T is a Σ_1-formula in one free variable, and $\mathrm{Th}(T)$ is precisely the set of $\mathcal{L}$-formulae derivable from the axiom set

$$\{A \mid A \text{ an } \mathcal{L}\text{-formula}, \mathbb{N} \models \mathsf{Ax}_T(\ulcorner A\urcorner)\}.$$

In other words, a theory is a set of $\mathcal{L}$-formulae that is closed under derivability in intuitionistic predicate logic with equality, together with a Σ_1-formula that defines an axiom set for the theory in the standard model. Usually, we will define a theory by giving its axioms, understanding that their is some natural Σ_1-formulation in $\mathcal{L}$ for axiomhood. For a set of $\mathcal{L}$-formulae Γ and an $\mathcal{L}$-formula A, we write $\Gamma \vdash_T A$ to indicate that A is provable using open assumptions from Γ and the axioms of T. Notice that $\vdash_T A$ just means $A \in \mathrm{Th}(T)$. Now we define three theories that will be of great interest to us.

Definition 2.1.3. (i) The theory HA, called *Heyting arithmetic*, has the axioms

$\neg(\mathsf{S}x = 0)$	$x \times 0 = 0$
$\mathsf{S}x = \mathsf{S}y \to x = y$	$x \times \mathsf{S}y = x \times y + x$
$x + 0 = 0$	$\mathsf{exp}(x, 0) = 1$
$x + \mathsf{S}y = \mathsf{S}(x + y)$	$\mathsf{exp}(x, \mathsf{S}y) = \mathsf{exp}(x, y) \times x$

and, for each $\mathcal{L}$-formula A, the *induction axiom* $A[0/x] \wedge \forall x\,(A \to A[\mathsf{S}x/x]) \to \forall x\, A$.

(ii) The theory EA, called *elementary arithmetic*, has the same axioms as HA, except that the induction scheme is restricted to formulae $A \in \Delta_0$.

(iii) The theory PA, called *Peano arithmetic*, has all the axioms of HA, together with the *Law of the Excluded Middle*: $A \vee \neg A$, where A is an $\mathcal{L}$-formula.

Even though the axiom set we presented for EA is infinite, the theory EA is actually finitely axiomatizable (see e.g. [3], Theorem V.5.6), and there in fact exists a choice for $\mathsf{Ax}_{\mathsf{EA}}$ such the finite axiomatizability of EA can be verified in EA itself. It is also well-known that EA, and hence any theory extending it, is Σ_1-*complete*. That is, every Σ_1-sentence true in the standard model can be proven inside EA.

Concerning HA, we have the following well-known result.

Proposition 2.1.1. *Let $F\colon \mathbb{N}^k \to \mathbb{N}$ be primitive recursive. Then there exists a Σ_1-formula $A_F(\vec{x}, y)$ satisfying:*

(i) $\vdash_{\mathsf{HA}} A_F(\vec{n}, F(\vec{n}))$ *for all* $\vec{n} \in \mathbb{N}^k$*;*

(ii) $\vdash_{\mathsf{HA}} \exists y \forall z\,(A_F(\vec{x}, z) \leftrightarrow y = z)$.

Moreover, this formula can be chosen in such a way that the definition of F as a primitive recursive function is verifiable in HA.

Notice that we have a primitive recursive function $\mathrm{Subst}\colon \mathbb{N}^2 \to \mathbb{N}$ that is defined as follows. If a is the Gödel number of some formula $A(v)$ in one free variable v, then $\mathrm{Subst}(a, b) = \ulcorner A(b)\urcorner$;

otherwise, $\mathrm{Subst}(a,b)=0$. We can represent this function in HA using Proposition 2.1.1. In fact, for this particular function the clauses (i) and (ii) even hold when HA is replaced by EA. If $A(v)$ is a formula with one free variable, we will write $\ulcorner A(\tilde{x})\urcorner$ for $\mathrm{Subst}(\ulcorner A(v)\urcorner, x)$, which makes sense when working in a theory extending EA. We apply similar conventions for multiple free variables. We will need the following famous result, that we will not prove.

Theorem 2.1.2 (Diagonalization Lemma). *Suppose $A(\vec{x}, y)$ is an $\mathcal{L}$-formula. Then there exists an $\mathcal{L}$-formula $B(\vec{x})$ such that $\vdash_{\mathsf{EA}} B(\vec{x}) \leftrightarrow A(\vec{x}, \ulcorner B(\vec{x})\urcorner)$.*

Concerning the Δ_0- and Σ_1-formulae, we have the following results.

Proposition 2.1.3. *(i) If $A \in \Delta_0$, then $\vdash_{\mathsf{HA}} A \vee \neg A$.*

(ii) If $S \in \Sigma_1$, then we have $\vdash_{\mathsf{HA}} S \leftrightarrow \exists x\,(s=t)$ for a certain variable x and certain $\mathcal{L}$-terms s and t.

Now suppose we have a theory T. Using the Σ_1-formula Ax_T, we can construct a Σ_1-formula $\mathsf{Bew}_T(x)$ that expresses 'x is the Gödel number of some formula A such that $\vdash_T A$' in a natural way. We can write $\mathsf{Bew}(x)$ as $\exists y\,\mathsf{Prf}_T(y,x)$ for some Δ_0-formula Prf_T. We think of $\mathsf{Prf}(y,x)$ as expressing the fact that y codes a T-proof of the formula that has x as its Gödel number. For a formula $A = A(x_1,\dots,x_n)$, we write $\Box_T A$ for $\mathsf{Bew}_T(\ulcorner A(\tilde{x}_1,\dots,\tilde{x}_n)\urcorner)$. In particular, $\Box_T A$ has the same free variables as A. Now we can define certain relations between theories.

Definition 2.1.4. Let U and T be theories. We write:

(i) $U \subseteq T$ if $\mathrm{Th}(U) \subseteq \mathrm{Th}(T)$;

(ii) $U = T$ if $\mathrm{Th}(U) = \mathrm{Th}(T)$;

(iii) $U \leq T$ if $\vdash_{\mathsf{HA}} \mathsf{Bew}_U(x) \to \mathsf{Bew}_T(x)$;

(iv) $U \equiv T$ if $\vdash_{\mathsf{HA}} \mathsf{Bew}_U(x) \leftrightarrow \mathsf{Bew}_T(x)$.

We emphasize that, then we write $U = T$, we do not mean an equality of the pairs $(\mathrm{Th}(U), \mathsf{Ax}_U)$ and $(\mathrm{Th}(T), \mathsf{Ax}_T)$, but only an equality of the first coordinate. Since HA is sound, we see that $U \leq T$ implies that $U \subseteq T$. We also notice that, if U and T are theories such that $\vdash_{\mathsf{HA}} \mathsf{Ax}_U(x) \to \mathsf{Ax}_T(x)$, then $U \leq T$ clearly holds. However, this requirement is not necessary: it can also be the case that every U-proof can (verifiably in HA) be transformed into a T-proof without the one axiom set being contained in the other. Before we can develop more theory, we need to restrict our investigation to theories that, verifiably in HA, can perform a minimal amount of arithmetic.

Convention 2.1.1. All the theories T we shall consider, will satisfy $\mathsf{EA} \leq T$.

Notice that this clearly holds for the three theories from Definition 2.1.3. With this requirement in place, we can state some basic properties of $\Box_T$, that we will not prove.

Proposition 2.1.4. *Let T be a theory and let A, B and S be $\mathcal{L}$-formulae. Then we have:*

(i) We have $\vdash_T A$ if and only if $\mathbb{N} \models \Box_T A$, if and only if $\vdash_{\mathsf{EA}} \Box_T A$;

(ii) $\vdash_{\mathsf{HA}} \Box_T(A \to B) \to (\Box_T A \to \Box_T B)$;

(iii) $\vdash_{\mathsf{HA}} \Box_T A \to \Box_T \Box_T A$;

(iv) (Formalized Σ_1-completeness) if $S \in \Sigma_1$, then $\vdash_{\mathsf{HA}} S \to \Box_T S$;

(*v*) (Löb's Principle) *if U is a theory such that* $\mathsf{HA} \subseteq U \subseteq T$*, and* $\vdash_U \Box_T A \to A$*, then* $\vdash_U A$.

(*vi*) (Löb's Theorem) $\vdash_{\mathsf{HA}} \Box_T(\Box_T A \to A) \to \Box_T A$.

Moreover, (ii), (iii), (iv) and (vi) are verifiable in HA.

We remark that for (iii)-(vi), we need Convention 2.1.1. In the next section, we will need the following facts.

Proposition 2.1.5. *Let U and T be theories.*

(*i*) *If $U \subseteq T$, then $\vdash_U A$ implies $\vdash_U \Box_T A$ for all $\mathcal{L}$-formulae A.*

(*ii*) *If* $\mathsf{HA} \subseteq U \leq T$*, then* $\vdash_{\mathsf{HA}} \mathsf{Bew}_U(x) \to \Box_U \mathsf{Bew}_T(x)$*. In particular,* $\vdash_{\mathsf{HA}} \Box_U A \to \Box_U \Box_T A$ *for all $\mathcal{L}$-formulae A.*

Proof. (i) If $\vdash_U A$, then also $\vdash_T A$, so $\vdash_{\mathsf{EA}} \Box_T A$. Since $\mathsf{EA} \subseteq U$, we also get $\vdash_U \Box_T A$.
(ii) Since $U \leq T$, we have $\vdash_{\mathsf{HA}} \mathsf{Bew}_U(x) \to \mathsf{Bew}_T(x)$. Since $\mathsf{HA} \subseteq U$, it follows from (i) that $\vdash_{\mathsf{HA}} \Box_U \mathsf{Bew}_U(x) \to \Box_U \mathsf{Bew}_T(x)$. We also have $\vdash_{\mathsf{HA}} \mathsf{Bew}_U(x) \to \Box_U \mathsf{Bew}_U(x)$ by formalized Σ_1-completeness, and now the result follows. □

For future use, we state the following definition.

Definition 2.1.5. Let $T \supseteq \mathsf{HA}$ be a theory and let $S(x)$ be a Σ_1-formula in one free variable. For an $\mathcal{L}$-sentence A, we write $\Box A$ for $S(\ulcorner A \urcorner)$. We say that S is a *provability predicate* for T if the following hold for all $\mathcal{L}$-sentences A, B and R:

(i) if $\vdash_T A$, then $\mathbb{N} \models \Box A$;

(ii) $\vdash_{\mathsf{HA}} \Box(A \to B) \to (\Box A \to \Box B)$;

(iii) if $R \in \Sigma_1$, then $\vdash_{\mathsf{HA}} R \to \Box R$.

Using Proposition 2.1.4, we see that Bew_T is always a provability predicate for T.

Remark 2.1.1. Our use of the term 'provability predicate' differs from its use in other literature. Usually, a provability predicate is defined as a formula that satisfies the Hilbert-Bernays-Löb derivability conditions. Our definition is stronger, since these conditions follow from (i)-(iii) above. Indeed let A be an $\mathcal{L}$-sentence. Since HA is Σ_1-complete and $\Box A$ is a Σ_1-sentence, we have that $\mathbb{N} \models \Box A$ implies $\vdash_{\mathsf{HA}} \Box A$. From (iii), it follows that $\vdash_{\mathsf{HA}} \Box A \to \Box\Box A$, again since $\Box A$ is a Σ_1-sentence. In particular, we can derive Löb's Principle for $\Box$, using the Diagonalization Lemma. That is, if U is a theory such that $\mathsf{HA} \subseteq U \subseteq T$, and A is an $\mathcal{L}$-sentence such that $\vdash_U \Box A \to A$, then $\vdash_U A$. We also have Löb's Theorem for $\Box$, that is, $\vdash_{\mathsf{HA}} \Box(\Box A \to A) \to \Box A$ for all $\mathcal{L}$-sentences A. ◊

2.2 The Completeness Principle

In this section, we introduce the T-translation, that will allow us to define theories that prove their own completeness. All results in this section are from Visser's paper *On the Completeness Principle* [10], but we have formulated some of them under weaker conditions.

Definition 2.2.1. Let T be a theory. We define the T-translation $(\cdot)^T$ from the set of $\mathcal{L}$-formulae to itself by recursion. For all $\mathcal{L}$-terms s and t and $\mathcal{L}$-formulae A and B, we set:

(i) $(s = t)^T$ is $s = t$ and $\perp^T$ is $\perp$;
(ii) $(A \circ B)^T$ is $A^T \circ B^T$ for $\circ \in \{\wedge, \vee\}$;
(iii) $(A \to B)^T$ is $(A^T \to B^T) \wedge \Box_T(A^T \to B^T)$;
(iv) $(\exists x\, A)^T$ is $\exists x\, A^T$;
(v) $(\forall x\, A)^T$ is $\forall x\, A^T \wedge \Box_T(\forall x\, A^T)$.

Based on the T-translation, we can construct new theories out of old ones.

Definition 2.2.2. Let U and T be theories. We define the theory U^T as $\mathsf{HA} + \{A \mid \vdash_U A^T\}$. For a theory U, we write U^* for U^U.

We make some remarks on how Ax_{U^T} can be defined. Clearly, the function $(\cdot)^T : \mathbb{N} \to \mathbb{N}$ that satisfies $x^T = \ulcorner A^T \urcorner$ is x is the Gödel number of an $\mathcal{L}$-formula A, and $x^T = 0$ otherwise, is primitive recursive. So we can represent this function in HA using Proposition 2.1.1. Now we define $\mathsf{Ax}_{U^T}(x)$ as $\mathsf{Ax}_{\mathsf{HA}}(x) \vee (\mathrm{Form}(x) \wedge \mathsf{Bew}_U(x^T))$, where $\mathrm{Form}(x)$ naturally expresses the fact that x is the Gödel number of an $\mathcal{L}$-formula. We study the relation between provability in U^T and provability in U through the following lemmata.

Lemma 2.2.1. *For all $\mathcal{L}$-formulae A, we have $\vdash_{\mathsf{HA}} A^T \to \Box_T A^T$, and this fact is verifiable in HA.*

Proof. We proceed by induction on the complexity of A.

At If A is atomic, then $A^T = A$ and the claim follows from Proposition 2.1.4(iv) since A is a Σ_1-formula.

$\wedge$ Suppose $A = B \wedge C$ and the claim holds for B and C. Then A^T is $B^T \wedge C^T$, and we have
$$\vdash_{\mathsf{HA}} B^T \wedge C^T \to \Box_T B^T \wedge \Box_T C^T \to \Box_T(B^T \wedge C^T),$$
as desired.

$\vee$ Suppose $A = B \vee C$ and the claim holds for B and C. Then A^T is $B^T \vee C^T$, and we have $\vdash_{\mathsf{HA}} B^T \to \Box_T B^T \to \Box_T(B^T \vee C^T)$ and $\vdash_{\mathsf{HA}} C^T \to \Box_T C^T \to \Box_T(B^T \vee C^T)$, which together yield $\vdash_{\mathsf{HA}} B^T \vee C^T \to \Box_T(B^T \vee C^T)$, as desired.

$\to$ Suppose $A = B \to C$ and the claim holds for B and C. Then the formula A^T is equal to $(B^T \to C^T) \wedge \Box_T(B^T \to C^T)$, and we have
$$\begin{aligned}\vdash_{\mathsf{HA}} (B^T \to C^T) \wedge \Box_T(B^T \to C^T) &\to \Box_T(B^T \to C^T)\\ &\to \Box_T(B^T \to C^T) \wedge \Box_T\Box_T(B^T \to C^T)\\ &\to \Box_T((B^T \to C^T) \wedge \Box_T(B^T \to C^T)),\end{aligned}$$
as desired.

$\exists$ Suppose $A = \exists x\, B$ and the claim holds for B. Then A^T is $\exists x\, B^T$. It is provable in intuitionistic predicate logic that $B^T \to \exists x\, B^T$, so we also have $\vdash_{\mathsf{HA}} \Box_T B^T \to \Box_T(\exists x\, B^T)$. We get $\vdash_{\mathsf{HA}} \exists x\, B^T \to \exists x \Box_T B^T \to \Box_T(\exists x\, B^T)$, as desired.

∀ Suppose $A = \forall x B$ and the claim holds for B. Then $A^T = \forall x B^T \wedge \Box_T(\forall x B^T)$, and we have

$$\begin{aligned}\vdash_{\mathsf{HA}} \forall x B^T \wedge \Box_T(\forall x B^T) &\to \Box_T(\forall x B^T)\\ &\to \Box_T(\forall x B^T) \wedge \Box_T\Box_T(\forall x B^T)\\ &\to \Box_T(\forall x B^T \wedge \Box_T(\forall x B^T)),\end{aligned}$$

as desired.

For the second statement, we should carry out this induction inside HA. One should notice that now we need that clauses (ii)-(iv) from Proposition 2.1.4 are verifiable in HA. □

Lemma 2.2.2. *Let A be a formula, let x be a variable, and let s be a term.*

(i) A and A^T have the same free variables;

(ii) s is free for x in A if and only if s is free for x in A^T;

(iii) if s is free for x in A, then $\vdash_{\mathsf{HA}} (A^T)[s/x] \leftrightarrow (A[s/x])^T$.

Moreover, these are all verifiable in HA.

Proof. All three statements can be proven by an easy induction on the complexity of A. For the induction steps for implication and universal quantification in statement (iii), one should observe that, verifiably in HA, we have $\vdash_{\mathsf{HA}} (\Box_T A)[s/x] \leftrightarrow \Box_T(A[s/x])$ for all $\mathcal{L}$-terms s and $\mathcal{L}$-formulae A. □

Using these two lemmata, we can prove the following crucial result.

Theorem 2.2.3. *Let U and T be a theories such that $\mathsf{HA} \subseteq U$ and: $\vdash_U B$ implies $\vdash_U \Box_T B$ for all $\mathcal{L}$-formulae B. For a set of $\mathcal{L}$-formulae Γ, write $\Gamma^T = \{B^T \mid B \in \Gamma\}$. Then for all $\mathcal{L}$-formulae A, we have $\Gamma \vdash_{U^T} A$ if and only if $\Gamma^T \vdash_U A^T$.*

Remark 2.2.1. (i) By Proposition 2.1.5(i), the conditions on U and T apply in particular when $\mathsf{HA} \subseteq U \subseteq T$. We formulate this theorem (and Corollary 2.2.6 below) in such a strong way in order to obtain Lemma 4.3.1 towards the end of this thesis.

(ii) We warn the reader that, under these conditions, we cannot necessarily verify the result '$\Gamma \vdash_{U^T} A$ if and only if $\Gamma^T \vdash_U A^T$' inside HA; see Corollary 2.2.6 below. ♢

Proof of Theorem 2.2.3. Suppose that $\Gamma^T \vdash_U A^T$. Then there exist $n \geq 0$ and $C_1, \cdots, C_n \in \Gamma$ such that $\vdash_U C_0^T \wedge \ldots \wedge C_n^T \to A^T$. Then we also have $\vdash_U (C_0 \wedge \ldots \wedge C_n)^T \to A^T$, and by our assumption, we also get $\vdash_U \Box_T((C_0 \wedge \ldots \wedge C_n)^T \to A^T)$. So $\vdash_U (C_0 \wedge \ldots \wedge C_n \to A)^T$, and therefore we get $\vdash_{U^T} C_0 \wedge \ldots \wedge C_n \to A$. Finally, this clearly yields that $\Gamma \vdash_{U^T} A$.

For the converse direction, we proceed by induction on the proof tree for $\Gamma \vdash_{U^T} A$. Before we start, we notice the following: if $\vdash_U B \to C$ for certain $\mathcal{L}$-formulae B and C, then by our assumption, $\vdash_U \Box_T(B \to C)$. Since $\mathsf{HA} \subseteq U$, we also have $\vdash_U \Box_T(B \to C) \to (\Box_T B \to \Box_T C)$, so in particular, we get $\vdash_U \Box_T B \to \Box_T C$. We also note: if $\vdash_{\mathsf{HA}} B$, then $\vdash_U B$, whence $\vdash_U \Box_T B$.

First, suppose that A is an axiom of U^T. That is, we suppose that A is an axiom of HA or $\vdash_U A^T$. In the latter case, we are done. So suppose that A is an axiom of HA. We need to show that $\vdash_U A^T$. If A is the axiom $x = x$, then A^T is also $x = x$, which is clearly provable in

U. If A is the axiom $\neg(\mathsf{S}x = 0)$, then A^T is $\neg(\mathsf{S}x = 0) \wedge \Box_T(\neg(\mathsf{S}x = 0))$. Since $\vdash_{\mathsf{HA}} \neg(\mathsf{S}x = 0)$, we also have $\vdash_U \neg(\mathsf{S}x = 0)$, and by our assumption, $\vdash_U \Box_T(\neg(\mathsf{S}x = 0))$. So we indeed have $\vdash_U A^T$. If A is another basic axiom of HA, then A is atomic, so A^T is equal to A itself again. So $\vdash_{\mathsf{HA}} A^T$, hence also $\vdash_U A^T$. It remains to prove the claim for the case where A is an induction axiom, say $B[0/x] \wedge \forall x (B \to B[\mathsf{S}x/x]) \to \forall x B$. First of all, we notice that

$$\begin{aligned}\vdash_U (\forall x (B \to B[\mathsf{S}x/x]))^T &\leftrightarrow \forall x (B \to B[\mathsf{S}x/x])^T \wedge \Box_T(\forall x (B \to B[\mathsf{S}x/x])^T) \\ &\to \forall x (B^T \to (B[\mathsf{S}x/x])^T) \wedge \Box_T(\forall x (B^T \to (B[\mathsf{S}x/x])^T)) \\ &\leftrightarrow \forall x (B^T \to (B^T)[\mathsf{S}x/x]) \wedge \Box_T(\forall x (B^T \to (B^T)[\mathsf{S}x/x])). \quad (2.1)\end{aligned}$$

Furthermore, we know that $\vdash_{\mathsf{HA}} (B[0/x])^T \leftrightarrow (B^T)[0/x]$ and that $(\forall x B)^T$ is the formula $\forall x B^T \wedge \Box_T(\forall x B^T)$. Define the formulae

$$\begin{aligned}C &= (B^T)[0/x] \wedge \forall x (B^T \to (B^T)[\mathsf{S}x/x]) \wedge \Box_T(\forall x (B^T \to (B^T)[\mathsf{S}x/x])) \to \forall x B^T \wedge \Box_T(\forall x B^T), \\ D &= (B[0/x])^T \wedge (\forall x (B \to B[\mathsf{S}x/x]))^T \to (\forall x B)^T.\end{aligned}$$

Then it follows from (2.1) that $\vdash_U C \to D$, so we also get $\vdash_U \Box_T C \to \Box_T D$. Since A^T is the formula $D \wedge \Box_T D$, we see that $\vdash_U C \wedge \Box_T C \to A^T$. So it suffices to show that $\vdash_U C$.
We have

$$\vdash_U (B^T)[0/x] \wedge \forall x (B^T \to (B^T)[\mathsf{S}x/x]) \to \forall x B^T, \quad (2.2)$$

since the displayed formula is an induction axiom and $\mathsf{HA} \subseteq U$. Now it follows that

$$\vdash_U \Box_T((B^T)[0/x]) \wedge \Box_T(\forall x (B^T \to (B^T)[\mathsf{S}x/x])) \to \Box_T(\forall x B^T). \quad (2.3)$$

Finally, since $\vdash_{\mathsf{HA}} (B^T)[0/x] \leftrightarrow (B[0/x])^T$, we can use Lemma 2.2.1 to see that

$$\vdash_U (B^T)[0/x] \to (B[0/x])^T \to \Box_T((B[0/x])^T) \to \Box_T((B^T)[0/x]). \quad (2.4)$$

From (2.2), (2.3) and (2.4), we may deduce that C is indeed provable in U, as desired.
Now we treat the rules of inference. Since the T-translation commutes with conjunction, disjunction and existential quantification, the induction steps for rules of inference for these operators are trivial. It remains to check the rules for implication and universal quantification, and the substitution rule.

- $\to$E Suppose that $\Gamma^T \vdash_U (B \to C)^T$ and $\Gamma^T \vdash_U B^T$. We need to show that $\Gamma^T \vdash_U C^T$. But this is obvious since $\vdash_U (B \to C)^T \to (B^T \to C^T)$.
- $\to$I Suppose that $\Gamma^T, B^T \vdash_U C^T$. We need to show that $\Gamma^T \vdash_U (B \to C)^T$. We certainly have $\Gamma^T \vdash_U B^T \to C^T$. But then we also have $\Box_T(\Gamma^T) \vdash_U \Box_T(B^T \to C^T)$, where $\Box_T(\Gamma^T) = \{\Box_T(D^T) \mid D \in \Gamma\}$. Since $\mathsf{HA} \subseteq U$ and $\vdash_{\mathsf{HA}} D^T \to \Box_T D^T$ for all $D \in \Gamma$, we get $\Gamma^T \vdash_U \Box_T(B^T \to C^T)$. Combining our results, we find

  $$\Gamma^T \vdash_U (B^T \to C^T) \wedge \Box_T(B^T \to C^T),$$

  as desired.
- $\forall$E Suppose that $\Gamma^T \vdash_U (\forall x B)^T$. We need to show that $\Gamma^T \vdash_U (B[s/x])^T$. Since $\vdash_U (\forall x B)^T \to \forall x B^T$, we see that $\Gamma^T \vdash_U (B^T)[s/x]$. Since we also know that $\mathsf{HA} \subseteq U$ and $\vdash_{\mathsf{HA}} (B^T)[s/x] \leftrightarrow (B[s/x])^T$, we get $\Gamma^T \vdash_U (B[s/x])^T$, as desired.

∀I Suppose that $\Gamma^T \vdash_U B^T$, where the variable x does not occur anywhere in Γ. We need to show that $\Gamma^T \vdash_U (\forall x B)^T$. First of all, we certainly have $\Gamma^T \vdash_U \forall x B^T$, since x does not occur free anywhere in Γ^T. By applying the same reasoning as in the →I-case, we find $\Gamma^T \vdash_U \Box_T(\forall x B^T)$. We conclude that $\Gamma^T \vdash_U \forall x B^T \wedge \Box_T(\forall x B^T)$, as desired.

Subst Suppose that $\Gamma^T \vdash_U (B[s/x])^T$ and $\Gamma^T \vdash_U (s = t)^T$. We need to show that $\Gamma^T \vdash_U (B[t/x])^T$. We have $\Gamma^T \vdash_U s = t$ and by Lemma 2.2.2(iii) and the fact that $\mathsf{HA} \subseteq U$, we get $\Gamma^T \vdash_U (B^T)[s/x]$. This yields $\Gamma^T \vdash_U (B^T)[t/x]$, and thus $\Gamma^T \vdash_U (B[t/x])^T$, as desired.

This completes the induction. □

From this theorem, we can deduce the following results.

Corollary 2.2.4. *If* $\mathsf{HA} \subseteq U \subseteq T$, *then the theory* U^T *is consistent if and only if* U *is consistent.*

Corollary 2.2.5. *Suppose that* $\mathsf{HA} \subseteq T$. *If* $A \vdash_{\mathsf{HA}} B$, *then* $A^T \vdash_{\mathsf{HA}} B^T$.

Proof. If $A \vdash_{\mathsf{HA}} B$, then also $A \vdash_{\mathsf{HA}^T} B$. By applying Theorem 2.2.3 with $U \equiv \mathsf{HA}$, we find that $A^T \vdash_{\mathsf{HA}} B^T$. □

Corollary 2.2.6. *Let* U *and* T *be theories such that* $\mathsf{HA} \leq U$ *and* $\vdash_{\mathsf{HA}} \mathsf{Bew}_U(x) \to \Box_U \mathsf{Bew}_T(x)$. *Then* $\vdash_{\mathsf{HA}} \Box_{U^T} A \leftrightarrow \Box_U A^T$ *for all* $\mathcal{L}$*-formulae* A.

Remark 2.2.2. By Proposition 2.1.5(ii), the requirements on U and T are satisfied when $\mathsf{HA} \leq U \leq T$. ◊

Proof of Corollary 2.2.6. The '←'-direction is immediate as it follows from the definition of U^T, and it does not need the requirements on U and T. Concretely, we have

$$\vdash_{\mathsf{HA}} \mathrm{Form}(x) \wedge \mathsf{Bew}_U(x^T) \to \mathsf{Ax}_{U^T}(x) \to \mathsf{Bew}_{U^T}(x).$$

From this, the desired result follows.
For the '→'-direction, we formalize the proof of the left-to-right direction of Theorem 2.2.3 inside HA. We need that the statements of Proposition 2.1.4, Lemma 2.2.1 and Lemma 2.2.2 are verifiable in HA. If we restrict the result to the case where Γ is empty, we get $\vdash_{\mathsf{HA}} \mathsf{Bew}_{U^T}(x) \to \mathsf{Bew}_U(x^T)$, from which the desired result will follow. □

Next, we prove an important conservation result for U^T. First, we need the following definition.

Definition 2.2.3. The set $\mathcal{A}$ is the smallest set of $\mathcal{L}$-formulae such that

(i) $\mathcal{A}$ contains all atomic $\mathcal{L}$-formulae;

(ii) $\mathcal{A}$ is closed under conjunction, disjunction, and both existential and universal quantification;

(iii) if $S \in \Sigma_1$ and $A \in \mathcal{A}$, then $S \to A \in \mathcal{A}$.

Lemma 2.2.7. *Let* T *be a theory such that* $\mathsf{HA} \subseteq T$.

(i) We have $\vdash_{\mathsf{HA}} S \leftrightarrow S^T$ *for all* $S \in \Sigma_1$.

(ii) We have $\vdash_{\mathsf{HA}} A^T \to A$ *for all* $A \in \mathcal{A}$.

Now suppose that we have another theory U such that $\mathsf{HA} \subseteq U \subseteq T$.

(*iii*) *The theories U and U^T prove the same Σ_1-formulae.*

(*iv*) *The theory U^T is $\mathcal{A}$-conservative over U.*

Proof. (i) Let $S \in \Sigma_1$. By Proposition 2.1.3(ii), we have that $\vdash_{\mathsf{HA}} S \leftrightarrow \exists x\,(s = t)$ for certain $\mathcal{L}$-terms s and t. By Corollary 2.2.5, we also have $\vdash_{\mathsf{HA}} S^T \leftrightarrow (\exists x\,(s=t))^T$. Since $(\exists x\,(s=t))^T$ is just $\exists x\,(s=t)$, the result follows.
(ii) We proceed by induction on the complexity of A. Only clause (iii) in the definition of $\mathcal{A}$ is nontrivial. Suppose that A is $S \to B$, where $S \in \Sigma_1$, and that we already know the result for B. Then $\vdash_{\mathsf{HA}} S \leftrightarrow S^T$ and $\vdash_{\mathsf{HA}} B^T \to B$, so

$$\vdash_{\mathsf{HA}} (S \to B)^T \to (S^T \to B^T) \to (S \to B),$$

as desired.
(iii) Let $S \in \Sigma_1$. Then by item (i) and Theorem 2.2.3, we have that $\vdash_{U^T} S$ if and only if $\vdash_U S^T$, if and only if $\vdash_U S$.
(iv) Suppose we have $A \in \mathcal{A}$ such that $\vdash_{U^T} A$. By Theorem 2.2.3, we have $\vdash_U A^T$, and by item (ii), we get $\vdash_U A$. ☐

At the beginning of this section, we promised to construct theories that prove their own completeness. We now make this precise.

Definition 2.2.4. Let $S(x)$ be a provability predicate for a theory T. Again, if A is an $\mathcal{L}$-sentence, we write $\Box A$ for $S(\ulcorner A \urcorner)$.

(i) The *completeness principle* $\mathrm{CP}_\Box$ is the axiom scheme $A \to \Box A$, where A is an $\mathcal{L}$-sentence.

(ii) The *strong Löb principle* $\mathrm{SLP}_\Box$ is the axiom scheme $(\Box A \to A) \to A$, where A is an $\mathcal{L}$-sentence.

We write CP_T for $\mathrm{CP}_{\Box_T}$, and similarly for SLP.

Notice that '$\mathrm{CP}_\Box$' is a slight abuse of notation, since the $\Box$ is just an abbreviation and $\mathrm{CP}_\Box$ actually depends on $S(x)$.

Lemma 2.2.8. *Let $S(x)$ be a provability predicate for a theory T. Then the completeness principle and the strong Löb principle for S are interderivable over* HA.

Proof. Define $\Box$ as above, and let A be an $\mathcal{L}$-sentence. First, we show that $\vdash_{\mathsf{HA}+\mathrm{CP}_\Box} \mathrm{SLP}_\Box$. Since $S(x)$ is a provability predicate, we have the formalized Löb Theorem for $\Box$, so

$$\vdash_{\mathsf{HA}+\mathrm{CP}_\Box} (\Box A \to A) \to \Box(\Box A \to A) \to \Box A,$$

from which $\vdash_{\mathsf{HA}+\mathrm{CP}_\Box} (\Box A \to A) \to A$ follows.
Now we show that $\vdash_{\mathsf{HA}+\mathrm{SLP}_\Box} \mathrm{CP}_\Box$. Clearly, we have $\vdash_{\mathsf{HA}} \Box(A \wedge \Box A) \to \Box A$, so

$$\begin{aligned}\vdash_{\mathsf{HA}+\mathrm{SLP}_\Box} A &\to (\Box(A \wedge \Box A) \to A \wedge \Box A)\\ &\to A \wedge \Box A\\ &\to \Box A,\end{aligned}$$

as desired. ☐

Lemma 2.2.9. *Let U and T be theories such that $\mathsf{HA} \subseteq U, T$. Then for all $\mathcal{L}$-sentences A, we have $\vdash_{U^T} A \to \Box_T A^T$. In particular, $\vdash_{U^T} \mathrm{CP}_{T^*}$.*

Proof. Let A be an $\mathcal{L}$-sentence. Since $\mathsf{HA} \subseteq T$ and $\Box_T A^T \in \Sigma_1$, we have that $\vdash_{\mathsf{HA}} \Box_T A^T \to (\Box_T A^T)^T$. So we get $\vdash_{\mathsf{HA}} A^T \to \Box_T A^T \to (\Box_T A^T)^T$, and since $\mathsf{HA} \subseteq T$, we also get $\vdash_{\mathsf{HA}} \Box_T(A^T \to (\Box_T A^T)^T)$. So we find $\vdash_{\mathsf{HA}} (A \to \Box_T A^T)^T$. This means that $\vdash_U (A \to \Box_T A^T)^T$, since $\mathsf{HA} \subseteq U$. It follows that $\vdash_{U^T} A \to \Box_T A^T \to \Box_{T^*} A$, as desired. □

Remark 2.2.3. Notice that under the present assumptions, $\Box_T A^T$ is not necessarily *equivalent*, over HA, to $\Box_{T^*} A$. ◊

Corollary 2.2.10. *Let U be a theory such that $\mathsf{HA} \subseteq U$. Then $\vdash_{U^*} \mathrm{CP}_{U^*}$.*

Proof. This follows by taking $U \equiv T$ in Lemma 2.2.9. □

Remark 2.2.4. We remark that the proof of Lemma 2.2.9 also goes through if we replace the first line with 'Let A be an $\mathcal{L}$-formula.' We will not need this greater generality. ◊

2.3 Fast and Slow Provability

In this section, we introduce two nonstandard notions of provability. The first of these is *fast provability*, which can be seen as iterated provability. The second is *slow provability*, a notion of provability that puts a certain size restriction on the axioms that may be used in a proof. For developing the theory of fast provability, the following technique, that is also used in [4], will prove useful.

Lemma 2.3.1 (Reflexive induction)**.** *Let $T \supseteq \mathsf{HA}$ be a theory. Suppose A is an $\mathcal{L}$-formula in one free variable such that $\vdash_{\mathsf{HA}} A[0/x]$ and $\vdash_{\mathsf{HA}} \Box_T A \to A[\mathsf{S}x/x]$. Then $\vdash_{\mathsf{HA}} A$.*

Proof. It is provable in intuitionistic predicate logic that $\forall x\, A \to A$. So from our assumptions, it follows that $\vdash_{\mathsf{HA}} \Box_T(\forall x\, A) \to \Box_T A \to A[\mathsf{S}x/x]$. Since we also know that $\vdash_{\mathsf{HA}} A[0/x]$, we get $\vdash_{\mathsf{HA}} \Box_T(\forall x\, A) \to \forall x\, A$. Using Löb's Principle, we can conclude that $\vdash_{\mathsf{HA}} \forall x\, A$, so $\vdash_{\mathsf{HA}} A$. □

Definition 2.3.1. Let T be a theory.

(i) We define $\mathsf{IBew}_T(u, x)$ as a formula satisfying

$$\vdash_{\mathsf{EA}} \mathsf{IBew}_T(u, x) \leftrightarrow ((u = 0 \wedge \mathsf{Bew}_T(x)) \vee \exists v\, (u = \mathsf{S}v \wedge \Box_T \mathsf{IBew}_T(v, x)))$$

as provided by the Diagonalization Lemma.

(ii) For an $\mathcal{L}$-formula $A = A(x_1, \ldots, x_n)$, we write $\Box_T^{u+1} A$ for $\mathsf{IBew}_T(u, \ulcorner A(\tilde{x}_1, \ldots, \tilde{x}_n) \urcorner)$

(iii) We write $\mathsf{Bew}_T^f(x)$ for $\exists u \mathsf{IBew}_T(u, x)$. For an $\mathcal{L}$-formula A, we write $\Box_T^f A$ for $\exists u \Box_T^{u+1} A$.

Remark 2.3.1. As we shall see shortly, Bew_T^f is a provability predicate for T. This notion of provability is called *fast provability* and was introduced by Parikh in [8]. In this paper, fast provability is introduced in a different way, namely by closing the set of theorems of T under Parikh's rule 'from $\vdash \Box_T A$, infer $\vdash A$', where A is an $\mathcal{L}$-sentence. This yields, verifiably in HA, the same notion of provability we defined here. If T is Σ_1-sound, then Parikh's rule does not lead to any new theorems, so the notions of ordinary provability and fast provability coincide. However, the use of Parikh's rule can lead to much shorter proofs, which explains the name 'fast provability'. Later in this section, we will show that, if T is consistent, it is never verifiable in HA that fast provability coincides with ordinary provability. $\Diamond$

We notice that IBew_T is equivalent, over EA, to a Σ_1-formula. Informally, $\mathsf{IBew}_T(u,x)$ can be thought of as the formula $\mathsf{Bew}_T(\cdots(\mathsf{Bew}_T(x))\cdots)$, where the Bew_T occurs $u+1$ times, so we can see IBew_T as representing 'iterated provability'. Notice that we write '$u+1$' in the superscript of $\Box_T$, to indicate that the $\Box_T$ 'occurs' $u+1$ times. We prove a number of technical facts about $\Box_T^{u+1}$ and $\Box_T^f$.

Lemma 2.3.2. *Let $T \supseteq \mathsf{HA}$ be a theory and let A and B be $\mathcal{L}$-formulae. Then we have:*

(i) $\vdash_{\mathsf{HA}} \Box_T\Box_T^{u+1}A \leftrightarrow \Box_T^{\mathsf{S}u+1}A \leftrightarrow \Box_T^{u+1}\Box_T A$;

(ii) $\vdash_{\mathsf{HA}} u \leq v \to (\Box_T^{u+1}A \to \Box_T^{v+1}A)$;

(iii) $\vdash_{\mathsf{HA}} \Box_T^{u+1}(A \to B) \to (\Box_T^{u+1}A \to \Box_T^{u+1}B)$,

(iv) $\vdash_{\mathsf{HA}} \Box_T A \to \Box_T^f A$;

(v) Bew_T^f *is a provability predicate for T;*

(vi) if T is Σ_1-sound, then $\mathbb{N} \models \Box_T^f A$ if and only if $\mathbb{N} \models \Box_T A$, if and only if $\vdash_T A$;

(vii) $\vdash_{\mathsf{HA}} \Box_T^f\Box_T A \leftrightarrow \Box_T^f A$.

(viii) if T is consistent, then $\nvdash_{\mathsf{HA}} \Box_T^f\bot \to \Box_T\bot$.

Proof. (i) From the definition of IBew_T, it follows that $\vdash_{\mathsf{HA}} \Box_T\mathsf{IBew}_T(u,x) \leftrightarrow \mathsf{IBew}_T(\mathsf{S}u,x)$, so the first equivalence is immediate. For the second equivalence, we proceed by reflexive induction. First of all, we have $\vdash_{\mathsf{HA}} \Box_T^{\mathsf{S}0+1}A \leftrightarrow \Box_T\Box_T^{0+1}A \leftrightarrow \Box_T\Box_T A \leftrightarrow \Box_T^{0+1}\Box_T A$. Furthermore,

$$\begin{aligned} \vdash_{\mathsf{HA}} \Box_T(\Box_T^{\mathsf{S}u+1}A \leftrightarrow \Box_T^{u+1}\Box_T A) \to \\ [\Box_T^{\mathsf{SS}u+1}A &\leftrightarrow \Box_T\Box_T^{\mathsf{S}u+1}A \\ &\leftrightarrow \Box_T\Box_T^{u+1}\Box_T A \\ &\leftrightarrow \Box_T^{\mathsf{S}u+1}\Box_T A], \end{aligned}$$

which completes the proof.
(ii) Since $\Box_T^{u+1}A$ is a Σ_1-formula, we have $\vdash_{\mathsf{HA}} \Box_T^{u+1}A \to \Box_T\Box_T^{u+1}A \to \Box_T^{\mathsf{S}u+1}A$. Now the claim follows by induction on v inside HA.
(iii) We proceed by reflexive induction. First of all, we have

$$\vdash_{\mathsf{HA}} \Box_T^{0+1}(A \to B) \leftrightarrow \Box_T(A \to B) \to (\Box_T A \to \Box_T B) \leftrightarrow (\Box_T^{0+1}A \to \Box_T^{0+1}B).$$

Furthermore,

$$\begin{aligned} \vdash_{\mathsf{HA}} \Box_T(\Box_T^{u+1}(A \to B) \to (\Box_T^{u+1}A \to \Box_T^{u+1}B)) \to \\ [\Box_T^{\mathsf{S}u+1}(A \to B) &\leftrightarrow \Box_T\Box_T^{u+1}(A \to B) \\ &\to \Box_T(\Box_T^{u+1}A \to \Box_T^{u+1}B) \\ &\to (\Box_T\Box_T^{u+1}A \to \Box_T\Box_T^{u+1}B) \\ &\leftrightarrow (\Box_T^{\mathsf{S}u+1}A \to \Box_T^{\mathsf{S}u+1}B)], \end{aligned}$$

which completes the proof.
(iv) This is immediate as $\vdash_{\mathsf{HA}} \Box_T A \leftrightarrow \Box_T^{0+1} A$.
(v) This follows easily from (ii), (iii) and (iv).
(vi) The second statement was already asserted in Proposition 2.1.4(i). So we prove the first statement. The right-to-left direction follows from (iv). For the converse, suppose that $\mathbb{N} \models \Box^{n+1} A$ for a certain $n \in \mathbb{N}$. If $n = 0$, then we are done. So suppose that $n = m + 1$ for a certain $m \geq 0$. Then $\mathbb{N} \models \Box_T \Box_T^{m+1} A$, so $\vdash_T \Box_T^{m+1} A$. Since T is Σ_1-sound, we see that $\mathbb{N} \models \Box_T^{m+1} A$. By repeating this argument, we find $\mathbb{N} \models \Box_T A$, as desired.
(vii) This follows from $\vdash_{\mathsf{HA}} \Box_T^{u+1} \Box_T A \to \Box_T^{\mathsf{S}u+1} A$ and $\vdash_{\mathsf{HA}} \Box_T^{u+1} A \to \Box_T^{\mathsf{S}u+1} \to \Box_T^{u+1} \Box_T A$.
(viii) Suppose that $\vdash_{\mathsf{HA}} \Box_T^f \bot \to \Box_T \bot$. Then by (iv) and (vii), we have

$$\vdash_{\mathsf{HA}} \Box_T \Box_T \bot \to \Box_T^f \Box_T \bot \to \Box_T^f \bot \to \Box_T \bot,$$

so by Löb's Theorem, we get $\vdash_{\mathsf{HA}} \Box_T \bot$. We conclude that $\vdash_T \bot$. □

Now we prove the analogue of Corollary 2.2.6 for fast provability.

Lemma 2.3.3. *Suppose* $\mathsf{HA} \leq U \leq T$ *are theories and let* A *be an* $\mathcal{L}$*-formula. Then we have* $\vdash_{\mathsf{HA}} \Box_{U^T}^{u+1} A \leftrightarrow \Box_U^{u+1} A^T$. *In particular,* $\vdash_{\mathsf{HA}} \Box_{U^T}^f A \leftrightarrow \Box_U^f A^T$

Proof. We proceed by reflexive induction. First of all, by Corollary 2.2.6, we have

$$\vdash_{\mathsf{HA}} \Box_{U^T}^{0+1} A \leftrightarrow \Box_{U^T} A \leftrightarrow \Box_U A^T \leftrightarrow \Box_U^{0+1} A^T.$$

Furthermore, we have

$$\begin{aligned}
\vdash_{\mathsf{HA}} \Box_U(\Box_{U^T}^{u+1} A \leftrightarrow \Box_U^{u+1} A^T) \to \\
[\Box_{U^T}^{\mathsf{S}u+1} A &\leftrightarrow \Box_{U^T} \Box_{U^T}^{u+1} A \\
&\leftrightarrow \Box_U (\Box_{U^T}^{u+1} A)^T \\
&\leftrightarrow \Box_U \Box_{U^T}^{u+1} A \\
&\leftrightarrow \Box_U \Box_U^{u+1} A^T \\
&\leftrightarrow \Box_U^{\mathsf{S}u+1} A^T].
\end{aligned}$$

This completes the proof. □

Now we turn to *slow provability.* We will not give as many details as we did for fast provability, but instead we will refer to the paper [5]. There are two reasons for this. First of all, developing the theory of slow provability is rather involved, so reasons of space do not permit us to provide all the details. The second reason involves our intended usage of fast and slow provability. In Chapter 4, we will obtain results about the provability logic of fast provability. In order to understand and appreciate these results, it is important to know what fast provability is, exactly. Slow provability, on the other hand, will only be used as a tool to obtain results that themselves do not mention slow provability. In order to understand these results, it is not necessary to know all the details about slow provability.

In the paper [5], the authors define a certain 'fast-growing' total recursive function $F\colon \mathbb{N} \to \mathbb{N}$. There exists a Σ_1-formula $\varphi_F(x, y)$ representing F in HA. This means that the definition of F as a total recursive function is verifiable in HA, and we have

$$\vdash_{\mathsf{HA}} \forall y\, (\varphi_F(n, y) \leftrightarrow y = F(n)) \quad \text{for all } n \in \mathbb{N}.$$

The Σ_1-formula $F(x)\downarrow$, which we read as '$F(x)$ is defined', is shorthand for $\exists y\, \varphi_F(x,y)$. We clearly have that $\vdash_{\mathsf{HA}} F(n)\downarrow$ for all $n \in \mathbb{N}$. However, the fast-growing function F is constructed is such a way that F is not provably total. That is, we do *not* have $\vdash_{\mathsf{HA}} F(x)\downarrow$. Now we are ready to define slow provability.

Definition 2.3.2. The theory *slow Heyting Arithmetic*, denoted sHA, is given by the axiom formula

$$\mathsf{Ax}_{\mathsf{sHA}}(x) :\leftrightarrow \mathsf{Ax}_{\mathsf{EA}}(x) \vee (\mathsf{Ax}_{\mathsf{HA}}(x) \wedge \exists y \geq x\,(F(y)\downarrow)).$$

Intuitively, we demand that the axioms we use must not be 'too large': they must be not be so large that they are entirely beyond the domain of F. Since F is in fact total, we have $\mathbb{N} \models \mathsf{Ax}_{\mathsf{sHA}}(x) \leftrightarrow \mathsf{Ax}_{\mathsf{HA}}(x)$, which means that HA = sHA. We also clearly have that $\vdash_{\mathsf{HA}} \mathsf{Ax}_{\mathsf{sHA}}(x) \to \mathsf{Ax}_{\mathsf{HA}}(x)$, so sHA $\leq$ HA. However, as we shall show shortly, we do *not* have HA $\leq$ sHA. So from the viewpoint of HA, the requirement that the axioms must not be too large is a genuine one.

Even though the base theory used in the paper [5] is the classical theory PA, many results carry over to the present case. The most important of these is:

Proposition 2.3.4. *We have* $\vdash_{\mathsf{HA}} \mathsf{Bew}_{\mathsf{HA}}(x) \to \Box_{\mathsf{HA}}\mathsf{Bew}_{\mathsf{sHA}}(x)$*, and in particular, we have* $\vdash_{\mathsf{HA}} \Box_{\mathsf{HA}^{\mathsf{sHA}}} A \leftrightarrow \Box_{\mathsf{HA}} A^{\mathsf{sHA}}$ *for all $\mathcal{L}$-formulae A.*

Proof. The first statement is proven as in [5], Corollary 15, taking S_n to be the theory axiomatized by the axioms of HA having Gödel number at most n. The second statement follows from Corollary 2.2.6 with $U \equiv$ HA and $T \equiv$ sHA. □

The converse of this result, which is valid for the classical case, does *not* carry over to the current setting, because the authors use a model theoretic argument to derive this result. However, we will only need a very weak version of this converse, which we can 'steal' from the classical case. This proof was suggested by A. Visser.

Proposition 2.3.5. *(i) For all Σ_1-sentences S, we have* $\vdash_{\mathsf{HA}} \Box_{\mathsf{HA}}\Box_{\mathsf{sHA}} S \to \Box_{\mathsf{HA}} S$.

(ii) We have $\nvdash_{\mathsf{HA}} \Box_{\mathsf{HA}}\bot \to \Box_{\mathsf{sHA}}\bot$*. In particular,* HA $\not\leq$ sHA.

Proof. (i) We define the analogue of slow provability for PA, e.g. by setting

$$\mathsf{Ax}_{\mathsf{sPA}}(x) :\leftrightarrow \mathsf{Ax}_{\mathsf{EA}}(x) \vee (\mathsf{Ax}_{\mathsf{PA}}(x) \wedge \exists y \geq x\,(F(y)\downarrow)).$$

Since $\vdash_{\mathsf{HA}} \mathsf{Ax}_{\mathsf{HA}}(x) \to \mathsf{Ax}_{\mathsf{PA}}(x)$, it is clear that HA $\leq$ PA and sHA $\leq$ sPA. We know from [5], Theorem 4, that $\vdash_{\mathsf{PA}} \Box_{\mathsf{PA}}\Box_{\mathsf{sPA}} S \to \Box_{\mathsf{PA}} S$. So we get

$$\vdash_{\mathsf{PA}} \Box_{\mathsf{HA}}\Box_{\mathsf{sHA}} S \to \Box_{\mathsf{PA}}\Box_{\mathsf{sPA}} S \to \Box_{\mathsf{PA}} S \to \Box_{\mathsf{HA}} S,$$

where the final step holds since PA is, verifiably in HA, Σ_1-conservative over HA. We notice that $\Box_{\mathsf{HA}}\Box_{\mathsf{sHA}} S \to \Box_{\mathsf{HA}} S$ is equivalent, over HA, to a Π_2-sentence, that is, a sentence of the form $\forall x\, R(x)$, where $R \in \Sigma_1$. Since PA is Π_2-conservative over HA, we also find that $\vdash_{\mathsf{HA}} \Box_{\mathsf{HA}}\Box_{\mathsf{sHA}} S \to \Box_{\mathsf{HA}} S$.
(ii) Suppose that $\vdash_{\mathsf{HA}} \Box_{\mathsf{HA}}\bot \to \Box_{\mathsf{sHA}}\bot$. Since $\bot \in \Sigma_1$, we have

$$\vdash_{\mathsf{HA}} \Box_{\mathsf{HA}}\Box_{\mathsf{HA}}\bot \to \Box_{\mathsf{HA}}\Box_{\mathsf{sHA}}\bot \to \Box_{\mathsf{HA}}\bot,$$

so by Löb's Theorem, we get $\vdash_{\mathsf{HA}} \Box_{\mathsf{HA}}\bot$. But then HA is inconsistent, contradiction. □

2.4 Intuitionistic Modal Logic

In this section, we briefly review intuitionistic modal logic, abbreviated IML, and we define the system of IML that will be relevant to us. The language $\mathcal{L}_\Box$ of IML has a countable set of propositional constants, the absurdity sign $\bot$, the usual binary connectives $\wedge$, $\vee$ and $\to$, and the unary sentential operator $\Box$. We shall also use $\mathcal{L}_\Box$ to denote the set of all $\mathcal{L}_\Box$-formulae. As our proof system, we pick a Hilbert-style system that has two inference rules:

$$\frac{A \qquad A \to B}{B} \to\mathrm{E} \qquad \text{and} \qquad \frac{A}{\Box A}\,\mathrm{Nec}\ .$$

Definition 2.4.1. (i) The set $\mathsf{iK} \subseteq \mathcal{L}_\Box$ is the smallest set that contains:

(a) all ($\mathcal{L}_\Box$-substitution instances of) tautologies of intuitionistic propositional logic;

(b) all $\mathcal{L}_\Box$-sentences of the form $\Box(A \to B) \to (\Box A \to \Box B)$, where $A, B \in \mathcal{L}_\Box$,

and is closed under $\to$E and Nec.

(ii) A *theory for IML* will be a set T that satisfies $\mathsf{iK} \subseteq T \subseteq \mathcal{L}_\Box$ and is closed under $\to$E and Nec. If $A \in \mathcal{L}_\Box$ and $\Gamma \subseteq \mathcal{L}_\Box$, we write $\Gamma \vdash_T A$ if there exists a finite subset $\Gamma_0 \subseteq \Gamma$ such that $\bigwedge \Gamma_0 \to A$ is in T..

(iii) The theory iGL is the smallest theory for IML that contains iK and all sentences of the form $\Box(\Box A \to A) \to \Box A$, where $A \in \mathcal{L}_\Box$.

(iv) The theory iGLC is the smallest theory for IML contains iGL and all sentences of the forms $A \to \Box A$, where $A \in \mathcal{L}_\Box$.

We now proceed to define the semantics of modal intuitionistic logic.

Definition 2.4.2. (i) Consider a triple $\langle W, \preceq, \sqsubset \rangle$, where W is a nonempty set and $\preceq$ and $\sqsubset$ are binary relations on W. We say that this triple satisfies the *model property* if $\preceq \circ \sqsubset$ is a subrelation of $\sqsubset$. That is, for all $w, v, u \in W$ we should have: if $w \preceq v \sqsubset u$, then $w \sqsubset u$.

(ii) A *frame for IML* is a triple $\langle W, \preceq, \sqsubset \rangle$, where W is a nonempty set and $\preceq$ and $\sqsubset$ are binary relations on W, such that: $\langle W, \preceq \rangle$ is a poset and $\langle W, \preceq, \sqsubset \rangle$ satisfies the model property.

(iii) A *model for IML* is a quadruple $\langle W, \preceq, \sqsubset, V \rangle$, where $\langle W, \preceq, \sqsubset \rangle$ is a frame for IML and V is a relation (called the *valuation*) between W and the proposition letters from $\mathcal{L}_\Box$ satisfying:

$$w \preceq v \text{ and } wVp \text{ implies } vVp,$$

for all $w, v \in W$ and proposition letters p.

(iv) Let $M = \langle W, \preceq, \sqsubset, V \rangle$ be a model for IML, let $w \in W$ and let $A \in \mathcal{L}_\Box$. We define the forcing relation $M, w \Vdash A$ by recursion on A, as follows. For all $B, C \in \mathcal{L}_\Box$, we set:

(a) $M, w \Vdash p$ iff wVp for all proposition letters p;

(b) $M, w \Vdash B \wedge C$ iff $M, w \Vdash B$ and $M, w \Vdash C$;

(c) $M, w \Vdash B \vee C$ iff $M, w \Vdash B$ or $M, w \Vdash C$;

(d) $M, w \Vdash B \to C$ iff for all $v \in W$ such that $w \preceq v$ and $M, v \Vdash B$, we have $M, v \Vdash C$;

(e) $M, w \Vdash \Box B$ iff for all $v \in W$ such that $w \sqsubset v$, we have $M, v \Vdash B$.

If M is understood, we just write $w \Vdash A$ instead of $M, w \Vdash A$. We write $M \Vdash A$ if $M, w \Vdash A$ for all $w \in W$, in which case we say that A is *valid* on M. Given a frame $\langle W, \preceq, \sqsubset \rangle$ for IML, we say that $A \in \mathcal{L}_\Box$ is valid on this frame iff for all models $M = \langle W, \preceq, \sqsubset, V \rangle$ for IML, we have that A is valid on M.

Usually, one writes 'R' for the modal relation we call '$\sqsubset$' here. Our notation has certain advantages that will become apparent in the next chapter. We impose the model property on our frames because we want the following result:

Proposition 2.4.1 (Preservativity of Knowledge)**.** *Let $M = \langle W, \preceq, \sqsubset, V \rangle$ be a model for IML. If we have $w, v \in W$ and $A \in \mathcal{L}_\Box$ such that $w \Vdash A$ and $w \preceq v$, then $v \Vdash A$.*

Proof. We proceed by induction on the complexity of A. The base case and the induction steps for conjunction, disjunction and implication are trivial. So suppose that A is $\Box B$ and that we have $w, v \in W$ such that $w \preceq v$ and $w \Vdash \Box B$. Consider any $u \in W$ such that $v \sqsubset u$. Then $w \preceq v \sqsubset u$, so since $\langle W, \preceq, \sqsubset \rangle$ has the model property, we get $w \sqsubset u$. Since $w \Vdash \Box B$, it follows that $u \Vdash B$. Since u was arbitrary, we can conclude that $v \Vdash \Box B$, as desired. □

For our purposes, the relevant frame properties are the following.

Definition 2.4.3. Let $\langle W, \preceq, \sqsubset \rangle$ be a frame for IML.

(i) We say that this frame if *transitive* if $\sqsubset \circ \sqsubset$ is a subrelation of $\sqsubset$.

(ii) We say that this frame if *semi-transitive* if $\sqsubset \circ \sqsubset$ is a subrelation of $\sqsubset \circ \preceq$.

(iii) We say that this frame is *realistic* if $\sqsubset$ is a subrelation of $\preceq$.

(iv) We say that this frame is *conversely well-founded* if every nonempty subset of W has a maximal element w.r.t. $\sqsubset$.

We say that a model for IML has one of the properties mentioned above iff the underlying frame has it.

The terminology from (iii) is not standard and was suggested by R. Iemhoff. The idea behind it is as follows. We can view $\sqsubset$ as an accessibility relation that is relative to the various worlds, while $\preceq$ represents the 'real' accessibility between worlds. If, in a realistic frame, a world w thinks that some world v is accessible, then v is also *really* accessible from w. We observe that, due to the model property, a realistic frame is automatically transitive. Indeed, suppose that $\langle W, \preceq, \sqsubset \rangle$ is a realistic frame for IML and suppose we have $w, v, u \in W$ such that $w \sqsubset v \sqsubset u$. Then we also have $w \preceq v \sqsubset u$, so $w \sqsubset u$ follows, as desired.

Now we relate our frame properties to the axioms of iGLC.

Proposition 2.4.2. *Let $F = \langle W, \preceq, \sqsubset \rangle$ be a frame for intuitionistic modal logic.*

(*i*) *The sentence $\Box(\Box p \to p) \to \Box p$ is valid on F if and only if F is semi-transitive and conversely well-founded.*

(*ii*) *The sentence $p \to \Box p$ is valid on F if and only if F is realistic.*

In particular, all theorems of iGLC *are valid on all realistic and conversely well-founded frames.*

Proof. (i) This result is known from the literature. We refer the reader to the paper [6], Lemma 8.

(ii) First, suppose that F is realistic. Let V be a valuation on F, and suppose we have $w \in W$ such that $w \Vdash p$. If $v \in W$ is such that $w \sqsubset v$, then also $w \preceq v$, so by preservativity of knowledge, we get $v \Vdash p$. We conclude that $w \Vdash \Box p$, and thus that $p \to \Box p$ is valid on F.
Conversely, suppose that F is not realistic. Then there exist $w, v \in K$ such that $w \sqsubset v$, but also $w \not\preceq v$. We define a valuation V on F such that

$$xVp \quad \text{if and only if} \quad w \preceq x.$$

Then wVp, but since $w \sqsubset v$ and $\neg(vVp)$, we also have $w \nVdash \Box p$. We conclude that $w \nVdash p \to \Box p$ and thus that $p \to \Box p$ is not valid on F.
The final statement is easily proven by an induction on iGLC-proofs. □

In order to get a completeness theorem, we need the following terminology.

Definition 2.4.4. Let T be a theory for intuitionistic modal logic.

(i) A set $X \subseteq \mathcal{L}_\Box$ is called *adequate* if it is closed under taking subformulae.

(ii) Suppose $X \subseteq \mathcal{L}_\Box$ is adequate. A set $S \subseteq X$ is called *X-saturated* if the following hold:

  (a) S is consistent, that is, $S \nvdash_T \bot$;

  (b) if $A \in X$ and $S \vdash_T A$, then $A \in S$;

  (c) if $A \vee B \in S$, then $A \in S$ or $B \in S$.

Notice that the converse of item (b) also holds: if $A \in S$, then clearly $A \in X$ and $S \vdash_T A$. We will need the following result.

Lemma 2.4.3 (Extension Lemma)**.** *Let T be a theory for intuitionistic modal logic and let $X \subseteq \mathcal{L}_\Box$ be an adequate set. Suppose we have $R \subseteq X$ and $A \in \mathcal{L}_\Box$ such that $R \nvdash_T A$. Then there exists an X-saturated set $S \supseteq R$ such that $S \nvdash_T A$.*

Proof. We fix an enumeration $B_0, B_1, B_2, \ldots$ of the formulae in X such that every element of X occurs infinitely many times in the enumeration. We define the sequence $S_0 \subseteq S_1 \subseteq S_2 \subseteq \ldots$ by recursion. First of all, we set $S_0 = R$. Now suppose that S_n has been defined. If $S_n \nvdash_T B_n$, then S_{n+1} is just S_n. If $S_n \vdash_T B_n$, then

$$S_{n+1} = \begin{cases} S_n \cup \{B_n\} & \text{if } B_n \text{ is not a disjunction;} \\ S_n \cup \{B_n, C\} & \text{if } B_n \text{ is } C \vee D, \text{ and } S_n \cup \{C\} \nvdash_T A; \\ S_n \cup \{B_n, D\} & \text{if } B_n \text{ is } C \vee D, \text{ and } S_n \cup \{C\} \vdash_T A; \end{cases}$$

We define S as $\bigcup_{n \in \mathbb{N}} S_n$. Clearly, we have $S_n \subseteq X$ for all $n \in \mathbb{N}$, so $S \subseteq X$.
Now we use induction on n to prove that $S_n \nvdash_T A$ for all $n \in \mathbb{N}$. For $n = 0$, this holds by assumption. Now suppose that $S_n \nvdash_T A$ for a certain $n \in \mathbb{N}$; we need to show that $S_{n+1} \nvdash_T A$. If $S_n \nvdash_T B_n$, then this holds trivially. So suppose that $S_n \vdash_T B_n$. Then we must have that $S_n \cup \{B_n\} \nvdash_T A$, so if B_n is not a disjunction, then we are also done. So suppose that B_n is $C \vee D$. If $S_n \cup \{C\} \nvdash_T A$, then we also have $S_{n+1} = S_n \cup \{B_n, C\} \nvdash_T A$, so we are done. Finally, suppose that $S_n \cup \{C\} \vdash_T A$. Then we cannot have $S_n \cup \{D\} \vdash_T A$. Indeed, if we have both $S_n \cup \{C\} \vdash_T A$ and $S_n \cup \{D\} \vdash_T A$, then also $S_n \cup \{C \vee D\} \vdash_T A$, which is not the case. So $S_n \cup \{D\} \nvdash_T A$, and it follows that $S_{n+1} = S_n \cup \{B_n, D\} \nvdash_T A$, as desired. This completes the induction.

It follows that $S \nvdash_T A$, and in particular, S is consistent. We check that S is X-saturated. Now suppose that $C \in X$ and $S \vdash_T C$. Then there must be an $n \in \mathbb{N}$ such that $S_n \vdash_T C$. Let $m \geq n$ be minimal such that B_m is C. Then $S_m \vdash_T B_m$, so we get $B_m \in S_{m+1} \subseteq S$, that is $C \in S$. Finally, supppse that $C \vee D \in S$. Then there must be an $n \in \mathbb{N}$ such that $C \vee D \in S_n$. Let $m \geq n$ be minimal such that B_m is $C \vee D$. Then $B_m \in S_n \subseteq S_m$, so we certainly have $S_m \vdash_T B_m$. It follows that $C \in S_m \subseteq S$ or $D \in S_m \subseteq S$. This concludes the proof. □

Using the Extension Lemma, we can prove a sound- and completeness theorem for iGLC. This result also appears, in a stronger form, as Theorem 4.25 in [1].

Theorem 2.4.4. *Let A be an $\mathcal{L}_\Box$-sentence. Then $\vdash_{\mathsf{iGLC}} A$ if and only if A is valid on all finite irreflexive realistic frames.*

Proof. It is well-known that any finite irreflexive transitive frame is conversely well-founded. So if $\vdash_{\mathsf{iGLC}} A$, then A is indeed valid on all finite irreflexive realistic frames, by Proposition 2.4.2. Conversely, suppose that $\nvdash_{\mathsf{iGLC}} A$. Let X_0 be the set of subformulae of A, and let $X_1 = \{\Box B \mid B \in X_0\}$. Then $X := X_0 \cup X_1$ is an adequate set. We let W be the set of all X-saturated sets. Clearly, W is finite, and we have the subset relation $\subseteq$ on W. For $w, v \in W$, we write $w \sqsubset v$ if:

(i) whenever $B \in \mathcal{L}_\Box$ and $\Box B \in w$, we have $B \in v$;

(ii) there exists a $C \in \mathcal{L}_\Box$ such that $\Box C \notin w$ and $\Box C \in v$.

For $w \in W$ and p a proposition letter, we say that wVp if and only if $p \in w$. We clearly have: if wVp and $w \subseteq v$, then vVp. It is also not difficult to check that $\langle W, \subseteq, \sqsubset\rangle$ satisfies the model property. Finally, since $\nvdash_{\mathsf{iGLC}} A$, there exists a $w_0 \in W$ such that $w_0 \nVdash A$, by the Extension Lemma. In particular, W is nonempty, so $M = \langle W, \subseteq, \sqsubset, V\rangle$ is a model for intuitionistic modal logic.

We claim that the frame $\langle W, \subseteq, \sqsubset\rangle$ is irreflexive and realistic. Irreflexivity is immediate from the definition. Now suppose we have $w, v \in W$ such that $w \sqsubset v$, and $B \in w$. If $B \in X_0$, then $\Box B \in X_1 \subseteq X$ and $w \vdash_{\mathsf{iGLC}} \Box B$, so $\Box B \in w$. Since $w \sqsubset v$ we get $B \in v$. Now suppose that $B \in X_1$. Then B is $\Box C$ for some $C \in X_0$. Since $w \sqsubset v$, we get $C \in v$. This means that $v \vdash_{\mathsf{iGLC}} B$, so $B \in v$. In both cases, we get $B \in v$, so we conclude that $w \subseteq v$, as desired.

Now we show that $w \Vdash B$ if and only if $B \in w$, for all $w \in W$ and $B \in X$. We proceed by induction on the complexity of B.

At For proposition letters, the result holds by the definition of V.

$\wedge$ Suppose that B is $C \wedge D$ and that the result holds for C and D. If $w \in W$, then $w \Vdash C \wedge D$ iff $w \Vdash C$ and $w \Vdash D$, iff $C \in w$ and $D \in w$. Now suppose that $C \in w$ and $D \in w$. Then $w \vdash_{\mathsf{iGLC}} C \wedge D$ and $C \wedge D \in X$, so $C \wedge D \in w$. Conversely, suppose that $C \wedge D \in w$. Then $w \vdash_{\mathsf{iGLC}} C, D$ and $C, D \in X$, so we get $C \in w$ and $D \in w$.

$\vee$ Suppose that B is $C \vee D$ and that the result holds for C and D. If $w \in W$, then $w \Vdash C \vee D$ iff $w \Vdash C$ or $w \Vdash D$, iff $C \in w$ or $D \in w$. Suppose that $C \in w$ or $D \in w$. Then in both cases, we have $w \vdash_{\mathsf{iGLC}} C \vee D$. Since $C \vee D \in X$, we get $C \vee D \in w$. Conversely, if $C \vee D \in w$, then $C \in w$ or $D \in w$ since w is X-saturated.

$\to$ Suppose that B is $C \to D$ and that the result holds for C and D. If $w \in W$, then $w \Vdash C \to D$ iff for all $v \supseteq w$, we have that $v \Vdash C$ implies $v \Vdash D$. And this holds iff for all $v \supseteq w$, we have that $C \in v$ implies $D \in v$. Now suppose that $C \to D \in w$ and that we have $v \supseteq w$ such that $C \in v$. Then also $C \to D \in v$, so $v \vdash_{\mathsf{iGLC}} D$. Since $D \in X$, we get $D \in v$. Conversely, suppose that $C \to D \notin w$. Since $C \to D \in X$, this means that $w \nvdash_{\mathsf{iGLC}} C \to D$, and hence $w \cup \{C\} \nvdash_{\mathsf{iGLC}} D$. Since $w \cup \{C\} \subseteq X$, we can use the Extension Lemma to find a $v \in W$ such that $w \cup \{C\} \subseteq v$ and $v \nvdash_{\mathsf{iGLC}} D$. Then $w \subseteq v$, $C \in v$, and $D \notin v$, so it follows that $w \nVdash B \to C$.

- $\Box$ Suppose that B is $\Box C$ and that the result holds for C. If $w \in W$, then $w \Vdash \Box C$ iff for all $v \sqsupset w$, we have $v \Vdash C$. And this holds iff for all $v \sqsupset w$, we have $C \in v$. Now suppose that $\Box C \in w$ and that we have $v \sqsupset w$. Then by the definition of $\sqsubset$, we get $C \in v$. Conversely, suppose that $\Box C \notin w$. Consider the set $R = \{D \in \mathcal{L}_\Box \mid \Box D \in w\} \cup \{\Box C\} \subseteq X$. Suppose that $R \vdash_{\mathsf{iGLC}} C$. Then $\{D \in \mathcal{L}_\Box \mid \Box D \in w\} \vdash_{\mathsf{iGLC}} \Box C \to C$, so we also get $\{\Box D \in \mathcal{L}_\Box \mid \Box D \in w\} \vdash_{\mathsf{iGLC}} \Box(\Box C \to C)$. In particular, $w \vdash_{\mathsf{iGLC}} \Box(\Box C \to C)$, which yields $w \vdash_{\mathsf{iGLC}} \Box C$. However, we also have $\Box C \in X$, so we get $\Box C \in w$, contradiction. So $R \nvdash_{\mathsf{iGLC}} C$. By the Extension Lemma, there exists a $v \in W$ such that $R \subseteq v$ and $v \nvdash_{\mathsf{iGLC}} C$. We have $\{D \in \mathcal{L}_\Box \mid \Box D \in w\} \subseteq v$, $\Box C \notin w$ and $\Box C \in v$, so $w \sqsubset v$. Furthermore, we have $C \notin v$, so $w \nVdash \Box C$.

This completes the induction. Since $w_0 \nvdash_{\mathsf{iGLC}} A$, we have $A \notin w$. Since $A \in X$, we can apply the above result to conclude that $w_0 \nVdash A$. So A is not valid on the finite irreflexive realistic frame $\langle W, \subseteq, \sqsubset \rangle$. $\Box$

Chapter 3

An Abstract Arithmetical Completeness Theorem

In this chapter, we prove a completeness theorem for certain kinds of provability logics. We prove the theorem in a rather abstract form, not yet mentioning any specific provabiility predicates. In Section 3.1, we introduce the general framework and define the required Solovay function along with the intended realization of the propositional letters of $\mathcal{L}_\Box$. Section 3.2 is of a rather technical nature and forms the heart of the proof. Here we show that the realization we defined commutes with the logical operators of $\mathcal{L}_\Box$. In Section 3.3, we formulate the completeness theorem and use the preceding material to prove it.

3.1 Definition of the Solovay Function

The general setting of this chapter is given by the following definition.

Definition 3.1.1. Let $T \supseteq \mathsf{HA}$ be a theory and let $S(x)$ and $R(x)$ be Σ_1-formulae in one free variable. If A is an $\mathcal{L}$-sentence, we write $\Box A$ for $S(\ulcorner A \urcorner)$. We also write $\triangle A$ for $R(\ulcorner A \urcorner)$. We say that (S, R) is a *good pair* for T if the following conditions are satisfied:

(i) S and R are provability predicates for T;

(ii) if $\mathbb{N} \models \Box A$, then $\vdash_T A$, for all $\mathcal{L}$-sentences A;

(iii) $\vdash_T \mathrm{SLP}_\triangle$ (or equivalently, $\vdash_T \mathrm{CP}_\triangle$);

(iv) $\vdash_{\mathsf{HA}} \Box\triangle S \to \Box S$ for all Σ_1-sentences S.

We immediately observe that, if these clauses apply and S is a Σ_1-sentence, then we also have $\vdash_{\mathsf{HA}} \triangle S \to \Box\triangle S \to \Box S$. We also notice that, since $\mathsf{HA} \subseteq T$, we have that $\vdash_{\mathsf{HA}} A$ implies $\vdash_T A$, which implies $\vdash_{\mathsf{HA}} \Box A$ and $\vdash_{\mathsf{HA}} \triangle A$ for all $\mathcal{L}$-sentences A.

Remark 3.1.1. We remark that the definition of a good pair does not occur anywhere in the literature. This definition is extremely artificial and tailor made to obtain the result of this chapter. $\Diamond$

In the remainder of this chapter, we suppose that a theory T extending HA and a good pair (S, R) for T are given. We also use $\Box$ and $\triangle$ as defined above.

Let $M_0 = \langle W_0, \preceq_0, \sqsubset_0, V_0\rangle$ be a finite irreflexive realistic model for IML such that W_0 has a least element w.r.t. $\preceq_0$. Let $r > 0$ be the cardinality of W_0. We assume that $W_0 = \{1, \ldots, r\}$ and that the node r is the least element of W_0 w.r.t. $\preceq_0$. Now we expand M_0 to a new model $M = \langle W, \preceq, \sqsubset, V\rangle$ for IML. Intuitively, we append a copy of $1 + \omega^{\mathrm{op}}$ (in the $\sqsubset$-order relation) to the node r. Formally, we do this as follows. We take $W = \mathbb{N} \supset W_0$. The relation $\preceq$ is defined by:

$$\begin{aligned} i \preceq j \quad \text{iff} \quad & 1 \le i, j \le r \text{ and } i \preceq_0 j, \\ & \text{or } i > r \text{ and } 1 \le j \le i, \\ & \text{or } i = 0, \end{aligned}$$

for all $i, j \in \mathbb{N}$. The relation $\sqsubset$ is defined by:

$$\begin{aligned} i \sqsubset j \quad \text{iff} \quad & 1 \le i, j \le r \text{ and } i \sqsubset_0 j, \\ & \text{or } i > r \text{ and } 1 \le j < i, \\ & \text{or } i = 0 \text{ and } j > 0, \end{aligned}$$

for all $i, j \in \mathbb{N}$. Finally, V is defined by:

$$iVp \quad \text{iff} \quad 1 \le i \le r \text{ and } iV_0p,$$

for all $i \in \mathbb{N}$ and proposition letters p.

We can prove that M is again a realistic irreflexive model for IML; but of course M is no longer finite. However, M *is* conversely well-founded, so M still validates all theorems of iGLC. Since $\preceq_0$ and $\sqsubset_0$ are finite relations, we can give Δ_0-definitions of these relations inside HA. Now we can formalize the definitions of $\preceq$ and $\sqsubset$ given above in order to obtain Δ_0-definitions of $\preceq$ and $\sqsubset$ inside HA. Then HA verifies the relevant properties of M: that $\preceq$ is a poset, that $\sqsubset$ is irreflexive, that $\langle W, \preceq, \sqsubset\rangle$ has the model property, and that this frame is realistic. E.g. by verification of the model property we mean that $\vdash_{\mathsf{HA}} x \preceq y \wedge y \sqsubset z \to x \sqsubset z$. Since $\preceq$ is defined by a Δ_0-formula, we have: if $i \preceq j$, then $\vdash_{\mathsf{HA}} i \preceq j$, and if $i \not\preceq j$, then $\vdash_{\mathsf{HA}} \neg(i \preceq j)$. A similar result holds for $\sqsubset$. Moreover, by Proposition 2.1.3(i), we have safely make case distinctions like $x \preceq y \vee \neg(x \preceq y)$ inside HA. Since we assumed that $\mathsf{HA} \subseteq T$, all these remarks also hold for T instead of HA.

For an $A \in \mathcal{L}_\Box$, we define the set $[\![A]\!]$ as $\{i \in \mathbb{N} \mid i \Vdash A\}$. The model M is constructed in such a way that the following result holds.

Lemma 3.1.1. *If $A \in \mathcal{L}_\Box$, then $[\![A]\!]$ is finite or $[\![A]\!] = \mathbb{N}$.*

Proof. We have to show the following: if $i \in [\![A]\!]$ for all $i > 0$, then $0 \in [\![A]\!]$. We proceed by induction on the complexity of A. The atomic case clearly holds, and the steps for $\wedge$ and $\vee$ are trivial. Now suppose that A is $B \to C$ and that the claim holds for B and C. Suppose that $i \in [\![B \to C]\!]$ for all $i > 0$, and that $0 \notin [\![B \to C]\!]$. Then we must have $0 \in [\![B]\!]$ and $0 \notin [\![C]\!]$ as well. By the induction hypothesis, $i \notin C$ for some $i > 0$. However, since $0 \preceq i$, we also have $i \in [\![B]\!]$, so $i \notin [\![B \to C]\!]$, contradiction. Finally, suppose that A is $\Box B$ and that the claim holds for B. Suppose that $i \in [\![\Box B]\!]$ for all $i > 0$. We should show that $0 \in [\![\Box B]\!]$. By preservativity of knowledge, it suffices to show that $j \in [\![B]\!]$ for all $j \geq r$. But for such j, we have $j+1 \in [\![\Box B]\!]$ by assumption, and $j+1 \sqsubset j$, so we indeed have $j \in [\![B]\!]$. $\Box$

We now proceed to define the Solovay function. Our models are equipped with *two* relations, as opposed to just one in the classical case, and we need to find some way to incorporate this into the Solovay function. A. Visser suggested to use two separate provability predicates to take care of the relations $\preceq$ and $\sqsubset$. This is where our good pair comes in. Since $S(x)$ and $R(x)$ are Σ_1-formulae, we can write $S(x)$ as $\exists y \mathsf{Prf}_\Box(y, x)$ and $R(x)$ as $\exists y \mathsf{Prf}_\triangle(y, x)$, where $\mathsf{Prf}_\Box$ and $\mathsf{Prf}_\triangle$ are Δ_0-formulae.

Let $\langle \cdot, \cdot \rangle \colon \mathbb{N}^2 \to \mathbb{N}$ be a primitive recursive pairing function that can be formulated inside HA using a Δ_0-formula. Let $p_0 \colon \mathbb{N} \to \mathbb{N}$ be the primitive recursive function that gives the projection onto the first coordinate. By replacing $\mathsf{Prf}_\Box(y, x)$ with $\exists z \leq y\, (y = \langle x, z \rangle \wedge \mathsf{Prf}_\Box(z, x))$, we may assume without loss of generality that

$$\vdash_{\mathsf{HA}} \mathsf{Prf}_\Box(y, x) \to x = p_0(y). \tag{3.1}$$

We do the same for $\mathsf{Prf}_\triangle$.

In the sequel, we write $x \prec y$ for $x \preceq y \wedge \neg(x = y)$ and $x \sqsubseteq y$ for $x \sqsubset y \vee x = y$. We define the function $h \colon \mathbb{N} \to \mathbb{N}$ by $h(0) = 0$ and

$$h(k+1) = \begin{cases} m & \text{if } h(k) \sqsubset m \text{ and } \mathsf{Prf}_\Box(k, \ulcorner \exists x \neg(h(x) \sqsubseteq m) \urcorner); \\ n & \text{if } h(k) \prec n \text{ and } \mathsf{Prf}_\triangle(k, \ulcorner \exists y \neg(h(y) \preceq n) \urcorner); \\ h(k) & \text{if neither of these apply.} \end{cases}$$

Here x and y are two (syntactically) distinct variables, so by our assumption (3.1) above, the first two clauses can never apply simultaneously. Using (3.1) again, we also see that m as in the first clause, if it exists, is unique, and similarly for the second clause. So h is well-defined. Using the Diagonalization Lemma, we can give a Σ_1-definition of h inside EA. The argument that h is well-defined above can then be formalized inside HA, so HA proves that h is a function. We also have $\vdash_{\mathsf{HA}} x \leq y \to h(x) \preceq h(y)$.

Notice that it is in some sense 'easier' to move along $\sqsubset$ than it is to move along $\preceq$. We have $\vdash_{\mathsf{HA}} \exists y \neg(h(y) \preceq m) \to \exists x \neg(h(x) \sqsubseteq m)$. Since $\mathsf{HA} \subseteq T$ and R is a provability predicate for T, we also find that $\vdash_{\mathsf{HA}} \triangle(\exists y \neg(h(y) \preceq m)) \to \triangle(\exists x \neg(h(x) \sqsubseteq m))$. We also observe that $\exists x \neg(h(x) \sqsubseteq m)$ is equivalent, over EA, to a Σ_1-sentence. This means that we also have $\vdash_{\mathsf{HA}} \triangle(\exists x \neg(h(x) \sqsubseteq m)) \to \Box(\exists x \neg(h(x) \sqsubseteq m))$. We conclude that

$$\vdash_{\mathsf{HA}} \triangle(\exists y \neg(h(y) \preceq m)) \to \Box(\exists x \neg(h(x) \sqsubseteq m)), \tag{3.2}$$

for any $m \in \mathbb{N}$. We will need this in the sequel. We also need the following observation: if $i \neq 0$ is a natural number, then

$$\vdash_{\mathsf{HA}} \neg(x \preceq i) \leftrightarrow \bigvee_{j \in U} x = j, \tag{3.3}$$

where $U = \{j \in \mathbb{N} \mid j \not\preceq i\}$ is finite. In other words, if HA knows that $x \not\preceq i$ for some standard $i \neq 0$, then HA knows that x is some standard number as well. For $\sqsubseteq$, a similar remark applies.

To close this section, we define, for any $A \in \mathcal{L}_\Box$, the $\mathcal{L}$-sentence

$$[A] = \begin{cases} \bigvee_{i \in [\![A]\!]} \exists x\,(h(x) = i) & \text{if } [\![A]\!] \text{ is finite;} \\ \top & \text{if } [\![A]\!] = \mathbb{N}. \end{cases}$$

We notice that $[A]$ is always a Σ_1-sentence.

3.2 Preservation of the Logical Structure

In this rather technical section, we show that $[\cdot]$ commutes with all the logical operators figuring in $\mathcal{L}_\Box$. The proofs in this section will become increasingly difficult. We adopt all the notation introduced in the previous section.

Lemma 3.2.1. *We have* $\vdash_{\mathsf{HA}} [B \vee C] \leftrightarrow [B] \vee [C]$ *for* $B, C \in \mathcal{L}_\Box$.

Proof. This is immediate from the definition of $[\cdot]$. □

Lemma 3.2.2. *We have* $\vdash_{\mathsf{HA}} [B \wedge C] \leftrightarrow [B] \wedge [C]$ *for* $B, C \in \mathcal{L}_\Box$.

Proof. If $[\![B]\!] = \mathbb{N}$, then $[\![B \wedge C]\!] = [\![C]\!]$, so we have $\vdash_{\mathsf{HA}} [B \wedge C] \leftrightarrow [C] \leftrightarrow [B] \wedge [C]$. Similarly, the result follows if $[\![C]\!] = \mathbb{N}$. So suppose that $[\![B]\!]$ and $[\![C]\!]$ are both finite; then $[\![B \wedge C]\!]$ is finite as well.
The '$\rightarrow$'-statement is immediate in this case. For, the other direction, we should show that $\vdash_{\mathsf{HA}} \exists x\,(h(x) = i) \wedge \exists y\,(h(y) = j) \rightarrow [B \wedge C]$ whenever $i \in [\![B]\!]$ and $j \in [\![C]\!]$. First of all, we notice that $\vdash_{\mathsf{HA}} \exists x\,(h(x) = i) \wedge \exists y\,(h(y) = j) \rightarrow i \preceq j \vee j \preceq i$. Indeed, reason inside HA and suppose we have x and y such that $h(x) = i$ and $h(y) = j$. Since $x \leq y \vee y \leq x$ and $x \leq y \rightarrow h(x) \preceq h(y)$, we can conclude that $i \preceq j \vee j \preceq i$, as desired.
Now, if i and j are incomparable w.r.t. $\preceq$, then $\vdash_{\mathsf{HA}} \neg(i \preceq j \vee j \preceq i)$, so by the above we have $\vdash_{\mathsf{HA}} \neg(\exists x\,(h(x) = i) \wedge \exists y\,(h(y) = j))$, in which case the result is clear. If i and j are comparable w.r.t. $\preceq$, then assume without loss of generality that $i \preceq j$. Then $j \in [\![B \wedge C]\!]$, so

$$\vdash_{\mathsf{HA}} \exists x\,(h(x) = i) \wedge \exists y\,(h(y) = j) \rightarrow \exists y\,(h(y) = j) \rightarrow [B \wedge C],$$

as desired. □

Lemma 3.2.3. *We have* $\vdash_T [B \rightarrow C] \leftrightarrow ([B] \rightarrow [C])$ *for* $B, C \in \mathcal{L}_\Box$.

Proof. If $[\![B \rightarrow C]\!] = \mathbb{N}$, then $[\![B]\!] \subseteq [\![C]\!]$, so $[B \rightarrow C]$ and $[B] \rightarrow [C]$ are both equivalent to $\top$, even over HA. Now suppose that $[\![B \rightarrow C]\!]$ is finite.
We first treat the $\leftarrow$-direction. Let $j_0, \ldots, j_{s-1} \neq 0$ be the $\preceq$-maximal elements j of $\mathbb{N}$ such that $j \notin [\![B \rightarrow C]\!]$. Then for all $t < s$, we have $j_t \in [\![B]\!]$ and $j_t \notin [\![C]\!]$. Using the fact that $\prec$ is also a conversely well-founded relation, we can show that for all $i \in \mathbb{N}$, we have $i \in [\![B \rightarrow C]\!]$ if and only if $i \not\preceq j_t$ for all $t < s$.
Now we reason inside HA. Suppose that $[B] \rightarrow [C]$ and $\triangle[B \rightarrow C]$. Since $[B \rightarrow C] \rightarrow \exists y \neg(h(y) \preceq j_t)$, we have $\triangle(\exists y \neg(h(y) \preceq j_t))$. Now let k_t satisfy $\mathsf{Prf}_\triangle(k_t, \ulcorner \exists y \neg(h(y) \preceq j_t) \urcorner)$. We distinguish three cases (which is constructively acceptable).

1. Suppose that $h(k_t) \prec j_t$ for some $t < s$. Then then by the definition of h, we get $h(k_t + 1) = j_t$. But $j_t \in [\![B]\!]$, so $[B]$ holds, so $[C]$ holds, and therefore $[B \to C]$ also holds.
2. Suppose that $h(k_t) = j_t$ for some $t < s$. Then $[B \to C]$ again follows.
3. Suppose that $\neg(h(k_t) \preceq j_t)$ for all $t < s$. Let $k = \max_{t<s} k_t$. Then we also know that $\neg(h(k) \preceq j_t)$ for all $t < s$. Indeed, suppose that $h(k) \preceq j_t$ for some t. Since $k_t \leq k$, we get $h(k_t) \preceq h(k) \preceq j_t$, so, since $\preceq$ is (provably) transitive, $h(k_t) \preceq j_t$, which we already excluded. So we indeed have $\neg(h(k) \preceq j_t)$ for all $t < s$. But then by (3.3) applied to $j_0, \ldots, j_{s-1}$, we see that $\bigvee_{j \in U} h(k) = j$, where $U = \{j \in \mathbb{N} \mid j \not\preceq j_t \text{ for all } t < s\}$ is a finite set. We see (outside HA) that $U \subseteq [\![B \to C]\!]$, so (inside HA again) we get $[B \to C]$.

We conclude that $\vdash_{\mathsf{HA}} ([B] \to [C]) \to (\triangle[B \to C] \to [B \to C])$. Since $\mathsf{HA} \subseteq T$ and $\vdash_T \mathrm{SLP}_\triangle$, we may conclude that $\vdash_T ([B] \to [C]) \to [B \to C]$.
The $\to$-direction is even provable in HA. Notice that $[\![B \wedge (B \to C)]\!] \subseteq [\![C]\!]$, so by Lemma 3.2.2, we have

$$\vdash_{\mathsf{HA}} ([B] \wedge [B \to C]) \to [B \wedge (B \to C)] \to [C],$$

so $\vdash_{\mathsf{HA}} [B \to C] \to ([B] \to [C])$. □

The idea of defining M from M_0 and the techniques used in the previous three lemmata were already introduced by A. Visser. The proof that $[\cdot]$ commutes with $\Box$ contains some new ideas. First, we need some auxilliary results.

Lemma 3.2.4. *Suppose $i > 0$ is a natural number. Then*

$$\vdash_{\mathsf{HA}} \exists x\,(h(x) = i) \to \Box(\exists y\,(i \prec h(y))).$$

Proof. Before we start proving the displayed sentence inside HA, we need to verify two auxilliary facts inside HA. First of all, we claim that

$$\vdash_{\mathsf{HA}} (\neg(h(y) \preceq i) \wedge h(x) = i) \to i \prec h(y).$$

Reason inside HA and assume the antecedent. If $y < x$, then $h(y) \preceq h(x) = i$, quod non. So $x \leq y$, which means that $i = h(x) \preceq h(y)$. But $h(y)$ cannot be equal to i, so $i \prec h(y)$, as desired. Now we also have:

$$\vdash_{\mathsf{HA}} (\exists y \neg(h(y) \preceq i) \wedge \exists x\,(h(x) = i)) \to \exists y\,(i \prec h(y)). \tag{3.4}$$

Secondly, we claim that

$$\vdash_{\mathsf{HA}} (\neg(h(y) \sqsubseteq i) \wedge h(x) = i \wedge h(x-1) \sqsubset i) \to i \prec h(y).$$

Again, reason inside HA and assume the antecedent. Suppose that $y < x$. Then $y \leq x - 1$, so $h(y) \preceq h(x-1) \sqsubset i$. Since our frame (provably) has the model property, we get $h(y) \sqsubset i$, contradiction. So $y \geq x$. But then $i = h(x) \preceq h(y)$ and $\neg(i = h(y))$, so $i \prec h(y)$, as desired. We also find:

$$\vdash_{\mathsf{HA}} \exists y\,(\neg(h(y) \sqsubseteq i)) \wedge \exists x\,(h(x) = i \wedge h(x-1) \sqsubset i) \to \exists y\,(i \prec h(y)). \tag{3.5}$$

Now we start the main part of the proof. Reason inside HA, and suppose that we have an x such that $h(x) = i$. Since h is (provably) a function, we can consider the *least* x such that $h(x) = i$. Then $x > 0$, and $h(x-1) \prec i$. Again, we make a constructively acceptable case distinction.

1. Suppose that $\neg(h(x-1) \sqsubset i)$. Then $\triangle(\exists y \neg(h(y) \preceq i))$ (otherwise, we wouldn't have moved up to i). Since $\exists x\,(h(x) = i)$ is a Σ_1-sentence, we also know get $\triangle(\exists x\,(h(x) = i))$. Using (3.4) and the properties of $\triangle$, we can conclude that $\triangle(\exists y\,(i \prec h(y))$. Since $\exists y\,(i \prec h(y))$ is a Σ_1-sentence, we also get $\Box(\exists y\,(i \prec h(y))$, as desired.
2. Suppose that $h(x-1) \sqsubset i$. Then, from the fact that we moved up to i, we can deduce that $\Box(\exists x \neg(h(x) \sqsubseteq i))$ or $\triangle(\exists y \neg(h(y) \preceq i))$. By (3.2), we can conclude that $\Box(\exists y \neg(h(y) \sqsubseteq i))$ in both cases. At this point, we have $\exists x\,(h(x) = i \wedge h(x-1) \sqsubset i)$. Since this is a Σ_1-sentence, we also get $\Box(\exists x\, h(x) = i \wedge h(x-1) \sqsubset i))$. Using (3.5) and the properties of $\Box$, we again find $\Box(\exists y\,(i \prec h(y))$, as desired. $\Box$

Lemma 3.2.5. *Let i, j be natural numbers such that $i \prec j$ and $\neg(i \sqsubset j)$. Then*

$$\vdash_{\mathsf{HA}} \exists x\,(h(x) = i) \wedge \exists y\,(h(y) = j) \rightarrow \triangle(\exists z\,(j \prec h(z)).$$

Proof. First of all, we notice that we also know that $i \prec j$ and $\neg(i \sqsubset j)$ inside HA. Now reason inside HA, and suppose that $\exists x\,(h(x) = i)$ and $\exists y\,(h(y) = j)$. Since h is (provably) a function, we can consider the least y such that $h(y) = j$. Then $y > 0$, and $h(y-1) \prec j$. Consider an x such that $h(x) = i$. Suppose that $y \leq x$. Then $j = h(y) \prec h(x) = i \prec j$, which is a contradiction since $\preceq$ is (provably) antisymmetric. So $x < y$, which also means $x \leq y - 1$. Now we get $i = h(x) \preceq h(y-1)$.
If $h(y-1) \sqsubset j$, then $i \preceq h(y-1) \sqsubset j$, so $i \sqsubset j$. But we also have $\neg(i \sqsubset j)$, contradiction. So $\neg(h(y-1) \sqsubset j)$. Now we can use the exact same reasoning as in case 1 in the proof of Lemma 3.2.4 (with j instead of i, and y instead of x) to arrive at $\triangle(\exists z\,(j \prec h(z))$, as desired. $\Box$

Now that we have proven these tedious lemmata, we can derive our crucial result.

Lemma 3.2.6. *We have $\vdash_{\mathsf{HA}} [\Box B] \leftrightarrow \Box[B]$ for all $B \in \mathcal{L}_\Box$.*

Proof. If $[\![B]\!] = \mathbb{N}$, then $[\![\Box B]\!] = \mathbb{N}$ as well, and we see that $[\Box B]$ and $\Box[B]$ are both equivalent to $\top$ over HA. Now suppose that $[\![B]\!]$ is finite.
We first treat the $\leftarrow$-direction. Let $j_0, \ldots, j_{s-1} \neq 0$ be the $\sqsubset$-maximal elements j of $\mathbb{N}$ such that $j \notin [\![B]\!]$. Notice that $j_t \in [\![\Box B]\!]$ for all $t < s$. Suppose that we have $i \in [\![B]\!]$ and $t < s$ such that $i \sqsubseteq j_t$. Since M is realistic, we get $i \preceq j_t$, so by preservativity of knowledge, $j_t \in [\![B]\!]$, contradiction. So if $i \in [\![B]\!]$, then $i \not\sqsubseteq j_t$. In particular, we have $\vdash_{\mathsf{HA}} [B] \rightarrow \exists x \neg(h(x) \sqsubseteq j_t)$ for all $t < s$. Using the fact that $\sqsubset$ is a conversely well-founded relation, we can also show: if $i \not\sqsubseteq j_t$ for all $t < s$, then $i \in [\![B]\!]$.
Now we reason inside HA and suppose that $\Box[B]$. Then $\Box(\exists x \neg(h(x) \sqsubseteq j_t))$ also holds. Let k_t satisfy $\mathsf{Prf}_\Box(k_t, \ulcorner \exists x \neg(h(x) \sqsubseteq j_t) \urcorner)$. We distinguish three cases.

1. Suppose $h(k_t) \sqsubset j_t$ for some $t < s$. Then by the definition of h, we have $h(k_t + 1) = j_t$, and $[\Box B]$ follows.
2. Suppose $h(k_t) = j_t$ for some $t < s$. Then $[\Box B]$ again follows.
3. Suppose that $\neg(h(k_t) \sqsubseteq j_t)$ for all $t < s$. Let $k = \max_{t<s} k_t$. If $h(k) = j_t$ for some $t < s$, then $[\Box B]$ again follows. Suppose $h(k) \sqsubset j_t$ for some $t < s$. Since $h(k_t) \preceq h(k) \sqsubset j_t$ and our frame (provably) has the model property, we get $h(k_t) \sqsubset j_t$, which we already excluded. So we have $\neg(h(k) \sqsubseteq j_t)$ for all $t < s$. But then using the $\sqsubseteq$-analogue of (3.3) for $j_0, \ldots, j_{s-1}$, we see that $\bigvee_{j \in U} h(k) = j$, where $U = \{j \in \mathbb{N} \mid j \not\sqsubseteq j_t$ for all $t < s\}$ is a finite set. We see (outside HA) that $U \subseteq [\![B]\!] \subseteq [\![\Box B]\!]$, where the latter inclusion holds

since M is realistic. So (inside HA again), we get $[\Box B]$, as desired.

Now we treat the $\rightarrow$-direction. Consider an $i \in [\![\Box B]\!]$. Then $i > 0$, since $[\![B]\!]$ is finite. So by Lemma 3.2.4, we have

$$\vdash_{\mathsf{HA}} \exists x\,(h(x) = i) \rightarrow \Box(\exists y\,(i \prec h(y))). \tag{3.6}$$

Every nonzero node k of M has a finite $\prec$-rank, which is the greatest n such that there exists a sequence $k = k_0 \prec k_1 \prec \cdots \prec k_n$. Let $a \in \mathbb{N}$ be the $\prec$-rank of i. For $b \in \mathbb{N}$, we define the finite set

$$U_b = \{j \in \mathbb{N} \mid i \prec j, i \not\sqsubset j \text{ and } \mathrm{rank}(j) < b\}.$$

We know (inside HA) that $i \prec h(y)$ implies that $h(y)$ is a standard number. Moreover, such a standard number must have rank smaller than a, so it is either in $[\![B]\!]$ (if $i \sqsubset h(y)$) or in U_a (if $i \not\sqsubset h(y)$). That is, we have

$$\vdash_{\mathsf{HA}} i \prec h(y) \rightarrow \bigvee_{j \in [\![B]\!]} h(y) = j \vee \bigvee_{j \in U_a} h(y) = j.$$

From this, it follows that

$$\vdash_{\mathsf{HA}} \exists y\,(i \prec h(y)) \rightarrow [B] \vee \bigvee_{j \in U_a} \exists y\,(h(y) = j).$$

So (3.6) together with the properties of the $\Box$ implies that

$$\vdash_{\mathsf{HA}} \exists x\,(h(x) = i) \rightarrow \Box\left([B] \vee \bigvee_{j \in U_a} \exists y\,(h(y) = j)\right). \tag{3.7}$$

Suppose that $j \in U_b$ for a certain $b \geq 1$. By Lemma 3.2.5, we know that

$$\vdash_{\mathsf{HA}} \exists x\,(h(x) = i) \wedge \exists y\,(h(y) = j) \rightarrow \triangle(\exists z\,(j \prec h(z)). \tag{3.8}$$

Furthermore, if $j \prec h(z)$, then we know (inside HA) that $h(z)$ is some standard number. Moreover, such a standard number must have lower $\prec$-rank than j, so it is either in $[\![B]\!]$ (if $i \sqsubset h(z)$) or in U_{b-1} (if $i \not\sqsubset h(z)$). That is, we have

$$\vdash_{\mathsf{HA}} j \prec h(z) \rightarrow \bigvee_{k \in [\![B]\!]} h(z) = k \vee \bigvee_{k \in U_{b-1}} h(z) = k.$$

From this, it follows that

$$\vdash_{\mathsf{HA}} \exists z\,(j \prec h(z)) \rightarrow [B] \vee \bigvee_{k \in U_{b-1}} \exists z\,(h(z) = k).$$

So using (3.8) and the properties of $\triangle$, we get

$$\vdash_{\mathsf{HA}} \exists x\,(h(x) = i) \wedge \exists y\,(h(y) = j) \rightarrow \triangle\left([B] \vee \bigvee_{k \in U_{b-1}} \exists z\,(h(z) = k)\right),$$

This holds for all $j \in U_b$, so

$$\vdash_{\mathsf{HA}} \exists x\,(h(x) = i) \wedge \bigvee_{j \in U_b} (\exists y\,(h(y) = j)) \to \triangle \left([B] \vee \bigvee_{j \in U_{b-1}} \exists y\,(h(y) = j) \right).$$

(We changed some bound variables on the right hand side.) Since $[B]$ is equivalent, over HA, to a Σ_1-sentence, we also have

$$\vdash_{\mathsf{HA}} [B] \to \triangle[B] \to \triangle \left([B] \vee \bigvee_{j \in U_{b-1}} \exists y\,(h(y) = j) \right).$$

So we conclude that

$$\vdash_{\mathsf{HA}} \exists x\,(h(x) = i) \wedge \left([B] \vee \bigvee_{j \in U_b} (\exists y\,(h(y) = j)) \right) \to \triangle \left([B] \vee \bigvee_{j \in U_{b-1}} \exists y\,(h(y) = j) \right).$$

Since $\exists x\,(h(x) = i)$ is equivalent, over HA, to a Σ_1-sentence, we have $\vdash_{\mathsf{HA}} \exists x\,(h(x) = i) \to \Box(\exists x\,(h(x) = i))$. Now we see:

$$\begin{aligned}
\vdash_{\mathsf{HA}} \exists x(h(x) = i) \wedge \Box \left([B] \vee \bigvee_{j \in U_b} \exists y\,(h(y) = j) \right) & \\
\to \Box \left(\exists x\,(h(x) = i) \wedge \left([B] \vee \bigvee_{j \in U_b} (\exists y\,(h(y) = j)) \right) \right) & \\
\to \Box\triangle \left([B] \vee \bigvee_{j \in U_{b-1}} \exists y\,(h(y) = j) \right) & \\
\to \Box \left([B] \vee \bigvee_{j \in U_{b-1}} \exists y\,(h(y) = j) \right), &
\end{aligned}$$

where the final step holds since $[B] \vee \bigvee_{j \in U_{b-1}} \exists y\,(h(y) = j)$ is equivalent, over HA, to a Σ_1-sentence. Now we can apply this repeatedly to (3.7) in order to obtain

$$\begin{aligned}
\vdash_{\mathsf{HA}} \exists x\,(h(x) = i) \to \Box \left([B] \vee \bigvee_{j \in U_0} \exists y\,(h(y) = j) \right) & \\
\leftrightarrow \Box([B] \vee \bot) & \\
\leftrightarrow \Box[B], &
\end{aligned}$$

where we used that $U_0 = \emptyset$.
Since this holds for all $i \in [\![\Box B]\!]$, we can conclude that $\vdash_{\mathsf{HA}} [\Box B] \to \Box[B]$, as desired. $\Box$

3.3 The Completeness Theorem

In this section, we formulate and prove our completeness theorem in its abstract form. First, we define provability logics.

Definition 3.3.1. Let T be a theory and let $S(x)$ be a provability predicate for T. If A is an $\mathcal{L}$-sentence, we write $\Box A$ for $S(\ulcorner A \urcorner)$.

(i) A *realization* is a function σ that assigns, to each proposition letters p in $\mathcal{L}_\Box$, an $\mathcal{L}$-sentence $\sigma(p)$. We call σ a Σ_1-realization if $\sigma(p) \in \Sigma_1$ for every proposition letter p.

(ii) Given a realization σ, we define the function $\sigma_\Box \colon \mathcal{L}_\Box \to \mathcal{L}$ by:
   (a) $\sigma_\Box(\bot)$ is $\bot$ and $\sigma_\Box(p)$ is $\sigma(p)$ for every proposition letter p;
   (b) $\sigma_\Box(B \circ C)$ is $\sigma_\Box(B) \circ \sigma_\Box(C)$ for all $B, C \in \mathcal{L}_\Box$ and $\circ \in \{\wedge, \vee, \to\}$.
   (c) $\sigma_\Box(\Box B)$ is $\Box(\sigma_\Box(B))$ for all $B \in \mathcal{L}_\Box$.

   (As in Definition 2.2.4, this is a slight abuse of notation.) Notice that $\sigma_\Box(B)$ is a sentence for every $A \in \mathcal{L}_\Box$.

(iii) The *logic for* $\Box$ is defined as the set of all $A \in \mathcal{L}_\Box$ such that $\vdash_T \sigma_\Box(A)$ for every realization σ. The Σ_1-logic for $\Box$ is the set of all $A \in \mathcal{L}_\Box$ such that $\vdash_T \sigma_\Box(A)$ for every Σ_1-realization σ.

We write σ_T for $\sigma_{\Box_T}$ and σ_T^f for $\sigma_{\Box_T^f}$. The (Σ_1-)logic for $\Box_T$ is called the (Σ_1-)*provability logic of* T, and the (Σ_1-)logic for $\Box_T^f$ is called the *fast* (Σ_1-)*provability logic of* T.

Now, we again adopt the conventions and notation from Section 3.1. All the work from Section 3.2 now leads to the following result.

Theorem 3.3.1. *Define the Σ_1-realization σ by $\sigma(p) = [p]$ for every proposition letter p. Then $\vdash_T \sigma_\Box(A) \leftrightarrow [A]$ for all $A \in \mathcal{L}_\Box$.*

Proof. This follows by induction on the complexity of A using Lemma 3.2.1, Lemma 3.2.2, Lemma 3.2.3 and Lemma 3.2.6. □

The following result tells us what the 'real' behaviour of the Solovay function h is, in the case that T is Σ_1-sound.

Proposition 3.3.2. *Suppose that T is Σ_1-sound. Then $\mathbb{N} \models h(x) = 0$.*

Proof. Since $\prec$ is conversely well-founded, we know that h must have a certain limit $i \in \mathbb{N}$. Suppose that $i > 0$. Then $\vdash_{\mathsf{HA}} \exists x\,(h(x) = i) \to \Box(\exists y\,(i \preceq h(y)))$ by Lemma 3.2.4, so since HA is sound, we get $\mathbb{N} \models \exists x\,(h(x) = i) \to \Box(\exists y\,(i \prec h(y)))$. By assumption, $\mathbb{N} \models \exists x\,(h(x) = i)$, so $\mathbb{N} \models \Box(\exists y\,(i \preceq h(y)))$. By requirement (ii) for a good pair, we have $\vdash_T \exists y\,(i \prec h(y))$. Since $\exists y\,(i \prec h(y))$ is a Σ_1-sentence and T is Σ_1-sound, we get $\mathbb{N} \models \exists y\,(i \prec h(y))$. However, this is impossible as i is supposed to be the limit of h. So $i = 0$, and the result follows. □

Now we can finally formulate and prove our main result.

Theorem 3.3.3. *Let $T \supseteq \mathsf{HA}$ be a Σ_1-sound theory and suppose we have a good pair (S, R) for T. If A is an $\mathcal{L}$-sentence, we write $\Box A$ for $S(\ulcorner A \urcorner)$.*

(*i*) *The Σ_1-logic for $\Box$ is equal to the set of theorems of* iGLC.

(*ii*) *If $\vdash_T \mathrm{CP}_\Box$, then the logic for $\Box$ is equal to the set of theorems of* iGLC.

Proof. (i) Since $S(x)$ is a provability predicate for T, we see that the Σ_1-logic for $\Box$ contains iGLC and is closed under $\to$E and Nec. So the logic Σ_1-logic for $\Box$ contains all theorems of iGLC.

Now suppose that we have $A \in \mathcal{L}_\Box$ such that iGLC $\nvdash A$. Then by Theorem 2.4.4, there exists a finite, irreflexive, realistic model $M_0 = \langle W_0, \preceq_0, \sqsubset_0, V_0 \rangle$ in which A is not valid. We label the nodes of M_0 as $W_0 = \{1, \ldots, r\}$ in such a way that $M_0, r \nVdash A$. By shrinking W_0 to $\{i \in W_0 \mid r \preceq_0 i\}$ if necessary, we may assume without loss of generality that r is the $\preceq_0$-least element of W_0.

Now define the model M, the Solovay function h, and the Σ_1-sentences $[B]$ for $B \in \mathcal{L}_\Box$ as above. It is easy to show that $M_0, i \Vdash B$ iff $M, i \Vdash B$ for all $B \in \mathcal{L}_\Box$ and all i with $1 \leq i \leq r$. So we have $M, r \nVdash A$, that is, $r \notin [\![A]\!]$. Now define the Σ_1-realization σ by $\sigma(p) = [p]$ for every proposition letter p. By Theorem 3.3.1, we have $\vdash_T \sigma_\Box(B) \leftrightarrow [B]$ for all $B \in \mathcal{L}_\Box$.

Now suppose for the sake of contradiction that $\vdash_T \sigma_\Box(A)$. Then we also get $\vdash_T [A]$. Since $[A]$ is (equivalent to) a Σ_1-sentence and T is Σ_1-sound, we see that $\mathbb{N} \models [A]$. By Proposition 3.3.2, we also know that $\mathbb{N} \models h(x) = 0$. This implies that $0 \in [\![A]\!]$. However, we also have $0 \preceq r$ and $r \notin [\![A]\!]$, which yields a contradiction. We conclude that A is not in the Σ_1-logic for $\Box$, as desired.

(ii) Since $S(x)$ is a provability predicate for T, we see that the logic for $\Box$ contains iGL and is closed under $\to$E and Nec. Since $\vdash_T \mathrm{CP}_\Box$, we also have that $A \to \Box A$ is in the logic for $\Box$, for every $A \in \mathcal{L}_\Box$. So the logic for $\Box$ contains all theorems of iGLC. Conversely, the logic for $\Box$ is contained in the Σ_1-logic for $\Box$, which is contained in the set of theorems of iGLC. $\Box$

Chapter 4

Applications of the Completeness Theorem

In the previous chapter, we proved a completeness theorem in a very abstract form. In this chapter, we provide several applications of this theorem. In particular, we will determine the fast provability logics of the theories U^*, for Σ_1-sound theories $U \supseteq \mathsf{HA}$, and we will determine the fast and ordinary Σ_1-provability logics of HA. First of all, we lay some further groundwork in Section 4.1. Then, in Section 4.2, we determine the fast provability logics mentioned above. Finally, in Section 4.3, we determine the Σ_1-provability logic of HA.

4.1 The Sets NNIL and TNNIL

In the sequel, $\mathcal{L}_p$ is the language of propositional logic, and for $A \in \mathcal{L}_p$, we write '$\vdash_{\mathsf{IPC}} A$' to indicate that A is provable in intuitionistic propositional logic. We notice that, if σ is a substitution, $A \in \mathcal{L}_p$ and $S(x)$ is a provability predicate for a certain theory T, then $\sigma_\Box(A)$ does not depend on the provability predicate S. So we will just write $\sigma(A)$ instead of $\sigma_\Box(A)$. We will also drop the brackets in expressions of the form $\sigma(A)$ and $\sigma_T(A)$.

Like the authors of [1], we introduce the set of NNIL-formulae.

Definition 4.1.1. The set $\mathsf{NNIL} \subseteq \mathcal{L}_p$ ('no nested implications on the left') is defined recursively, as follows:

(i) all proposition letters are in NNIL, as is $\bot$;

(ii) if $A, B \in \mathsf{NNIL}$, then $A \wedge B, A \vee B \in \mathsf{NNIL}$;

(iii) if $A \in \mathcal{L}_p$ contains no implications and $B \in \mathsf{NNIL}$, then $A \to B \in \mathsf{NNIL}$.

That is, a NNIL-formula is a propositional formulae in which no implication occurs in the antcedent of another implication. In the paper [11], we find the following result, that we will not prove here.

Theorem 4.1.1. *There exists a computable function* $(\cdot)^* \colon \mathcal{L}_p \to \mathsf{NNIL}$*, called the* NNIL-algorithm*, such that for every* $A \in \mathcal{L}_p$*, the following hold:*

(i) $\vdash_{\mathsf{IPC}} A^* \to A$*;*

(ii) if $B \in \mathsf{NNIL}$ *and* $\vdash_{\mathsf{IPC}} B \to A$*, then* $\vdash_{\mathsf{IPC}} B \to A^*$*;*

(iii) if σ *is a* Σ_1*-realization, then* $\vdash_{\mathsf{HA}} \Box_{\mathsf{HA}}(\sigma A) \leftrightarrow \Box_{\mathsf{HA}}(\sigma A^*)$*.*

Remark 4.1.1. Consider the preorder $(\mathcal{L}_p, \leq)$, there $\leq$ is defined by: $A \leq B$ if and only if $\vdash_{\mathsf{IPC}} A \to B$, for $A, B \in \mathcal{L}_p$. Consider also the subpreorder $(\mathsf{NNIL}, \leq)$. Then items (i) and (ii) above say that the NNIL-algorithm is left adjoint to the inclusion $\mathsf{NNIL} \to \mathcal{L}_p$. $\Diamond$

We can get an analogue of (iii) for fast provability.

Corollary 4.1.2. *Let* $A \in \mathcal{L}_p$ *and let* σ *be a* Σ_1*-realization. Then*

$$\vdash_{\mathsf{HA}} \Box^f_{\mathsf{HA}}(\sigma A) \leftrightarrow \Box^f_{\mathsf{HA}}(\sigma A^*).$$

Proof. Since $\mathsf{Bew}^f_{\mathsf{HA}}$ is a provability predicate for HA, we can derive from Theorem 4.1.1(iii) that

$$\vdash_{\mathsf{HA}} \Box^f_{\mathsf{HA}}(\sigma A) \leftrightarrow \Box^f_{\mathsf{HA}}\Box_{\mathsf{HA}}(\sigma A) \leftrightarrow \Box^f_{\mathsf{HA}}\Box_{\mathsf{HA}}(\sigma A^*) \leftrightarrow \Box^f_{\mathsf{HA}}(\sigma A^*),$$

where we also used Lemma 2.3.2(vii). $\Box$

The NNIL-algorithm behaves nicely with respect to the theories U^T and Σ_1-realizations.

Proposition 4.1.3. *Suppose that* U *and* T *are theories such that* $\mathsf{HA} \leq U \leq U^T$*,* $\mathsf{HA} \subseteq T$ *and* $\vdash_{\mathsf{HA}} \mathsf{Bew}_U(x) \to \Box_U \mathsf{Bew}_T(x)$*. Then for all* Σ_1*-realizations* σ *and* $C \in \mathsf{NNIL}$*, we have*

$$\vdash_{\mathsf{HA}} \Box_U(\sigma C) \leftrightarrow \Box_{U^T}(\sigma C). \tag{4.1}$$

Proof. The '$\to$'-direction holds since $U \leq U^T$. For the converse, we notice that, since σ is a Σ_1-realization and $C \in \mathsf{NNIL}$, we have that σC is equivalent, over HA, to a sentence in $\mathcal{A}$. Since $\mathsf{HA} \subseteq T$, we have $\vdash_{\mathsf{HA}} (\sigma C)^T \to \sigma C$, by Corollary 2.2.5 and Lemma 2.2.7(ii). Since $\mathsf{HA} \subseteq U$, we get $\vdash_{\mathsf{HA}} \Box_U(\sigma C)^T \to \Box_U(\sigma C)$. Finally, we notice that the conditions of Corollary 2.2.6 hold, so we get $\vdash_{\mathsf{HA}} \Box_{U^T}(\sigma C) \to \Box_U(\sigma C)^T \to \Box_U(\sigma C)$, as desired. $\Box$

Corollary 4.1.4. *Suppose that* U *and* T *are theories such that* $\mathsf{HA} \leq U \leq T$ *and* $U \leq U^T$*. Then for all* Σ_1*-realizations* σ *and* $C \in \mathsf{NNIL}$*, we have*

$$\vdash_{\mathsf{HA}} \Box^f_U(\sigma C) \leftrightarrow \Box^f_{U^T}(\sigma C). \tag{4.2}$$

Proof. We notice that the conditions from Proposition 4.1.3 apply, so we have

$$\vdash_{\mathsf{HA}} \Box_U(\sigma C) \leftrightarrow \Box_{U^T}(\sigma C) \leftrightarrow (\Box_{U^T}(\sigma C))^T,$$

where the latter equivalence holds by Lemma 2.2.7(i) and the fact that $\mathsf{HA} \subseteq T$. Since $\mathsf{HA} \subseteq U$, we also get $\vdash_U \Box_U(\sigma C) \leftrightarrow (\Box_{U^T}(\sigma C))^T$. We know that Bew^f_U is a provability predicate for U, so we also get $\vdash_{\mathsf{HA}} \Box^f_U \Box_U(\sigma C) \leftrightarrow \Box^f_U(\Box_{U^T}(\sigma C))^T$. Using Lemma 2.3.2(vii) and Lemma 2.3.3, we find

$$\begin{aligned}\vdash_{\mathsf{HA}} \Box^f_U(\sigma C) &\leftrightarrow \Box^f_U \Box_U(\sigma C)\\ &\leftrightarrow \Box^f_U(\Box_{U^T}(\sigma C))^T\\ &\leftrightarrow \Box^f_{U^T}\Box_{U^T}(\sigma C)\\ &\leftrightarrow \Box^f_{U^T}(\sigma C),\end{aligned}$$

as desired. □

Following [1], we now extend the notion of 'no nested implication on the left' to modal formulae.

Definition 4.1.2. The set TNNIL $\subseteq \mathcal{L}_\Box$ ('thoroughly no nested implications on the left') is defined by recursion, as follows:

(i) all proposition letters are in TNNIL;

(ii) if $A, B \in$ TNNIL, then $A \wedge B, A \vee B, \Box A \in$ TNNIL;

(iii) if $A, B \in$ TNNIL and A contains no implications outside a box, then $A \to B \in$ TNNIL.

We notice that every $A \in \mathcal{L}_\Box$ can, in a unique way, be written as $C(\vec{p}, \Box B_1, \ldots, \Box B_k)$, for certain $C(\vec{p}, q_1, \ldots, q_k) \in \mathcal{L}_p$ and *distinct* $B_1, \ldots, B_k \in \mathcal{L}_\Box$. It is easy to show that, with this notation, we have $A \in$ TNNIL if and only if $C \in$ NNIL and $B_i \in$ TNNIL for $1 \le i \le k$. Now we define an operation on modal formulae as in [1].

Definition 4.1.3. The TNNIL-*algorithm* $(\cdot)^+ \colon \mathcal{L}_\Box \to$ TNNIL is defined by recursion, as follows. For $A \in \mathcal{L}_\Box$, write $A = C(\vec{p}, \Box B_1, \ldots, \Box B_k)$, where $C(\vec{p}, q_1, \ldots, q_k) \in \mathcal{L}_p$ and $B_1, \ldots, B_k \in \mathcal{L}_\Box$ are distinct. Then

$$A^+ := C^*(\vec{p}, \Box B_1^+, \ldots, \Box B_k^+).$$

Notice that, since all the B_i have lower complexity than A, the operation $(\cdot)^+$ is well-defined.

We can use our results about NNIL and $(\cdot)^*$ to obtain the following lemmata about TNNIL and $(\cdot)^+$. We notice that Lemma 4.1.5(i) also occurs in [1] as Corollary 4.7.1.

Lemma 4.1.5. *Let $A \in \mathcal{L}_p$ and let σ be a Σ_1-realization. Then the following hold:*

(i) $\vdash_{\mathsf{HA}} \Box_{\mathsf{HA}}(\sigma_{\mathsf{HA}} A) \leftrightarrow \Box_{\mathsf{HA}}(\sigma_{\mathsf{HA}} A^+)$;

(ii) $\vdash_{\mathsf{HA}} \Box^f_{\mathsf{HA}}(\sigma^f_{\mathsf{HA}} A) \leftrightarrow \Box^f_{\mathsf{HA}}(\sigma^f_{\mathsf{HA}} A^+)$.

Proof. (i) We proceed by strong induction on the boxdepth of A. As above, we write A as $C(\vec{p}, \Box B_1, \ldots, \Box B_k)$, where $C(\vec{p}, q_1, \ldots, q_k) \in \mathcal{L}_p$ and $B_1, \ldots, B_k \in \mathcal{L}_\Box$ are distinct. Then all the B_i have smaller boxdepth than A, so we assume by induction hypothesis that

$$\vdash_{\mathsf{HA}} \Box_{\mathsf{HA}}(\sigma_{\mathsf{HA}} B_i) \leftrightarrow \Box_{\mathsf{HA}}(\sigma_{\mathsf{HA}} B_i^+) \quad \text{for} \quad 1 \le i \le k. \tag{4.3}$$

If $\vec{p} = p_1, \ldots, p_l$, then we write $\sigma\vec{p}$ as a shorthand for $\sigma(p_1), \ldots, \sigma(p_l)$. Now we take a Σ_1-realization τ such that $\tau\vec{p} = \sigma\vec{p}$ and $\tau(q_i) = \Box_{\mathsf{HA}}(\sigma_{\mathsf{HA}} B_i)$ for $1 \leq i \leq k$. Now we observe that

$$\begin{aligned} \sigma_{\mathsf{HA}} A &= C(\sigma\vec{p}, \Box_{\mathsf{HA}}(\sigma_{\mathsf{HA}} B_1), \ldots, \Box_{\mathsf{HA}}(\sigma_{\mathsf{HA}} B_k)) = \tau C, \\ \sigma_{\mathsf{HA}} A^+ &= C^*(\sigma\vec{p}, \Box_{\mathsf{HA}}(\sigma_{\mathsf{HA}} B_1^+), \ldots, \Box_{\mathsf{HA}}(\sigma_{\mathsf{HA}} B_k^+)) \quad \text{and} \\ \tau C^* &= C^*(\sigma\vec{p}, \Box_{\mathsf{HA}}(\sigma_{\mathsf{HA}} B_1), \ldots, \Box_{\mathsf{HA}}(\sigma_{\mathsf{HA}} B_k)). \end{aligned}$$

So (4.3) gives $\vdash_{\mathsf{HA}} \sigma_{\mathsf{HA}} A^+ \leftrightarrow \tau C^*$. Since $\mathsf{Bew}_{\mathsf{HA}}$ is a provability predicate for HA, we conclude that

$$\vdash_{\mathsf{HA}} \Box_{\mathsf{HA}}(\sigma_{\mathsf{HA}} A) \leftrightarrow \Box_{\mathsf{HA}}(\tau C) \leftrightarrow \Box_{\mathsf{HA}}(\tau C^*) \leftrightarrow \Box_{\mathsf{HA}}(\sigma_{\mathsf{HA}} A^+),$$

where we used Theorem 4.1.1(iii). This completes the induction.
(ii) The proof is completely analogous, but with an appeal to Corollary 4.1.2 instead of Theorem 4.1.1(iii). □

Lemma 4.1.6. *Let U and T be theories.*

(i) If (4.1) holds for all $C \in \mathsf{NNIL}$ and Σ_1-realizations σ, then for every $A \in \mathsf{TNNIL}$, we have:

$$\vdash_{\mathsf{HA}} \Box_U(\sigma_U A) \leftrightarrow \Box_{U^T}(\sigma_{U^T} A)$$

(ii) If (4.2) holds for all $C \in \mathsf{NNIL}$ and Σ_1-realizations σ, then for every $A \in \mathsf{TNNIL}$, we have:

$$\vdash_{\mathsf{HA}} \Box_U^f(\sigma_U A) \leftrightarrow \Box_{U^T}^f(\sigma_{U^T} A)$$

Proof. (i) The proof is very similar to the proof of Lemma 4.1.5.
We proceed by strong induction on the boxdepth of A. Write $A = C(\vec{p}, \Box B_1, \ldots, \Box B_k)$, where $C(\vec{p}, q_1, \ldots, q_k) \in \mathsf{NNIL}$ and $B_1, \ldots, B_k \in \mathsf{TNNIL}$ are distinct. Then all the B_i have smaller boxdepth than A, so we assume by induction hypothesis that

$$\vdash_{\mathsf{HA}} \Box_U(\sigma_U B_i) \leftrightarrow \Box_{U^T}(\sigma_{U^T} B_i) \quad \text{for} \quad 1 \leq i \leq k. \tag{4.4}$$

Now we take a Σ_1-realization τ such that $\tau\vec{p} = \sigma\vec{p}$ and $\tau(q_i) = \Box_U(\sigma_U B_i)$ for $1 \leq i \leq k$. Now we observe that

$$\begin{aligned} \sigma_U A &= C(\sigma\vec{p}, \Box_U(\sigma_U B_1), \ldots, \Box_U(\sigma_U B_k)) = \tau C \quad \text{and} \\ \sigma_{U^T} A &= C(\sigma\vec{p}, \Box_{U^T}(\sigma_{U^T} B_1), \ldots, \Box_{U^T}(\sigma_{U^T} B_k)). \end{aligned}$$

So (4.4) gives $\vdash_{\mathsf{HA}} \sigma_{U^T} A \leftrightarrow \tau C$. Since $\mathsf{HA} \subseteq U^T$, we also get $\vdash_{U^T} \sigma_{U^T} A \leftrightarrow \tau C$. We also know that Bew_{U^T} is a provability predicate for U^T, so we also find $\vdash_{\mathsf{HA}} \Box_{U^T}(\sigma_{U^T} A) \leftrightarrow \Box_{U^T}(\tau C)$. Using (4.1), we get,

$$\vdash_{\mathsf{HA}} \Box_U(\sigma_U A) \leftrightarrow \Box_U(\tau C) \leftrightarrow \Box_{U^T}(\tau C) \leftrightarrow \Box_{U^T}(\sigma_{U^T} A),$$

which completes the induction.
(ii) The proof is again completely analogous, but of course with an appeal to (4.2) instead of (4.1). □

4.2 Some Fast (Σ_1-)Provability Logics

Let $U \supseteq \mathsf{HA}$ be a Σ_1-sound theory. We consider the theory $U^* \supseteq \mathsf{HA}$. First of all, we prove the following.

Lemma 4.2.1. *The pair* $\left(\mathsf{Bew}^f_{U^*}(x), \mathsf{Bew}_{U^*}(x)\right)$ *is good for* U^*.

Proof. By Lemma 2.3.2(v), we know that $\mathsf{Bew}^f_{U^*}$ is a provability predicate for U^*, and we also know that Bew_{U^*} is a provability predicate for U^*.
By Lemma 2.2.7(iii), the theories U and U^* prove the same Σ_1-formulae, so U^* is Σ_1-sound as well. By Lemma 2.3.2(vi), we see that $\mathbb{N} \models \Box^f_{U^*} A$ implies that $\mathbb{N} \models \Box_{U^*} A$, which implies $\vdash_{U^*} A$, for all $\mathcal{L}$-sentences A.
By Corollary 2.2.10, we have $\vdash_{U^*} \mathrm{CP}_{U^*}$.
The final requirement for a good pair follows from Lemma 2.3.2(vii). □

Theorem 4.2.2. *Let* $U \supseteq \mathsf{HA}$ *be a* Σ_1*-sound theory. Then the fast* (Σ_1-)*provability logic of* U^* *is equal to the set of theorems of* iGLC.

Proof. By Corollary 2.2.10, we have $\vdash_{U^*} A \to \Box_{U^*} A \to \Box^f_{U^*} A$ for all $\mathcal{L}$-sentences A. Now both statements follow from Theorem 3.3.3 and Lemma 4.2.1. □

Remark 4.2.1. Presently, the ordinary (Σ_1-)provability logic of HA^* is unknown. We conjecture that it is equal to iGLC as well. For PA^*, more is known. Since PA is a classical theory, we have $\vdash_{\mathsf{PA}} B \vee (B \to A)$ for all $\mathcal{L}$-formulae A and B. This means that we also have

$$\vdash_{\mathsf{PA}} \Box_{\mathsf{PA}} A^{\mathsf{PA}} \to (B^{\mathsf{PA}} \vee ((B^{\mathsf{PA}} \to A^{\mathsf{PA}}) \wedge \Box_{\mathsf{PA}}(B^{\mathsf{PA}} \to A^{\mathsf{PA}})))$$

for all $\mathcal{L}$-formulae A and B. This, in turn, implies that

$$\vdash_{\mathsf{PA}^*} \Box_{\mathsf{PA}^*} A \to (B \vee (B \to A)),$$

for all $\mathcal{L}$-formulae A and B. So the (Σ_1-)provability logic of PA^* contains at least the theorems of iGLC extended with the axiom scheme $\Box A \to (B \vee (B \to A))$. This scheme is called the *propositional trace principle*, or PTP for short. The theory $\mathsf{iGLC} + \mathrm{PTP}$ for IML is sound and complete with respect to finite frames $\langle W, \preceq, \sqsubset \rangle$, such that $w \sqsubset v$ iff $w \prec v$ for all $w, v \in W$. A. Visser used a proof in the style of Chapter 3 to show that the (Σ_1-)provability logic of PA^* contains *exactly* the theorems of $\mathsf{iGLC} + \mathrm{PTP}$. Since in this case, the model relation $\sqsubset$ can be defined in terms of the intuitionistic relation $\preceq$, the definition of the Solovay function and the induction step for $\Box$ (our Lemma 3.2.6) are somewhat easier. Since $\mathsf{iGLC} + \mathrm{PTP}$ is a proper extension of iGLC (i.e. has more theorems), we have an example of a theory for which the fast and ordinary provability logics do not coincide. ◊

We now turn our attention to determining the fast Σ_1-provability logic of HA.

Theorem 4.2.3. *Let* $A \in \mathcal{L}_\Box$. *Then* A *is in the fast* Σ_1*-provability logic of* HA *if and only if* $\mathsf{iGLC} \vdash A^+$.

Remark 4.2.2. This result gives an 'indirect' characterization of the fast Σ_1-provability logic of HA, since we first have to apply the TNNIL-algorithm, and then see whether the result is provable in iGLC. But we can already see that the fast Σ_1-provability logic of HA is decidable, since iGLC is decidable (this follows from the proof of Theorem 2.4.4). In the paper [1], the authors give a direct characterization of the set $\{A \in \mathcal{L}_\Box \mid \vdash_{\mathsf{iGLC}} A^+\}$, by providing an axiomatization for it. $\Diamond$

Proof of Theorem 4.2.3. Since $\mathsf{HA} \leq \mathsf{HA}$ and $\mathsf{HA} \leq \mathsf{HA}^*$, the conditions of Corollary 4.1.4 apply for $U \equiv T \equiv \mathsf{HA}$, so (4.2) holds for $U \equiv T \equiv \mathsf{HA}$. Now let $A \in \mathcal{L}_\Box$ and let σ be a Σ_1-realization. Using Lemma 4.1.5(ii) and Lemma 4.1.6(ii), we find that

$$\vdash_{\mathsf{HA}} \Box^f_{\mathsf{HA}}(\sigma^f_{\mathsf{HA}} A) \leftrightarrow \Box^f_{\mathsf{HA}}(\sigma^f_{\mathsf{HA}} A^+) \leftrightarrow \Box^f_{\mathsf{HA}^*}(\sigma^f_{\mathsf{HA}^*} A^+).$$

Since HA is sound, we see that $\mathbb{N} \models \Box^f_{\mathsf{HA}}(\sigma_{\mathsf{HA}} A)$ if and only if $\mathbb{N} \models \Box^f_{\mathsf{HA}^*}(\sigma_{\mathsf{HA}^*} A^+)$. Using Lemma 2.2.7(iii), we see that HA^* is Σ_1-sound. Using Lemma 2.3.2(vi), we can now see that

$$\vdash_{\mathsf{HA}} \sigma_{\mathsf{HA}} A \quad \text{iff} \quad \mathbb{N} \models \Box^f_{\mathsf{HA}}(\sigma_{\mathsf{HA}} A) \quad \text{iff} \quad \mathbb{N} \models \Box^f_{\mathsf{HA}^*}(\sigma_{\mathsf{HA}^*} A^+) \quad \text{iff} \quad \vdash_{\mathsf{HA}^*} \sigma_{\mathsf{HA}^*} A^+.$$

This means that A is in the fast Σ_1-provability logic of HA if and only if A^+ is in the fast Σ_1-provability logic of HA^*. By Theorem 4.2.2, the latter holds if and only if $\vdash_{\mathsf{iGLC}} A^+$. $\Box$

4.3 The Σ_1-Provability Logic of HA

In this final section, we determine the (ordinary) Σ_1-provability logic of HA. This is also the main result of the paper [1], but the authors arrive at it using different methods.

Recall the theory slow Heyting Arithmetic sHA, that satisfies $\mathsf{sHA} = \mathsf{HA}$ and $\mathsf{sHA} \leq \mathsf{HA}$, but not $\mathsf{HA} \leq \mathsf{sHA}$. We consider the theory $\widehat{\mathsf{HA}} := \mathsf{HA}^{\mathsf{sHA}} \supseteq \mathsf{HA}$. By Proposition 2.3.4, we know that $\vdash_{\mathsf{HA}} \Box_{\widehat{\mathsf{HA}}} A \leftrightarrow \Box_{\mathsf{HA}} A^{\mathsf{sHA}}$ for all $\mathcal{L}$-formulae A.

Our first goal is to show that the (Σ_1-)provability logic of this theory is equal to the set of theorems of iGLC. In order to do this, we need to find a good pair for $\widehat{\mathsf{HA}}$. In the previous section, the role of $S(x)$ was fulfilled by *fast* provability. In this section, we put ordinary provability here. We define the Σ_1-formula $R(x)$ as $\mathsf{Bew}_{\mathsf{sHA}}(x^{\mathsf{sHA}})$. As usual, for an $\mathcal{L}$-sentence A, we write $\triangle A$ for $R(\ulcorner A \urcorner)$. Then we see that $\triangle A$ is equivalent, over EA, to $\Box_{\mathsf{sHA}} A^{\mathsf{sHA}}$.

Lemma 4.3.1. *The pair* $\big(\mathsf{Bew}_{\widehat{\mathsf{HA}}}(x), R(x)\big)$ *is good for* $\widehat{\mathsf{HA}}$.

Proof. We already know that $\mathsf{Bew}_{\widehat{\mathsf{HA}}}$ is a provability predicate for $\widehat{\mathsf{HA}}$. We show that $R(x)$ is a provability predicate for $\widehat{\mathsf{HA}}$ as well. First of all, let A and B be $\mathcal{L}$-sentences and suppose that $\vdash_{\widehat{\mathsf{HA}}} A$. Then by Theorem 2.2.3, we see that $\vdash_{\mathsf{HA}} A^{\mathsf{sHA}}$, since $\mathsf{HA} \subseteq \mathsf{HA} \subseteq \mathsf{sHA}$ holds. But $\mathsf{HA} = \mathsf{sHA}$, so $\vdash_{\mathsf{sHA}} A^{\mathsf{sHA}}$ as well, which yields $\mathbb{N} \models \Box_{\mathsf{sHA}} A^{\mathsf{sHA}}$. Moreover, since $\vdash_{\mathsf{EA}} (A \to B)^{\mathsf{sHA}} \to (A^{\mathsf{sHA}} \to B^{\mathsf{sHA}})$, we find

$$\vdash_{\mathsf{HA}} \Box_{\mathsf{sHA}}(A \to B)^{\mathsf{sHA}} \to \Box_{\mathsf{sHA}}(A^{\mathsf{sHA}} \to B^{\mathsf{sHA}}) \to (\Box_{\mathsf{sHA}} A^{\mathsf{sHA}} \to \Box_{\mathsf{sHA}} B^{\mathsf{sHA}}).$$

Finally, if S is a Σ_1-sentence, then by Lemma 2.2.7(i), we get $\vdash_{\mathsf{HA}} S \leftrightarrow S^{\mathsf{sHA}}$. Since $\mathsf{Bew}_{\mathsf{sHA}}$ is a provability predicate for sHA, hence also for HA, we get

$$\vdash_{\mathsf{HA}} S \to \Box_{\mathsf{sHA}} S \to \Box_{\mathsf{sHA}} S^{\mathsf{sHA}},$$

as desired.
Next, let A be an $\mathcal{L}$-sentence. We know from Proposition 2.1.4(i) that $\mathbb{N} \models \Box_{\widehat{\mathsf{HA}}} A$ implies $\vdash_{\widehat{\mathsf{HA}}} A$.
Moreover, by Lemma 2.2.9 with $U \equiv \mathsf{HA}$ and $T \equiv \mathsf{sHA}$, we have $\vdash_{\widehat{\mathsf{HA}}} A \to \Box_{\mathsf{sHA}} A^{\mathsf{sHA}}$.
Finally, let S be a Σ_1-sentence. We recall that $\vdash_{\mathsf{HA}} S \leftrightarrow S^{\mathsf{sHA}}$. We also have that $\vdash_{\mathsf{HA}} \Box_{\mathsf{sHA}} S^{\mathsf{sHA}} \leftrightarrow (\Box_{\mathsf{sHA}} S^{\mathsf{sHA}})^{\mathsf{sHA}}$. Now we use Proposition 2.3.5(i) to find that:

$$\begin{aligned} \vdash_{\mathsf{HA}} \Box_{\widehat{\mathsf{HA}}} \Box_{\mathsf{sHA}} S^{\mathsf{sHA}} &\leftrightarrow \Box_{\mathsf{HA}} (\Box_{\mathsf{sHA}} S^{\mathsf{sHA}})^{\mathsf{sHA}} \\ &\leftrightarrow \Box_{\mathsf{HA}} \Box_{\mathsf{sHA}} S^{\mathsf{sHA}} \\ &\leftrightarrow \Box_{\mathsf{HA}} \Box_{\mathsf{sHA}} S \\ &\to \Box_{\mathsf{HA}} S \\ &\leftrightarrow \Box_{\mathsf{HA}} S^{\mathsf{sHA}} \\ &\leftrightarrow \Box_{\widehat{\mathsf{HA}}} S, \end{aligned}$$

which finishes the proof. □

Now that we have our good pair, we can prove the following.

Lemma 4.3.2. *The* (Σ_1-)*provability logic of* $\widehat{\mathsf{HA}}$ *is exactly the set of theorems of* iGLC.

Proof. Since $\mathsf{sHA} \leq \mathsf{HA}$, we see that

$$\vdash_{\widehat{\mathsf{HA}}} A \to \Box_{\mathsf{sHA}} A^{\mathsf{sHA}} \to \Box_{\mathsf{HA}} A^{\mathsf{sHA}} \to \Box_{\widehat{\mathsf{HA}}} A$$

for every $\mathcal{L}$-sentence A. This means that $\vdash_{\widehat{\mathsf{HA}}} \mathrm{CP}_{\widehat{\mathsf{HA}}}$, so both statements follow from Theorem 3.3.3 and Lemma 4.3.1. □

Since the theory $\widehat{\mathsf{HA}}$ is rather *ad hoc*, this result is in itself not very interesting. But we can use it to obtain the theorem we were after.

Theorem 4.3.3. *Let* $A \in \mathcal{L}_\Box$. *Then* A *is in the* Σ_1-*provability logic of* HA *if and only if* $\mathsf{iGLC} \vdash A^+$.

Proof. We have $\mathsf{HA} \leq \mathsf{HA} \leq \widehat{\mathsf{HA}}$ and $\mathsf{HA} \subseteq \mathsf{sHA}$. Moreover, by Proposition 2.3.4, we have $\vdash_{\mathsf{HA}} \mathsf{Bew}_{\mathsf{HA}}(x) \to \Box_{\mathsf{HA}} \mathsf{Bew}_{\mathsf{sHA}}(x)$. That is, the conditions of Proposition 4.1.3 apply, so equation (4.1) holds for $U \equiv \mathsf{HA}$ and $T \equiv \mathsf{sHA}$. Now let σ be a Σ_1-realization. Using Lemma 4.1.5(i) and Lemma 4.1.6(i), we find that

$$\vdash_{\mathsf{HA}} \Box_{\mathsf{HA}}(\sigma_{\mathsf{HA}} A) \leftrightarrow \Box_{\mathsf{HA}}(\sigma_{\mathsf{HA}} A^+) \leftrightarrow \Box_{\widehat{\mathsf{HA}}}(\sigma_{\widehat{\mathsf{HA}}} A^+).$$

Since HA is sound, we get

$$\vdash_{\mathsf{HA}} \sigma_{\mathsf{HA}} A \quad \text{iff} \quad \mathbb{N} \models \Box_{\mathsf{HA}} \sigma_{\mathsf{HA}} A \quad \text{iff} \quad \mathbb{N} \models \Box_{\widehat{\mathsf{HA}}}(\sigma_{\widehat{\mathsf{HA}}} A^+) \quad \text{iff} \quad \vdash_{\widehat{\mathsf{HA}}} \sigma_{\widehat{\mathsf{HA}}} A^+.$$

This means that A is in the Σ_1-provability logic of HA if and only if A^+ is in the Σ_1-provability logic of $\widehat{\mathsf{HA}}$. By Lemma 4.3.2, the latter holds if and only if $\vdash_{\mathsf{iGLC}} A^+$. □

Chapter 5

Conclusion

In this thesis, our goal was to give a Solovay-style embedding of frames equipped with both an intuitionistic relation $\preceq$ and a modal relation $\sqsubset$. In order to approach this task, we considered theories that prove their own completeness principle. This project has led to the following results and insights.

(i) We were able to give a Solovay-style embedding of finite, irreflexive, realistic frames for IML, in the presence of the completeness principle and the principle $\Box\triangle S \to \Box S$ for $S \in \Sigma_1$.

(ii) We reproved the result from [1] that the Σ_1-provability logic of Heyting Arithmetic is equal to the set $\{A \in \mathcal{L}_\Box \mid \vdash_{\mathsf{iGLC}} A^+\}$.

(iii) We showed that the fast Σ_1-provability logic of HA is also equal to this set.

(iv) We showed that for any Σ_1-sound theory $U \supseteq \mathsf{HA}$, the fast (Σ_1-)provability logic of U^* is equal to the set of theorems of iGLC.

(v) We found an intuitionistic theory of arithmetic other than PA^*, namely the theory $\widehat{\mathsf{HA}}$, for which we were able to determine the provability logic.

(vi) We discovered that for the theory PA^*, the fast provability logic and the ordinary provability logic do not coincide.

Of course, a variety of questions remains open. We mention a few of them, ordered from small to large.

(i) Does the result $\vdash_{\mathsf{HA}} \Box_{\mathsf{HA}}\Box_{\mathsf{sHA}}S \to \Box_{\mathsf{HA}}S$ for $S \in \Sigma_1$ (Proposition 2.3.5(i)) also hold for sentences that are not Σ_1, as it does in the classical case? Or do we even have this principle uniformly over formulae, i.e. $\vdash_{\mathsf{HA}} \mathrm{Form}(x) \wedge \Box_{\mathsf{HA}}\mathsf{Bew}_{\mathsf{sHA}}(x) \to \mathsf{Bew}_{\mathsf{HA}}(x)$?

(ii) What is the ordinary provability logic of HA^* and of other theories U^*, where $U \supseteq \mathsf{HA}$ is Σ_1-sound? We conjecture that the former is equal to the set of theorems of iGLC. We would like to make two remarks about possible approaches to proving this.

   (a) If $T \supseteq \mathsf{HA}$ is a consistent theory, then a good pair (S, R) for T can never satisfy $\vdash_T S(x) \leftrightarrow R(x)$. Therefore, Theorem 3.3.3 can only provide us with the logic for $\Box$, and not with the logic for $\triangle$. Since HA^* is specifically designed to prove its own completeness principle $\mathrm{CT}_{\mathsf{HA}^*}$, i.e. $\mathrm{CT}_{\Box_{\mathsf{HA}^*}}$, Theorem 3.3.3 seems to be useless for determining the provability logic of HA^*, i.e. the logic for $\Box_{\mathsf{HA}^*}$.

(b) As we remarked in the Introduction, Solovay's original proof has an incredible upward stability. That is, we can use almost the very same proof to show that any Σ_1-sound theory extending PA has the same provability logic as PA. In the present context, this strength is worrisome. If one wishes to gives a Solovay-style proof, one might aim for a theorem like: 'if $T \supseteq \mathsf{HA}$ is a Σ_1-sound theory such that $\vdash_T \mathrm{CT}_T$, then the provability logic of T is equal to the set of theorems of iGLC'. This 'theorem' is false, however, since PA^* is a counterexample. If we want to show that the provability logic of HA^* is equal to the set of theorems of iGLC, we will need a way to distinguish HA^* from PA^* in our proof.

(iii) What is the provability logic of HA? We have not really touched upon this question, and it remains wide open. The Σ_1-provability logic of HA at least provides a non-trivial upper bound, but we know that this upper bound cannot be strict, since $A \to \Box A$, which is in the Σ_1-provability logic of HA, is definitely not in the provability logic of HA itself.

Index of symbols

$\mathcal{A}$, 10
$[A]$, 24
$[\![A]\!]$, 22
A^*, 32
A^+, 33
A^T, 7
Ax_T, 4
Bew_T, 5
Bew_T^f, 12
$\Box$
  as a modal operator, 16
  for a provability predicate, 6
$\Box_T$, 5
$\Box_T^f$, 12
$\Box_T^{u+1}$, 12
CP, 11
Δ_0, 3
EA, 4
exp, 3
$\Vdash$, 16
Form, 7
h, 23
HA, 4
$\widehat{\mathsf{HA}}$, 36
HA^*, 35
IBew_T, 12
iGL, 16
iGLC, 16
iK, 16
$\mathcal{L}$, 3
$\mathcal{L}_\Box$, 16
$\mathcal{L}_p$, 31
$\overline{n}$, 3
Nec, 16
NNIL, 31
PA, 4
PA^*, 35
$\mathsf{Prf}_\Box$, 23
$\mathsf{Prf}_\triangle$, 23
Prf_T, 5
S, 3
sHA, 15
sPA, 15
σ, $\sigma_\Box$, σ_T, σ_T^f, 29
Σ_1, 3
$s \leq t$, 3
$s < t$, 3
SLP, 11
Subst, 4
Th(T), 4
TNNIL, 33
$\triangle$, 21
U^*, 7
U^T, 7
$U \subseteq T$, 5
$U = T$, 5
$U \leq T$, 5
$U \equiv T$, 5
$\vdash_{\mathsf{IPC}}$, 31
$\vdash_T$
  for a theory, 4
  for a theory for IML, 16
$\tilde{x}$, 5

Index of terms

Adequate set, 18
Axiom, 3

Completeness principle, 11
Convention on theories, 5
Conversely well-founded, 17

Diagonalization Lemma, 5

Elementary Arithmetic, 4
Extension Lemma, 18

Fast-growing function, 14
Fast provability, 12
Forcing relation, 16
Frame for IML, 16

Gödel number, 4
Good pair, 21

Heyting Arithmetic, 4
    Slow Heyting Arithmetic, 15

Induction axiom, 4
Intuitionistic modal logic, 16
    IML, 16
Iterated provability, 12

Language
    of arithmetic, 3
    of IML, 16
    of intuitionistic propositional logic, 31
Law of the Excluded Middle, 4
Löb's Principle, 5
Löb's Theorem, 5

(Σ_1-)logic, 29

Model for IML, 16
Model property, 16

Natural deduction with equality, 3
NNIL-algorithm, 32
Numeral, 3

Peano Arithmetic, 4
Preservativity of Knowledge, 17
(Fast) (Σ_1-)provability logic, 29
Provability predicate, 6

Realistic, 17
(Σ_1-)realization, 29
Reflexive induction, 12

Semi-transitive, 17
Σ_1-completeness, 4
    Formalized Σ_1-completeness, 5
Slow provability, 14
Solovay function, 23
Standard model, 3
Strong Löb principle, 11

T-translation, 7
Theory, 4
    for IML, 16
TNNIL-algorithm, 33
Transitive, 17

Valid, 17

X-saturated set, 18

Bibliography

[1] M. Ardeshir and S. Mojtaba Mojtahedi. The Σ_1 Provability Logic of HA. *ArXiv e-prints*. Available at `arxiv.org/abs/1409.5699v1`.

[2] George Boolos. *The Logic of Provability.* Cambridge University Press, 1993.

[3] Petr Hájek and Pavel Pudlák. *Metamathematics of First-Order Arithmetic.* Springer-Verlag, 1993.

[4] Paula Henk. *Nonstandard Provability for Peano Arithmetic: A Modal Perspective*. PhD thesis, Institute for Language, Logic and Computation, University of Amsterdam, 2016.

[5] Paula Henk and Fedor Pakhomov. Slow and Ordinary Provability for Peano Arithmetic. *ArXiv e-prints*. Available at `arxiv.org/abs/1602.01822`.

[6] R. Iemhoff. A Modal Analysis of Some Principles of the Provability Logic of Heyting Arithmetic. In *Proceedings of AiML'98*, volume 2, 2001.

[7] Per Lindström. On Parikh Provability: An Exercise in Modal Logic. In S. Lindström H. Lagerlund and R. Sliwinski, editors, *Modality Matters: Twenty-Five Essays in Honour of Krister Segerberg*, volume 53 of *Uppsala Philosophical Studies*. Uppsala Universitet, 2006.

[8] Rohit Parikh. Existence and Feasibility. *Journal of Symbolic Logic*, 36(3):494–508, 1971.

[9] A.S. Troelstra and D. van Dalen. *Contructivism in Mathematics I*, volume 121 of *Studies in Logic and the Foundations of Mathematics.* Elsevier, 1988.

[10] Albert Visser. On the Completeness Principle: A Study of Provability in Heyting's Arithmetic and Extensions. *Annals of Mathematical Logic*, 22(3):263–295, 1982.

[11] Albert Visser. Substitutions of Σ^0_1-sentences: Explorations between Intuitionistic Propositional Logic and Intuitionistic Arithmetic. *Annals of Pure and Applied Logic*, 114:227–271, 2002.